Article

The CP2K Program Package Made Simple

Marcella Iannuzzi, Jan Wilhelm, Frederick Stein, Augustin Bussy, Hossam Elgabarty, Dorothea Golze, Anna-Sophia Hehn, Maximilian Graml, Stepan Marek, Beliz Sertcan Gökmen, Christoph Schran, Harald Forbert, Rustam Z. Khaliullin, Anton Kozhevnikov, Mathieu Taillefumier, Rocco Meli, Vladimir V. Rybkin, Martin Brehm, Robert Schade, Ole Schütt, Johann V. Pototschnig, Hossein Mirhosseini, Andreas Knüpfer, Dominik Marx, Matthias Krack, Jürg Hutter, and Thomas D. Kühne*

Cite This: https://doi.org/10.1021/acs.jpcb.5c05851

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: CP2K is a versatile open-source software package for simulations across a wide range of atomistic systems, from isolated molecules in the gas phase to low-dimensional functional materials and interfaces, as well as highly symmetric crystalline solids, disordered amorphous glasses, and weakly interacting soft-matter systems in the liquid state and in solution. This review highlights CP2K's capabilities for computing both static and dynamical properties using quantum-mechanical and classical simulation methods. In contrast to the accompanying theory and code paper [*J. Chem. Phys.* 152, 194103 (**2020**)], the focus here is on the practical usage and applications of CP2K, with underlying theoretical concepts introduced only as needed.
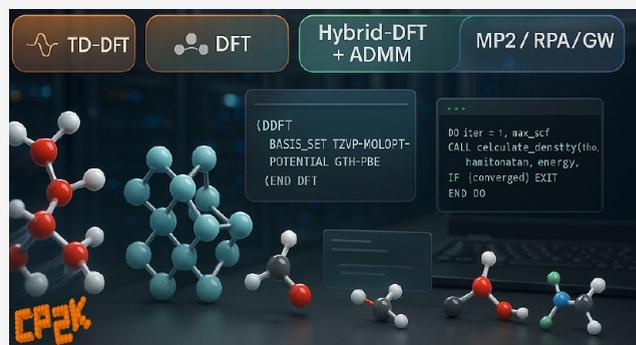
1. INTRODUCTION

The aim of CP2K is to predict average properties, such as those that arise in statistical mechanics and thermodynamics, of any substance made of interacting electrons and nuclei from first-principles. Although with the advent of quantum mechanics this became possible by solving the nonrelativistic many-body Schrödinger equation, early on Dirac recognized that the exact application of these laws leads to equations much too complicated to be soluble.[1] Yet, assuming the so-called Born–Oppenheimer approximation, which is underlying most of the methods within CP2K, the full many-body Schrödinger equation can be separated into its electronic and nuclear counterparts.

Consequently, the general structure of CP2K entails a wide variety of different classical and quantum mechanical energy and force methods that can be arbitrarily combined with geometry optimization and transition state search techniques,[2,3] as well as (rare event) sampling approaches, such as Monte Carlo (MC), meta- and molecular dynamics (MD),[4–6] to name just a few. The latter approaches dealing with the nuclear motion are specified in the &MOTION section, whereas the former force methods, which are due to the electronic structure of matter, are detailed in the &FORCE_EVAL section. However, since we have elected to limit ourselves to computational schemes that make CP2K unique, the techniques of the &MOTION section, which are well described

in various textbooks,[7,8] will be mostly neglected. The same also applies to ab initio MD (AIMD),[9,10] in particular the second-generation Car–Parrinello method,[11,12] which has been extensively covered elsewhere.[13–15]

Due to the user-focused nature of the present review, the paper is organized by the property to be calculated instead of the computational methods used, which will be introduced in the relevant sections where necessary. Hence, first how to compute the total energy and nuclear forces of a system using density functional theory (DFT)[16] and quantum chemical models[17] is presented in Section 2. Thereafter, the calculation of the electronic band structure using the pseudopotential plane wave (PP-PW),[18] projector augmented-wave (PAW),[19] full-potential linearized augmented plane wave (FP-LAPW),[20] as well as single-particle Green's function and the screened Coulomb interaction (*GW*) methods,[21] is detailed in Section 3. Embedding schemes such as implicit solvation and mixed classical-quantum mechanical approaches, as well as purely quantum mechanical embedding theories, are the content of

Section 4. Nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) spectroscopies are described in Section 5, whereas optical spectroscopy methods such as linear-response time-dependent DFT (LR-TDDFT) and the Bethe−Salpeter equation (BSE) are the subject matter of Section 6. Section 7 is dedicated to excited state dynamics in terms of real-time TDDFT (RT-TDDFT), Ehrenfest dynamics and real-time BSE (RT-BSE), whereas X-ray absorption (XAS) and X-ray emission (XES) spectroscopies are described in Section 8. Energy decomposition analysis (EDA) based on absolutely localized molecular orbitals (ALMO) for condensed phase systems is the focus of Section 9. Finite temperature effects using machine learning potentials (MLPs) including nuclear quantum effects (NQEs) by means of path-integral MD (PIMD) simulations are presented in Section 10, before concluding the paper on the extended topic of vibrational spectroscopy.

2. TOTAL ENERGY AND FORCE METHODS

Even though total energies and atomic forces are rarely relevant observables for themselves, they are the basis for the theoretical computation of all properties considered here. In the spirit of CP2K, to simulate all possible kinds of matter from isolated molecules to the extended condensed phase, all energy and force methods can not only be employed in vacuo (0D), but also in the presence of arbitrary periodic boundary conditions (PBC), i.e. 1D (chains and nanotubes), 2D (surfaces and interfaces) and 3D (solids and liquids). Moreover, to explicitly include finite temperature and pressure effects by means of dynamical sampling methods, analytical gradients are generally available for all classical and QM electronic structure methods, even in combination with PBCs.

This is made possible by the QUICKSTEP module,[14,22,23] which is based on the Gaussian and plane wave (GPW) method and its all-electron counterpart denoted as the Gaussian and augmented plane wave (GAPW) approach,[24,25] and contains all electronic structure methods described in Section 2, i.e. in particular DFT and Hartree−Fock (HF),[26] but also hybrid-DFT and post-HF schemes such as second-order Møller−Plesset perturbation theory (MP2)[27] and the random phase approximation (RPA).[28,29]

2.1. Density Functional Theory

The unique aspect of the GPW and GAPW methods is their usage of a mixed Gaussian and an auxiliary plane wave (PW) basis set, thereby unifying the computational efficiency of a compact localized basis set with the simplicity of PWs for periodic electronic structure calculations. Within CP2K/QUICKSTEP, the Kohn−Sham (KS) orbitals $\phi_i(\mathbf{r})$ are represented by atom-centered Gaussian functions, whereas the electron density is expanded in PWs on a regular grid. This is to say that the molecular orbitals (MO) are represented by the common linear combination of atomic orbitals (LCAO)

$$\phi_i(\mathbf{r}) = \sum_j c_{ji} \varphi_j(\mathbf{r}) \tag{1}$$

where c_{ji} are the so-called orbital coefficients, and the atomic orbitals (AO)

$$\varphi_j(\mathbf{r}) = \sum_k d_{kj} g_k(\mathbf{r}) \tag{2}$$

are themselves expanded in terms of primitive Gaussian functions $g_k(\mathbf{r})$ and their corresponding contraction coefficients d_{kj}.

The electron density $\rho(\mathbf{r})$ can either be written as

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P^{\mu\nu} \varphi_\mu(\mathbf{r}) \varphi_\nu(\mathbf{r}) \tag{3}$$

or alternatively in an auxiliary PW basis as

$$\tilde{\rho}(\mathbf{r}) = \frac{1}{\Omega} \sum_{\mathbf{G}} \rho(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \tag{4}$$

where Ω is the volume of the unit cell and the expansion coefficients $\rho(\mathbf{G})$ are such that $\rho(\mathbf{r}) = \tilde{\rho}(\mathbf{r})$.

Exploiting the efficiency of the fast Fourier transform (FFT), the KS matrix construction including the periodic long-range electrostatics can be performed in $O(M\log M)$ with M being the number of basis functions, via

$$\mathbf{P} \to \rho(\mathbf{r}) \xrightarrow{\text{FFT}} \rho(\mathbf{G}) \to 4\pi\frac{\rho(\mathbf{G})}{G^2}$$
$$\|$$
$$\mathbf{V} \leftarrow V_{\text{H}}(\mathbf{r}) \xleftarrow{\text{FFT}^{-1}} V_{\text{H}}(\mathbf{G}) \tag{5}$$

$V_{\text{H}}(\mathbf{G})$ is the Hartree potential that is the solution of the Poisson equation $\nabla^2 V_{\text{H}}(\mathbf{G}) = -4\pi\rho(\mathbf{r})$. By substituting $\rho(\mathbf{r})$ with the total charge density, which consists of the electron density plus Gaussian-smeared nuclear charges, all long-range electrostatic energy terms can be coalesced into a single Hartree-like expression that has to be augmented by compensating overlap and self-energy terms generated by these Gaussian distributions. Similarly, the exchange and correlation (XC) functional is evaluated on the very same uniform density grid as the Hartree energy, so that starting from $\rho(\mathbf{r})$, the combined potential for the long-range electrostatic and XC energies can be computed using FFT techniques.[14,24,25]

2.1.1. Basis Set Convergence Using the QUICKSTEP Method. This is why in a typical QUICKSTEP calculation, which is activated via &FORCE_EVAL%METHOD QUICKSTEP, the basis has to be defined in two sections. On the one hand, the localized Gaussian basis set is defined in the &FORCE_EVAL%SUBSYS%KIND section and, on the other hand, the PW density cutoff in the &FORCE_EVAL%DFT%MGRID section. It is important to recognize, however, that the latter is a density cutoff and therefore 4 times higher than a comparable energy cutoff typically specified in PW codes, which is a manifestation of Nyquist's sampling theorem together with the fact that the electron density is the squared modulus of the wave function (WF). Also, contrary to conventional PW codes, the complete basis set limit cannot be reached just by increasing the density cutoff of the auxiliary PW basis only. Instead, by increasing CUTOFF the limit of the finite Gaussian set is approached, which is why the convergence of the auxiliary PW basis depends on the underlying Gaussian basis and both have to be balanced and increased concurrently to yield the complete basis set limit.

An important aspect of the QUICKSTEP method is the usage of real-space integration grids to represent the electron density and product Gaussian functions in conjunction with multigrid techniques, so that wide and smooth Gaussian functions are mapped onto a coarser grid than narrow and sharp Gaussians.

However, the electron density is always mapped onto the finest grid. Hence, choosing a fine enough integration grid is crucial for any calculation to obtain highly accurate results as efficiently as possible. All settings related to multigrids are controlled within the `&MGRID` subsection. The number of multigrid levels is defined by `NGRIDS`, with 5 typically being a suitable value for most applications. As alluded to above, the keyword `CUTOFF` defines the PW density cutoff (in Rydbergs, Ry), the corresponding cutoffs for subsequent grid levels (from finer to coarser) are defined via

$$E_{cut}^i = \frac{E_{cut}^1}{\alpha^{i-1}} \tag{6}$$

where i is the corresponding multigrid level and α the progression factor controlled by `PROGRESSION_FACTOR`. Hence, the higher the value of `CUTOFF`, the finer the grids of all multigrid levels. To determine which product Gaussians are mapped onto which multigrid level, `REL_CUTOFF` defines the PW cutoff of a reference grid covered by a Gaussian with unit standard deviation, i.e. $\exp(|\mathbf{r}|^2)$. Therewith, a Gaussian is mapped onto the coarsest level of the multigrid, on which the function will require a number of grid points greater than or equal to the number of grid points $\exp(|\mathbf{r}|^2)$ will cover on a reference grid defined by `REL_CUTOFF`. In this way, CP2K tries to map each Gaussian onto a grid such that the number of grid points covered by the Gaussian, no matter how wide or narrow, is roughly the same. Hence, the two most important keywords affecting real-space integration grids and, as such, the convergence of the auxiliary PW basis are `CUTOFF` and `REL_CUTOFF`, respectively. A prototypical CP2K input section, using the GPW method, reads as follows:

```
&FORCE_EVAL
    METHOD QUICKSTEP
    &DFT
        &QS
            METHOD GPW
        &END QS
        &MGRID
            CUTOFF 400
            REL_CUTOFF 50
            NGRIDS 5
            PROGRESSION_FACTOR 3.0
        &END MGRID
    &END DFT
&END FORCE_EVAL
```

2.1.2. Pseudopotentials and Basis Sets. Besides the auxiliary PW basis, the primary Gaussian basis set and optionally also the employed pseudopotentials (PP) are essential for the accuracy of any QUICKSTEP calculation. For that purpose, in recent years the so-called "UZH protocol" has been developed that contains both molecularly optimized Gaussian basis sets for all-electron and PP calculations,[30] as well as corresponding separable dual-space Goedecker-Teter-Hutter (GTH) PPs that are highly transferable and norm-conserving.[31−33]

In addition to accelerating the calculation of the electronic structure by reducing the number of electrons and the basis set size, the usage of PPs also allows for the inclusion of relativistic effects. The separable dual-space GTH PPs of the "UZH protocol" are fully nonlocal with Gaussian-type projectors because of their simplicity and efficiency due to the usage of analytic integrals and FFTs. The local part, however, is represented by a Gaussian form in real-space, which can be

well combined with Gaussian basis sets. Hence, the fully analytical form of the GTH PPs requires only a small set of parameters for each element that is optimized with respect to an atomic all-electron WF of scalar relativistic DFT reference calculations,[34] as obtained by the CP2K `ATOM` code. The "UZH protocol" contains GTH PPs for the periodic table up to Rn for generalized gradient approximation (GGA) (PBE[35]), meta-GGA (SCAN[36]), and hybrid functionals (PBE0[37]). Special purpose GTH PPs for lanthanides and actinides,[38−41] as well as including nonlinear core corrections and spin−orbit coupling parameters, are also available at https://github.com/cp2k/cp2k/tree/master/data.[42] In addition, via the libgrpp library to evaluate molecular integrals over Gaussian functions,[43] most effective core potentials (ECP) provided by the EMSL Basis Set Exchange can be employed.[44]

The basic idea of the molecular-optimized MOLOPT basis set is to use generally contracted Gaussian basis sets, including diffuse primitives, fully optimized on molecular calculations.[30] The absence of lone diffuse functions ensures a low condition number of the overlap matrix, leading to better self-consistent field (SCF) convergence and a sparser density matrix, whereas the inclusion of diffuse primitive functions entails a particularly small basis set superposition error. By molecularly optimizing the basis functions, small but accurate basis sets are obtained, suitable for large-scale isolated gas and condensed phase calculations. As with PPs, in addition to the standard general purpose "UZH protocol", more specialized MOLOPT basis sets for lanthanides and actinides,[38−41] solids,[45−49] and for the later described auxiliary density matrix method (ADMM),[50] GW,[51,52] and resolution of the identity approaches (RI),[53,54] such as RI-MP2 and RI-RPA, are also available in the same repository. In addition, the "UZH protocol" also contains all-electron MOLOPT basis sets, although all other basis sets that can be downloaded at http://basissetexchange.org in CP2K format are also supported.[44]

A typical section for gold as an example looks as follows:

```
&FORCE_EVAL
    &DFT
        BASIS_SET_FILE_NAME BASIS_MOLOPT_UZH
        POTENTIAL_FILE_NAME POTENTIAL_UZH
    &END DFT
    &SUBSYS
        &KIND Au
            BASIS_SET TZV2P-MOLOPT-GGA-GTH-q19
            POTENTIAL GTH-GGA-q19
        &END KIND
    &SUBSYS
&END FORCE_EVAL
```

The accuracy of the "UZH protocol" can best be assessed by means of all-electron FP-LAPW calculations at the complete basis set limit using the CP2K-integrated PW code SIRIUS, which is described in Section 3.1.

The deviation between SIRIUS and QUICKSTEP employing a TZV2P MOLOPT basis set and corresponding GTH PPs is shown in Figure 1 for a large number of different unary crystal structures and quantified in terms of the previously suggested metric

$$\varepsilon(a, b) = \sqrt{\frac{\sum_i [E_a(V_i) - E_b(V_i)]^2}{\sqrt{\sum_i [E_a(V_i) - \langle E_a \rangle]^2 \ \sum_i [E_b(V_i) - \langle E_b \rangle]^2}}} \tag{7}$$

where the index i runs over multiple calculations of $E(V)$ within a given space group.[55] Since this quantity is insensitive
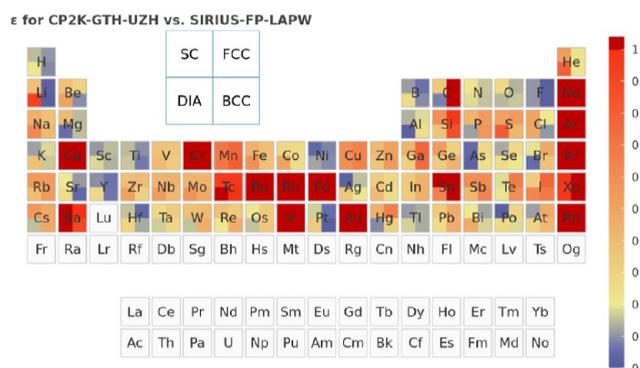
Figure 1. Values of the comparison metric ε for CP2K/Quickstep using the "UZH protocol" with respect to all-electron FP-LAPW calculations using the CP2K/SIRIUS code. For each element up to Rn four monoelemental cubic crystals were considered.[55]

to the magnitude of the bulk modulus, it provides a uniform metric across a large variety of structural and chemical environments.

2.1.3. All-Electron Calculations Using the GAPW Method.

All-electron calculations with CP2K/Quickstep necessitate the use of the GAPW method,[25,56] because a description of the core electrons with PWs becomes rapidly computationally prohibitive. Likewise, small-core and some-times even medium-core PPs can only be used with GAPW efficiently. There are various possibilities to devise a PP for heavy elements such as uranium with 92 electrons:

- large-core PP ([86 core] 6 valence electrons)
- [Rn] $7s^2$ $5f^3$ $6d^1$
- medium-core PP ([78 core] 14 valence electrons)
- [Xe $4f^{14}$ $5d^{10}$] $6s^2$ $6p^6$ $7s^2$ $5f^3$ $6d^1$
- small-core PP ([60 core] 32 valence electrons)
- [Kr $4d^{10}$ $4f^{14}$] $5s^2$ $5p^6$ $5d^{10}$ $6s^2$ $6p^6$ $7s^2$ $5f^3$ $6d^1$

The large-core PP for uranium is known to be inaccurate due to the non-negligible overlap of the outer valence orbitals with the semicore 6s and 6p orbitals, which does not allow keeping these semicore orbitals as frozen. Medium-core PPs for uranium are mostly used by PW codes[57] and with GPW,[58] because these PPs are sufficiently accurate and the semicore 6s and 6p orbitals are not too hard for an efficient expansion by PWs. However, this is not possible for small-core uranium PPs that require GAPW.

Furthermore, the calculation of several properties, especially spectroscopic properties involving core electrons, requires the explicit inclusion of the core electrons.[59,60] The GAPW method must be activated explicitly in the `&DFT%QS` section of the input, i.e.

```
&QS
    METHOD GAPW
    EPSFIT 1.0E-6
    EPSRHO0 1.0E-8
    EPSSVD 1.0E-12
&END QS
```

The GAPW specific parameters `EPSFIT`, `EPSRHO0` and `EPSSVD` are optional and their default values are 10^{-4}, 10^{-6} and 10^{-8}, respectively. These parameters are used to control the accuracy of a GAPW calculation as indicated by the electron count printed after SCF convergence has been achieved. `EPSFIT` controls the split of Gaussian basis sets into a hard and a soft basis set for each atomic kind in the

system. Smaller values for `EPSFIT` promote the inclusion of larger Gaussian exponents in the soft basis. Only the part of the electron density defined by the soft basis set is expanded in PWs, whereas the hard basis set describes the hard (frozen) part of the electron density close to the nuclei. A GAPW setup seamlessly turns into a GPW one when all exponents are included in the soft basis sets and all hard basis sets are "empty". In this case, the default `METHOD GPW` can be used, since GAPW is not required. Gaussian exponents up to 5 should always be included in the soft basis set. Check the values in the first column of the Gaussian basis set definition or the basis set printout in the CP2K output for `PRINT_LEVEL MEDIUM` or higher. The electron count should show at least an accuracy of 1.0×10^{-6}. For improved accuracy, Gaussian exponents in the range 5−10 can be added to the soft basis by lowering the value of `EPSFIT`. The inclusion of larger exponents should always be accompanied by an adequate increase of the density `CUTOFF` in `&MGRID`. `EPSRHO0` and `EPSSVD` improve the numerical precision of $V(\rho_0 - \rho_0^{\text{soft}})$ and the singular value decomposition (SVD) of the PAW projector matrix, respectively. In this way, an electron count of 1.0×10^{-8} or even less can be achieved. Gaussian exponents up to about 20 can still be expanded in PWs using large `CUTOFF` values, but the expansion of larger exponents quickly becomes computationally prohibitive or even impossible.

A few atomic kind-specific settings complete a GAPW all-electron setup:

```
&KIND O
    BASIS_SET ORB TZVPP-MOLOPT-GGA-ae
    POTENTIAL ALL
    LEBEDEV_GRID 400
    RADIAL_GRID 100
&END KIND
```

The atomic `POTENTIAL` must be set to `ALL` and an all-electron `ORB`ital basis set has to be selected. Note that most of these all-electron basis sets are made for isolated (0D) systems such as molecules or clusters in the gas phase, but there are also a few all-electron basis sets suited for condensed phase calculations with CP2K/Quickstep, as described in Section 2.1.2. For heavier elements, ECPs are employed that require GAPW like small core GTH PPs, since an all-electron description for condensed phase systems with heavy elements quickly becomes computationally intensive. There are usually Gaussian basis sets specifically optimized for such ECPs.

Optionally, the size of the atom-centered spherical Lebedev grids can be increased to improve the accuracy of the GAPW electron count for atomic densities and, therefore, also for total energies and nuclear forces. The number of `RADIAL_GRIDS` can be increased from the default value 50 to 100 or more, whereas only spherical Lebedev grids with certain sizes exist. The atomic grids are automatically generated in an adaptive manner.[61] The default value for the maximum `LEBEDEV_GRID` is 50 and thus is rather small. The following Lebedev grids are implemented in CP2K: 6, 14, 26, 38, 50, 86, 110, 146, 194, 302, 434, 590, 770, 974. The grids close to and far from the nuclei are small because the electron density is rather spherical at these distances, while the selected maximum grid is employed in the valence regions. For instance, `LEBEDEV_GRID 400` limits the maximum employed Lebedev grid for an atomic kind to 302.

2.1.4. Wave Function Optimization.

Since the Hamiltonian of effective single-particle theories, such as DFT, HF, semiempirical quantum chemistry (SQC) and tight-binding

(TB), depends on the electron density, i.e. on their own solution, the corresponding equations have to be solved iteratively using the SCF approach. The ability to converge this SCF cycle and to achieve self-consistency depends on multiple factors such as the band gap and dimensionality of the system, as well as the locality of the employed basis functions.

CP2K/QUICKSTEP provides two general techniques to yield self-consistency: traditional diagonalization and the orbital transformation (OT) method.[62] The former approach entails the direct diagonalization of the Hamiltonian followed by a mixing strategy, such as the Kerker, Pulay, multisecant, and Broyden techniques.[63−66] In these damped preconditioned fixed-point iteration schemes, the electron density of the next step is constructed in terms of a linear combination of previous densities, in order to achieve a robust SCF convergence with the fewest number of steps. For hardly converging systems, such as metallic solids for which finite electron temperature smearing and $\mathbf{k}$-point sampling may also be required, Broyden mixing is typically the best choice, and the associated settings in the following input snippet are a good starting point:

```
&QS
    EPS_DEFAULT 1.0E-12
&END QS
&SCF
    MAX_SCF 100
    EPS_SCF 1.0E-6
    SCF_GUESS RESTART
    ADDED_MOS 10
    &MIXING
        METHOD BROYDEN_MIXING
        ALPHA 0.2
        BETA 0.8
        NBROYDEN 10
    &END MIXING
    &SMEAR
        METHOD FERMI_DIRAC
        ELECTRONIC_TEMPERATURE [K] 300
    &END SMEAR
&END SCF
```

Therein, the keyword `EPS_SCF` denotes the SCF convergence threshold $\|\mathbf{P}_i - \mathbf{P}_{i-1}\|_{max}$, where $\mathbf{P}_i$ is the density matrix of the i-th iteration. Most other threshold values of the `&DFT%QS` section are connected to `EPS_DEFAULT`, whose value should be approximately that of `EPS_SCF` squared to ensure a sufficiently high numerical accuracy in order to reach the desired SCF convergence threshold. For metallic systems, finite electron temperature or smearing the density of states, which is specified in the `SMEAR` section, is often used to improve the convergence with respect to Brillouin zone sampling that is described in Section 2.1.5. Besides the temperature, however, the occupation function must be given, whose natural choice is the physical Fermi−Dirac distribution that would lead to the grand-canonical extension to DFT of Mermin.[67]

However, for gapped systems, a computationally much more efficient way to locate the electronic ground state is by variationally minimizing the total energy with respect to the orbital coefficient matrix $\mathbf{C}$ subject to the orthonormality constraint $\mathbf{C}^T\mathbf{S}\mathbf{C} = \mathbf{I}$ since electrons are Fermions. Inspired by the exponential transformation,[68] the occupied orbitals are parametrized in terms of an auxiliary variable $\mathbf{X}$, i.e.

$$\mathbf{C}(\mathbf{X}) = \mathbf{C}\mathbf{U}\cos(\mathbf{U}) + \mathbf{X}\sin(\mathbf{U}) \tag{8}$$

where

$$\mathbf{U} = \sqrt{\mathbf{X}^T\mathbf{S}\mathbf{X}} \tag{9}$$

This allows for an unconstrained optimization using the OT method, given that $\mathbf{X}$ obeys the linear constraint $\mathbf{X}^T\mathbf{S}\mathbf{C} = \mathbf{0}$.

In addition to the minimization scheme employed, the choice of a preconditioner is essential. In this regard, the ability to converge the KS equations has to be carefully balanced against the number of necessary minimization steps, which should not be confused with SCF iterations, and the computational effort to construct the preconditioner. Hence, for well converging large-scale systems, preconditioners based on the Cholesky inversion of $\mathbf{H} - \epsilon_0\mathbf{S}$ (`FULL_SINGLE_IN-VERSE`),[62] or $\mathbf{T} + \epsilon_0\mathbf{S}$ (`FULL_KINETIC`),[69] where ϵ_0 is an estimate of the highest eigenvalue of $\mathbf{C}^T\mathbf{H}\mathbf{C}$ and $\mathbf{T}$ the kinetic energy matrix, are typically offering the computationally best price/performance ratio. However, contrary to the aforementioned SCF approach, the convergence criterion employing OT is tested in terms of the preconditioned mean gradient deviation, which is why it has to be chosen approximately 1 order of magnitude tighter to achieve a comparable convergence than using direct diagonalization. In case convergence is not foreseeable after approximately 20 OT minimization steps, an additional loop can be introduced via an `&OUTER_SCF` section, whose purpose is not to conduct an outer-loop optimization, but solely to update the preconditioner. For even more difficult cases, the diagonalization-based `FULL_ALL` preconditioner should be used, which relies strongly on an estimate of the HOMO − LUMO gap. Since this must not overestimate the true band gap, `ENERGY_GAP` can hardly be set too low.

Among the various implemented optimization techniques, the direct inversion of the iterative subspace (DIIS) scheme is the method of choice as long as the fixed-point iteration is convergent.[64] The `SAFE_DIIS` option automatically switches to the steepest descent (SD) approach whenever a DIIS step points away from the minimum. But, since this leads to an unreasonably slow convergence behavior, this should be prevented from the outset by setting `STEPSIZE 0.1` or even smaller, i.e.

```
&OT
    MINIMIZER DIIS
    SAFE_DIIS TRUE
    PRECONDITIONER FULL_ALL
    ENERGY_GAP 0.0001
    STEPSIZE 0.1
&END OT
&OUTER_SCF
    EPS_SCF 1.0E-7
    MAX_SCF 10
&END OUTER_SCF
```

If this is not successful, it is advisable to shift to the conjugate gradient (CG) method in conjunction with `LINESEARCH ADAPT`, and otherwise to the traditional diagonalization alluded to above.

2.1.5. Brillouin-Zone Sampling Using k-Points. It is possible to run periodic calculations with $\mathbf{k}$-point sampling in CP2K, with both the GPW and GAPW methods. For each $\mathbf{k}$-point in the first Brillouin zone, a generalized complex Hermitian eigenvalue problem must be solved, i.e.

$$\mathbf{K}^\mathbf{k}\mathbf{C}^\mathbf{k} = \mathbf{S}^\mathbf{k}\mathbf{C}^\mathbf{k}\boldsymbol{\varepsilon}^\mathbf{k} \tag{10}$$

where $\mathbf{K}^\mathbf{k}$, $\mathbf{S}^\mathbf{k}$, $\mathbf{C}^\mathbf{k}$ and $\boldsymbol{\varepsilon}^\mathbf{k}$ are the $\mathbf{k}$-point dependent KS matrix, overlap matrix, MO coefficients and eigenvalues, respectively. In CP2K/QUICKSTEP, the KS and overlap matrices are first calculated in real-space, before being Fourier transformed to reciprocal-space. Hence, the overlap matrix elements read as

$$S^\mathbf{k}_{\mu\nu} = \sum_\mathbf{R} e^{i\mathbf{k}\cdot\mathbf{R}} S^\mathbf{R}_{\mu\nu} \tag{11}$$

with

$$S^\mathbf{R}_{\mu\nu} = \int \varphi_\mu(\mathbf{r}) \varphi_\nu(\mathbf{r} - \mathbf{R}) d\mathbf{r} \tag{12}$$

where $\mathbf{R}$ is a unit cell translation vector. Similarly, $K^\mathbf{R}_{\mu\nu}$ is the KS matrix element of an AO φ_μ in the unit cell and φ_ν in a periodic cell translated by $\mathbf{R}$, respectively. Note that the number of translation vectors $\mathbf{R}$ depends on the diffuseness of the basis set: periodic images are considered as long as one of their basis functions overlaps with the unit cell. The same amount of work is done in Γ-point calculations, except that contributions from neighboring cells are added up in a single matrix.

A standard Γ-point DFT input file can be modified for a $\mathbf{k}$-point calculation by adding a `&KPOINTS` subsection in `&DFT`:

```
&DFT
   &KPOINTS
      SCHEME MONKHORST-PACK 8 8 4
   &END KPOINTS
   # Rest of the DFT input
&END DFT
```

In this example, an $8 \times 8 \times 4$ Monkhorst−Pack $\mathbf{k}$-point mesh is generated.[70] Note that every $\mathbf{k}$-point can also be specified individually with its coordinates and weight using the `KPOINT` keyword. At the time of writing, CP2K does not automatically exploit symmetries to reduce the number of $\mathbf{k}$-points (except for time-reversal symmetry, i.e. $\mathbf{k} = -\mathbf{k}$). After a converged SCF calculation, band structure data can be generated by adding the following `&PRINT` subsection:

```
&PRINT
   &BAND_STRUCTURE
      # Output file:
      FILE_NAME graphene.bs
      # Number of empty bands:
      ADDED_MOS 2
      # Specifying the band path:
      &KPOINT_SET
         NPOINTS 10
         UNITS B_VECTOR
         SPECIAL_POINT G 0.0 0.0 0.0
         SPECIAL_POINT K 1/3 1/3 0.0
         SPECIAL_POINT M 1/2 0.0 0.0
         SPECIAL_POINT G 0.0 0.0 0.0
      &END KPOINT_SET
   &END BAND_STRUCTURE
&END PRINT
```

Here, a graphene band structure is calculated along the $\Gamma-$K$-$M$-\Gamma$ path, with 10 $\mathbf{k}$-points along each segment. Note that any path of any density can be generated, since all relevant quantities are calculated in real-space first, before being Fourier transformed to the specified path. The only requirement is that the original SCF uses a sufficiently converged $\mathbf{k}$-point mesh, as defined in `&KPOINTS`.

Finally, note that $\mathbf{k}$-points are not yet implemented for all electronic structure methods, but primarily for DFT, resolution-of-the-identity HF exchange (RI-HFX), and the *GW* approximation. It also has limited support with respect to SCF schemes: only full diagonalization (with DIIS) is available, but OT is not. As for Γ-point calculations, it is possible to print binary WF restart files via the `&SCF%PRINT%RESTART` section when running $\mathbf{k}$-point calculations. However, due to different format and content, restarting a $\mathbf{k}$-point calculation from a Γ-point restart file (and vice versa) is not possible.

2.2. Hartree−Fock and Hybrid Density Functional Theory

The central quantity in HF and hybrid DFT calculations is the Fock matrix $\mathbf{F}$. For nonperiodic calculations, it is simply defined as

$$F_{\mu\nu} = \sum_{\sigma,\lambda} P_{\sigma\lambda}(\mu\sigma|\nu\lambda) \tag{13}$$

where μ, ν, σ, λ defines elements of the AO basis, whereas $P_{\sigma\lambda}$ is the density matrix, and $(\mu\sigma|\nu\lambda)$ represents 2-electron 4-center electron repulsion integrals (ERI), which in the Mulliken notation reads as

$$(\mu\sigma|\nu\lambda) = \int d\mathbf{r} \int d\mathbf{r}' \varphi_\mu(\mathbf{r})\varphi_\sigma(\mathbf{r})$$
$$\times g(|\mathbf{r} - \mathbf{r}'|) \, \varphi_\nu(\mathbf{r}')\varphi_\lambda(\mathbf{r}') \tag{14}$$

Therein $g(|\mathbf{r} - \mathbf{r}'|)$ is the interaction potential, conventionally chosen to be the $1/|\mathbf{r} - \mathbf{r}'|$ Coulomb potential.

In periodic HF calculations, however, contributions from periodic images of the unit cell need to be explicitly accounted for. Elements of the Fock matrix that involve an AO $\varphi_\mu(\mathbf{r})$ in the unit cell and an AO $\varphi_\nu(\mathbf{r} - \mathbf{b})$ in a periodic image translated by $\mathbf{b}$ are defined as

$$F^\mathbf{b}_{\mu\nu} = \sum_{\sigma,\lambda} \sum_{\mathbf{a},\mathbf{c}} P^\mathbf{c}_{\sigma\lambda}(\mu^\mathbf{0}\sigma^\mathbf{a}|\nu^\mathbf{b}\lambda^{\mathbf{a}+\mathbf{c}}) \tag{15}$$

where contributions from AOs translated by $\mathbf{a}$ and $\mathbf{a} + \mathbf{c}$ are summed. In the special case of a Γ-point calculation, the sum over $\mathbf{c}$ can be carried out first, and the resulting periodic density matrix extracted from the sum. A subsequent sum over $\mathbf{a}$ and $\mathbf{b}$ yields the periodic Γ-point Fock matrix

$$F_{\mu\nu} = \sum_{\sigma,\lambda} P_{\sigma\lambda} \sum_{\mathbf{a},\mathbf{b},\mathbf{c}} (\mu^\mathbf{0}\sigma^\mathbf{a}|\nu^\mathbf{b}\lambda^{\mathbf{a}+\mathbf{c}})$$
$$= \sum_{\sigma,\lambda} P_{\sigma\lambda} \, T_{\mu,\sigma,\nu,\lambda} \tag{16}$$

where the 4-index tensor $T_{\mu,\sigma,\nu,\lambda}$ is obtained by summing ERIs involving AOs located in various periodic images. The number of periodic images necessary depends on the diffuseness of the basis set and the range of the interaction potential $g(|\mathbf{r} - \mathbf{r}'|)$. For a nonzero ERI, $\mu^\mathbf{0}$ and $\sigma^\mathbf{a}$ (respectively $\nu^\mathbf{b}$ and $\lambda^{\mathbf{a}+\mathbf{c}}$) must overlap and the product $\mu^\mathbf{0}\sigma^\mathbf{a}$ must be within the range of the product $\nu^\mathbf{b}\lambda^{\mathbf{a}+\mathbf{c}}$. The HF exchange (HFX) energy is obtained by contracting the density matrix with the Fock matrix

$$E_{\mathrm{HFX}} = -\frac{1}{2} \sum_{\mu,\nu} P_{\mu\nu} F_{\mu\nu} \tag{17}$$

and the nuclear forces are obtained by analytic differentiation of the ERIs with respect to the atomic positions.

The implementation of HFX in CP2K is documented in ref 71. Hybrid DFT and HF calculations require the `&HF` input

subsection of `&DFT%XC`. The `FRACTION` keyword defines the amount of exact exchange that is added to the calculation. Note that it is always necessary to define a XC functional in the `&XC_FUNCTIONAL` subsection. For example, the input for a pure HF calculation looks like

```
&XC
    &XC_FUNCTIONAL NONE
    &END XC_FUNCTIONAL
    &HF
        FRACTION 1.0
        # Additional HF inputs below
    &END HF
&END XC
```

whereas the input for the PBE0 functional reads as

```
&XC
    &XC_FUNCTIONAL
        &GGA_C_PBE
        &END GGA_C_PBE
        &GGA_X_PBE
            SCALE 0.75
        &END GGA_X_PBE
    &END XC_FUNCTIONAL
    &HF
        FRACTION 0.25
        # Additional HF inputs below
    &END HF
&END XC
```

An important input subsection within the `&DFT%XC%HF` section is `&INTERACTION_POTENTIAL`, which controls the parameters of $g(|\mathbf{r} - \mathbf{r}'|)$ in the 2-center 4-electron ERIs of eq 14. The `POTENTIAL_TYPE` keyword selects among the various interaction potentials implemented, while the remaining keywords in this subsection refer to the specifics of the chosen potential. The default value of `POTENTIAL_TYPE` is `COULOMB` for the standard $1/|\mathbf{r} - \mathbf{r}'|$ potential. However, this choice is typically not suitable for periodic calculations, as it introduces spurious exchange interactions between an electron and its periodic images. A possible choice is `TRUNCATED`, which refers to the truncated Coulomb potential

$$g(r) = \begin{cases} 1/r & \text{if } r \leq R_C \\ 0 & \text{if } r > R_C \end{cases} \tag{18}$$

which vanishes beyond a cutoff radius R_C. For a cubic cell of length L, $R_C \leq L/2$ must hold. This is to say that the truncation radius must be shorter than half the smallest cell dimension, but large enough for the exchange energy to converge. Values of 6 Å are typically enough for systems with a large band gap. Systems with smaller band gaps, however, might require larger cutoff values. If the simulation cell is not large enough for both requirements, a supercell must be constructed instead. For ideal performance, the shortest cutoff that converges the energy should be taken. The value of R_C is set with the keyword `CUTOFF_RADIUS` (in Angstroms). Periodic calculations with global hybrids like PBE0 or B3LYP should use this potential.[37,72,73]

An alternative choice is `POTENTIAL_TYPE SHORT-RANGE`, which comes from the decomposition of the Coulomb potential in a short- and a long-range term

$$\frac{1}{r} = \frac{1 - \text{erf}(\omega)}{r} + \frac{\text{erf}(\omega)}{r} \tag{19}$$

The `OMEGA` keyword controls the range-separation parameter. The HSE06 functional, for instance, involves the short-range potential with a value of 0.11 for ω. More involved range-separated hybrid functionals can be configured with the `MIX_CL_TRUNC` interaction potential, which mixes the long-range part of eq 19 and the truncated Coulomb potential. For example, the exact exchange part of the rCAM-B3LYP functional looks like:

```
&INTERACTION_POTENTIAL
    ! Should be <= than L/2 but large
    ! enough for the erf to decay
    CUTOFF_RADIUS 6.0
    OMEGA 0.33
    POTENTIAL_TYPE MIX_CL_TRUNC
    SCALE_COULOMB 0.18352
    SCALE_LONGRANGE 0.94979
&END INTERACTION_POTENTIAL
```

The evaluation of the ERIs $T_{\mu,\sigma,\nu,\lambda}$ in eq 16 can be very demanding and is often the bottleneck of calculations requiring exact exchange. Various screening methods can help reduce the computational cost. Most importantly, Schwarz screening provides an upper bound to any 2-electron 4-center ERI. The `EPS_SCHWARZ` keyword of the `&XC%HF%SCREEN-ING` section sets a threshold, and the evaluation of any ERI estimated to be smaller is skipped a priori. The default value of 10^{-10} is rather safe, and increasing that number can lead to significant performance gains. Yet, a too large value will lead to numerical instabilities and eventually even inaccurate results. The `EPS_SCHWARZ_FORCES` keyword does the same for ERI derivatives and is defined as a function of `EPS_SCHWARZ` by default.

Gaussian type orbital (GTO) basis sets have an exponentially vanishing tail, theoretically extending to infinity. Hence, another important screening threshold is `&DFT%QS%EPS_PGF_ORB`, which sets these tails to zero once a GTO reaches the value of `EPS_PGF_ORB`, thereby effectively defining the range D_μ of an AO

$$\begin{cases} \varphi_\mu(\mathbf{r}) = \varepsilon_{\text{PGF}} & \text{if } |\mathbf{r}| = D_\mu \\ \varphi_\mu(\mathbf{r}) = 0 & \text{if } |\mathbf{r}| > D_\mu \end{cases} \tag{20}$$

This is to say that two AOs, φ_μ centered on atom A at position $\mathbf{R_A}$ and φ_ν centered on atom B at position $\mathbf{R_B}$, only overlap if $|\mathbf{R_A} - \mathbf{R_B}| \leq D_\mu + D_\nu$. This and the choice of interaction potential effectively define the number of periodic cells to consider in the sum of eq 16. Hence, the impact on the computational performance can be significant. The default value of 10^{-5} for `EPS_PGF_ORB` is usually a good compromise between accuracy and speed. A value of 10^{-6} is already considered to be rather safe, whereas an even tighter threshold may negatively impact computational efficiency without necessarily improving accuracy. Since AO overlaps are key to screening, the choice of basis set can also affect performance. In particular, it is not recommended to use diffuse basis sets for HFX calculations, such as those from the previously described MOLOPT family. Instead, the ADMM and/or RI-HFX methods, described below, should be used with diffuse basis sets.

Since the ERIs only depend on atomic positions, it is in principle possible to precalculate and store them all in main memory to simply contract them with the new density matrix at each SCF step. The ERIs only need to be recalculated when

the atoms move, e.g. at each AIMD or geometry optimization step. However, due to the 4-index nature of the $T_{\mu,\sigma,\nu,\lambda}$ tensor, the memory requirements for the storage of the ERIs can become very large. The `MAX_MEMORY` keyword of the `&XC%HF%MEMORY` section caps the memory consumption of the whole HFX code. This value corresponds to the maximum allowed memory usage per MPI rank in MiB. After allocating various buffers and work arrays, the remaining memory is used to store as many ERIs as necessary and possible. To maximize the performance, the value of `MAX_MEMORY` should be as high as possible, while leaving enough space for the rest of the calculation. Usually, allocating 50−75% of the available memory to HFX works well. If there is not enough space to store all ERIs, some of them will be calculated on-the-fly at each SCF step, potentially slowing down the calculation. It is recommended to track the output of a HFX calculation to verify this (look for HFX_MEM_INFO in the output). If the number of spherical ERIs calculated on-the-fly is nonzero, significant gains can be expected by raising the value of `MAX_MEMORY`, or the total amount of memory by using more computing nodes. However, as of the publication of this paper, the HFX code runs only on CPUs. It is recommended to restart HFX calculations from a GGA level converged WF of the same system in order to reduce the number of necessary SCF steps.

2.2.1. Auxiliary Density Matrix Method. In most HFX based calculations, the evaluation of $T_{\mu,\sigma,\nu,\lambda}$ in eq 16 is the main bottleneck. Unfortunately, the cost of evaluating 2-electron 4-center ERIs is tightly connected to the size and diffuseness of the selected basis set. Performing calculations with high quality basis sets can become untractable, especially for periodic systems. The ADMM is designed to tackle this issue. In this approximation, the exact exchange energy is calculated as

$$
\begin{aligned}
E_{HFX} &= E_x^{HF}[\hat{\mathbf{P}}] + (E_x^{HF}[\mathbf{P}] - E_x^{HF}[\hat{\mathbf{P}}]) \\
&\approx E_x^{HF}[\hat{\mathbf{P}}] + (E_x^{GGA}[\mathbf{P}] - E_x^{GGA}[\hat{\mathbf{P}}])
\end{aligned} \tag{21}
$$

where $\mathbf{P}$ is the density matrix in the primary basis and $\hat{\mathbf{P}}$ is the density matrix expressed in an auxiliary basis that is smaller and/or less diffuse than the former basis set. The exact exchange energy is efficiently evaluated in the smaller auxiliary basis, while the difference in exchange energy due to the change of basis set is approximated at the DFT GGA level. The approximation is assumed to be accurate, provided that $\mathbf{P} \approx \hat{\mathbf{P}}$.

Given a choice of auxiliary basis set and GGA exchange functional, there exist different flavors of ADMM. In its simplest form, known as ADMM2, the density is simply projected from the orbital basis onto the auxiliary basis set. On the other hand, ADMM1 involves a purification of the auxiliary density matrix to ensure its idempotency and particle conservation. Both ADMM1 and ADMM2 are described in detail in the original publication by Guidon and co-workers.[50] Subsequent work by Merlot and co-workers introduced additional three flavors that are ADMMP, ADMMQ, and ADMMS, respectively.[74] The latter impose various physically motivated constraints to the approximation in a simpler way than the ADMM1 purification scheme. Note that most post-SCF methods, which can be used in conjunction with ADMM (RPA, MP2, TD-DFT, density-corrected DFT, etc.), only support ADMM2.

To run an ADMM calculation, a valid HFX input file can be modified as follows. The `&AUXILIARY_DENSITY_MA-TRIX_METHOD` subsection should be added to `&DFT`, i.e.

```
&AUXILIARY_DENSITY_MATRIX_METHOD
    # Choose between ADMM flavors
    ADMM_TYPE ADMMS
    # Use PBE exchange (example)
    EXCH_CORRECTION_FUNC PBEX
&END AUXILIARY_DENSITY_MATRIX_METHOD
```

Note that various other keywords exist, but their usage is not recommended. Indeed, `ADMM_TYPE` automatically sets their value for the desired ADMM flavor.

A critical aspect of running ADMM calculations is the choice of auxiliary basis set. The FIT basis family was originally developed to run in conjunction with MOLOPT basis sets. These basis sets can be found in the BASIS_ADMM_MO-LOPT file, and their hierarchy is described in ref 50. More recently, correlation-consistent ADMM basis sets were developed to be used with the ccGRB family, which can be found in the BASIS_ADMM_UZH and BASIS_ccGRB_UZH files, respectively. The usage of the more recent basis sets is recommended. Note that all-electron auxiliary basis sets are also available in the file BASIS_ADMM_ae to be used in conjunction with Jansen's pcseg basis family.[75,76] The ADMM basis set file must be added to the input file, and the specific choice of basis specified for each atomic species, hence:

```
&DFT
    BASIS_SET_FILE_NAME BASIS_ccGRB_UZH
    BASIS_SET_FILE_NAME BASIS_ADMM_UZH
    POTENTIAL_FILE_NAME POTENTIAL_UZH
    # Rest of DFT input
&END DFT
&SUBSYS
    &KIND O
        BASIS_SET ccGRB-D-q6
        BASIS_SET AUX_FIT admm-dz-q6
        POTENTIAL GTH-HYB-q6
    &END KIND
    # Rest of SUBSYS input
&END SUBSYS
```

Note that with the ADMM approximation, a full-fledged HFX calculation is still being carried out, albeit with a smaller base set. All input considerations discussed in the &HF section apply here as well.

2.2.2. Resolution-of-the-Identity Hartree−Fock Exchange. In the resolution-of-the-identity (RI) approximation, the 2-electron 4-center ERIs can be expressed in terms of 2- and 3-center integrals, i.e.

$$
\begin{aligned}
(\lambda\sigma|\mu\nu) &\approx (\lambda\sigma\lfloor P) \, (P\lfloor Q)^{-1}(Q\lfloor R) \, (R\lfloor S)^{-1}(S\lfloor\mu\nu) \\
&= (\lambda\sigma\lfloor P) \, V_{PQ}^{-1}(Q\lfloor R) \, V_{RS}^{-1} \, (S\lfloor\mu\nu)
\end{aligned} \tag{22}
$$

where the Einstein summation convention is used for the sum over the RI basis elements P, Q, R, S. The | and ⌊ symbols represent the HF interaction potential and RI metric, respectively. The RI approximation is exact if the RI basis spans the same space as all possible AO products, and if the RI metric is the same as the interaction potential. This approach allows to shift the computational effort from ERI evaluations to sparse tensor contractions, which can be efficiently performed on GPUs.[77] Note that RI-HFX is not always a suitable choice, and the standard 4-center approach remains relevant in most cases. In practice, RI-HFX shows a performance edge for nonperiodic calculations of molecules with large basis sets (where both the interaction potential and the RI metric are the standard $1/r$ Coulomb measure), and for periodic calculations of small to medium sized dense solids. The CP2K/QUICKSTEP

implementation, as well as a discussion on the choice of RI metric and application cases, can be found in ref 78. The RI-HFX method is compatible with all flavors of ADMM, and analytic gradients to compute the nuclear forces are implemented.

The RI-HFX method can be invoked by adding the following subsection to the `&HF` section of the CP2K input file:

```
&RI
   # Definition of the RI metric:
   RI_METRIC TRUNCATED
   CUTOFF_RADIUS 2.0
   # Optional but potentially
   # important keywords, here
   # with default values
   MEMORY_CUT 3
   EPS_FILTER 1.0E-9
&END RI
```

This input snippet defines the RI metric to be used in a similar fashion to the `&INTERACTION_POTENTIAL` subsection of the same parent `&HF` section. By default, the RI metric takes the same parameters as the latter. The RI metric is usually chosen to be a very short-ranged interaction in order to increase sparsity and to reduce the computational effort. The range of the RI metric is a trade-off between performance and accuracy, with the overlap as the most extreme case, which is known as the RI-SVS approximation.[79] The `MEMORY_CUT` keyword controls batching in the contraction of eq 22: potentially very large tensors are necessary to hold intermediate results, and batching allows processing them with relatively limited resources. Note that this keyword only concerns the previously mentioned contraction step, but it does not affect the total memory footprint of CP2K. Since batched tensor contractions come with an overhead, it is recommended to keep the value of `MEMORY_CUT` as low as the employed resources allow for. It is generally more efficient (but not always practical) to increase the amount of resources, rather than increasing the keyword's value. Note that RI-HFX calculations of large systems can be memory intensive, and that unlike the standard 4-center method, there is no easy way to predict the required memory. The `EPS_FILTER` keyword controls the filtering of small elements out of the sparse tensors. The default value of 10^{-9} is a good compromise between accuracy and efficiency. Increasing this value will lead to performance gains, but it is recommended to proceed with care. The rest of the `&HF` input section can remain the same. Note, however, that the `&MEMORY` and the `&SCREENING` subsections have no effect during a RI-HFX calculation. All considerations on the choice of interaction potential or `EPS_PGF_ORB` discussed for the original 4-center method also apply here.

The choice of RI basis set is important for this type of calculations, both in terms of performance and accuracy. It is generally recommended to use optimized bases when possible (e.g., cc-pVTZ-JKFIT with the cc-pVTZ orbital basis).[80] The cc-pVTZ all-electron basis set, however, is not distributed with CP2K: it would need to be downloaded independently at the EMSL Basis Set Exchange for instance.[44] If not available, the RI basis is generated on-the-fly using the method of Stoychev and co-workers.[81] This type of basis is safe to use and accurate, but is generally larger, and thus less efficient, than its optimized counterparts. The CP2K input file can be modified as follows to specify the RI basis:

```
&DFT
   SORT_BASIS EXP
   # Rest of the DFT input
&END DFT
&SUBSYS
   &KIND
      BASIS_SET cc-pVTZ
      # If the following line is not
      # present, the RI_HFX basis is
      # generated on-the-fly
      BASIS_SET RI_HFX cc-pVTZ-JKFIT
      POTENTIAL ALL
   &END KIND
   # Rest of the SUBSYS input
&END SUBSYS
```

Note the `SORT_BASIS` keyword, use is highly recommended, as it reorders the AO according to their Gaussian exponent, thus increasing sparsity.

2.2.3. Resolution-of-the-Identity Hartree−Fock Exchange with k-Point Brillouin-Zone Sampling. The RI-HFX with **k**-point sampling (RI-HFXk) method allows for periodic HFX calculations of small unit cells with **k**-point sampling. As for **k**-point sampling using DFT (see Section 2.1.5), the **k**-point dependent Fock matrix is first calculated in real-space before it is Fourier transformed, i.e.

$$F_{\mu\nu}^{\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} F_{\mu\nu}^{\mathbf{R}}$$

(23)

where **R** is a unit cell translation vector and $F_{\mu\nu}^{\mathbf{R}}$ is defined in eq 15. Contrary to Γ-point calculations, the sum over periodic images cannot be performed a priori, and the results are stored in a single 4-index tensor. As a result, storing the integrals is intractable, and all 4-center 2-electron ERIs would need to be calculated on-the-fly at each SCF step. Although possible, such an approach would be very expensive. Thus, an efficient alternative using an atom-specific RI scheme was developed.[82]

The RI-HFXk method can be invoked by adding a `&KPOINTS` section to a RI-HFX input file. The **k**-point specific input is the same as described in Section 2.1.5, including the optional band structure information. It is compatible with the ADMM approximation, and nuclear forces are implemented. As for any HFX method, it is generally recommended to restart from a converged GGA level (**k**-point) WF. Extra care must be paid to some parts of the input file, as listed below:

- `&HF%INTERACTION_POTENTIAL`: As for any periodic HF calculation, a short-range potential must be selected. The $L/2$ requirement now refers to the Born−von Kármán supercell, i.e. a supercell obtained by multiplying the unit cell by the number of **k**-points in each direction. As for Γ-point HFX, the potential range should be long enough for convergence, while kept as short as possible for performance.

- `&RI%RI_METRIC`: In the special case of RI-HFXk, the range of the RI metric barely impacts the performance. It is hence recommended to always take it to be the same as the interaction potential for maximal accuracy (default option).

- `&QS%PW_GRID_BLOCKED`: It is recommended to set this keyword to `FALSE`. This ensures that calculations of small unit cells with a lot of CPUs can go ahead without crashing.

- `&SCF%EPS_DIIS`: The default threshold value to start DIIS is 0.1. Experience has shown that better

convergence is achieved when this value is lower, typically around 0.05.

- `&RI%KP_NGROUPS`: There is a lot of computational work to be done, and most of it can be efficiently parallelized over MPI subgroups. This is controlled with the `KP_NGROUPS` keyword, which should be a divisor of the total number of ranks. Using N subgroups should accelerate the calculations by a factor $\sim N$ and increase its memory usage by the same factor.

- `&RI%KP_STACK_SIZE`: Most of the computational effort is spent on contracting tensors. In some cases, more efficient batched contractions can take place, at the cost of greater memory usage. The default value of 32 is meant for small systems of a few atoms. Larger systems will require more memory overall and might not have enough room for the storage of temporary tensors (in which case, use a smaller stack size).

Generally, the RI-HFXk method is very demanding, both in terms of compute and memory. It was designed and optimized for small to medium size systems with dense **k**-point meshes (e.g., 2D materials, highly symmetric crystals, etc.). The computational cost of the method scales as $O(N_{atoms}^3 N_{img}^2)$, where N_{img} refers to the number of periodic images required (depends on the diffuseness of the base set and the range of the interaction potential). The scaling of the memory usage is harder to predict, but experience suggests that it is steep, too. For large systems, a Γ-point calculation with an equivalent supercell will be much more efficient due to the asymptotical linear scaling of the 4-center HFX implementation. The details of the RI-HFXk method are not trivial, and it is recommended to read ref 82 before use.

2.3. Post-Hartree−Fock and Double-Hybrid Density Functional Theory

In addition to incorporating exact exchange via HFX, as described in Section 2.2, electron correlation beyond conventional DFT can be added via efficient MP2 and RPA methods described in this section. Most of them are available for condensed phase systems including nuclear forces, if not explicitly mentioned otherwise.

2.3.1. Second-Order Møller−Plesset Perturbation Theory.
The MP2 method computes

$$E^{(2)} = -\sum_{ijab} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{24}$$

employing canonical HF reference orbitals, each with the real-space representation $\phi_p(\vec{r})$ and an orbital energy ϵ_p. The ERIs are usually calculated in the AO basis first and then in four steps transformed to the MO basis. CP2K/QUICKSTEP provides three different implementations: a direct/canonical implementation, a GPW-based implementation[83] and a RI-based implementation.[84] All MP2 implementations scale with the fifth power of the number of atoms. The canonical implementation is available for GPW and GAPW reference calculations allowing core-corrections. However, it is the most costly implementation and analytical gradients (nuclear forces, stress tensors, etc.) are not available.

All flavors of MP2 calculations are available via the `&MP2` subsection in the `&DFT%XC%WF_CORRELATION` section and activating the respective method. For direct canonical MP2, it is `METHOD DIRECT_CANONICAL`. Further key-

words, which are also of interest for all other implementations, are

- `&WF_CORRELATION%MEMORY`: the allowed memory for the MP2 calculation.
- `&WF_CORRELATION%GROUP_SIZE`: it needs to be a divisor of the total number of processes. Lower values reduce communication costs, but increase memory requirements.
- `&WF_CORRELATION%INTEGRALS`: optional section to configure the ERI method via the `ERI_METHOD` keyword and by the `INTERACTION_POTENTIAL` subsection the corresponding interaction potential (default: Coulomb potential).

The direct MP2 implementation serves as a reference method for medium-sized systems, as its computational demands become easily untractable for larger systems. This implementation also does not benefit from GPU acceleration.

For better performance, the GPW-based implementation should be used.[83] It skips the first two transformation steps by calculating half-transformed ERIs from the product density $\phi_i(\vec{r})\phi_a(\vec{r})$ directly. It is available by setting `&MP2%METHOD MP2_GPW`. The GPW-integration is configured with the `&WFC_GPW` subsection inside the `&WF_CORRELATION%INTEGRALS` section. As with the original GPW method, there are two tunable parameters `CUTOFF` and `REL_CUT-OFF` with the same meaning and should be tuned as in the original DFT-based GPW approach described in Section 2.1.1. Yet, due to the different requirements on the accuracy and the properties of the densities, primary cutoffs of 150 to 300 Ry are typically sufficient for most applications. Using the grid and the distributed block compressed sparse row (DBCSR) libraries, the GPW-based implementation runs efficiently on GPUs, enabling MP2 calculations on systems containing hundreds of atoms.

The RI approximation to the ERIs allows for further computational savings.[85] If we expand products of orbitals in a set of auxiliary functions denoted by P, Q, ..., the ERI tensor factorizes approximately into a rank-3 tensor B, i.e.

$$(ia|jb) \approx \sum_P B_{iaP}B_{jbP} \tag{25}$$

where B is calculated from the respective three-center integrals $(P|Q)$ and $(P|ia)$. These can be calculated using the GPW-approach in a similar fashion as discussed above, or alternatively using a Minimax-Ewald (MME) approach or the Obara−Saika (OS) scheme, which can be set by the previously mentioned `&INTEGRALS%ERI_METHODS` keywords. For that, the `RI_MP2` section replaces the `MP2` section.

```
&RI_MP2
   # Larger block sizes require more
   # memory but reduce communication,
   # (default: -1, auto determination)
   BLOCK_SIZE 2
   # Large number of groups increase the
   # memory demands but reduce communication
   # (default: -1, auto determination)
   NUMBER_INTEGRATION_GROUPS 2
&END
```

The most important keywords are `BLOCK_SIZE` and `NUMBER_INTEGRATION_GROUPS`. By default, these parameters are determined automatically from the available memory, but their tuning may be of interest for calculations on a larger number of systems. Large block sizes and large

numbers of integration groups reduce communication costs at the expense of a higher memory demand. The number of integration groups must be a divisor of the number of subgroups (number of MPI ranks divided by the `GROUP_-SIZE`). See ref 84 for further information on the technical details and GPU acceleration.

As before, the RI approximation requires the user to specify an auxiliary basis set. In general, the auxiliary basis needs to be adapted to the primary basis set to yield an acceptable accuracy and sufficient speed-up compared to the canonical implementations. CP2K has several options: preoptimized, self-optimized, and auto-optimized basis sets. Optimization of the auxiliary basis sets is available according to ref 86. A compilation of preoptimized basis sets is available in the previously mentioned data directory of the CP2K GitHub repository. If none is available for a certain basis set or the optimization is not required, we recommend auto-optimized basis sets. They are available by omitting the specification of the auxiliary basis set in the `&SUBSYS%KIND` section and setting the `AUTO_BASIS` keyword in the `&DFT` section.

For the RI-MP2 implementation, analytical gradients and stress tensors are available.[87,88] Its implementation is based on the Lagrangian formalism of the MP2 energy by introducing the `&WF_CORRELATION%CANONICAL_GRADIENTS` section. The Lagrangian formalism requires the solution of the coupled-perturbed HF (CPHF) equations configured in the `&CANONICAL_GRADIENTS%CPHF` subsection:

```
&CANONICAL_GRADIENTS
    # The default is usually good enough
    EPS_CANONICAL 1E-6
    # This option which should always work
    # Try to set it to .TRUE.
    FREE_HFX_BUFFER .FALSE.
    &CPHF
        # Choose as tight as necessary
        EPS_CONV 1.0E-6
        # Smaller values of EPS_CONV
        # may require more iterations
        MAX_ITER 10
    &END CPHF
&END CANONICAL_GRADIENTS
```

Therein, `EPS_CANONICAL` is a MP2 specific parameter affecting the numerical stability of MP2 analytic gradient calculations. Higher values improve the accuracy, but increase the computational costs significantly. The `FREE_-HFX_BUFFER` keyword frees the HF buffers before the gradient calculations to increase the amount of available memory necessary for it. By default, it is turned off because the integrals cannot always be recovered for technical reasons. The parameters of the `&CPHF` section affect the numerical solution of the CPHF equations. The `EPS_CONV` parameter affects the accuracy of the solution of the CPHF equations, which are to be chosen as tightly as necessary and as high as possible. The `MAX_ITER` parameter determines the maximum number of iteration steps. If convergence is not observed, then the orbitals are not sufficiently converged, or the MP2 method is not suitable for the given system because of too high electron correlation.

The accuracy of MP2 calculations is mostly affected by the size of the primary basis set. In contrast to HF or DFT calculations, where the basis set error decays exponentially with the number of basis functions, the error of the MP2 energy decays with the reciprocal number of basis functions.[89] Thus, either sufficiently large basis sets or an extrapolation to the complete basis set limit is required. Independently of an extrapolation scheme, basis sets should have a quality of at least triple-ζ or augmented double-ζ for accurate results. Finally, the basis sets should contain enough polarization functions as in correlation-consistent basis sets.

The performance of MP2 calculations in CP2K depends on several factors. Conceptually, the number of operations increases linearly to quadratically with the number of auxiliary basis functions, suggesting that optimized auxiliary basis sets should be employed whenever possible. The calculation of the integrals using the GPW approach relies on the efficient grid library and sparse matrix-contractions with DBCSR.[90] The actual contraction of the ERIs to the MP2 energy relies on a well-optimized BLAS implementation for the matrix−matrix multiplications. For GPU acceleration, CP2K is either linked to an accelerated BLAS implementation or switches to the SpLA library.[91]

Although the calculation of the MP2 energy shows a steeper scaling than a HF calculation, the HF kernel can still dominate the computational costs in case of diffuse basis functions. This can be overcome by the ADMM,[50] which is available for MP2 energy and gradient calculations.[92]

Double-hybrid functionals improve upon MP2 by improving the description of short-ranged correlation effects using DFT.[93] Instead of starting from a HF description, double-hybrid functionals start from hybrid DFT and then add a certain amount of spin-rescaled MP2 energy. In CP2K/QUICKSTEP, this is achieved with the following modifications inside the `&XC` section:

- `&XC_FUNCTIONAL`: set up the respective DFT part. In addition to the natively implemented XC functionals, an even larger selection of functionals is available through the LibXC library.[94]
- `&HF`: set up the respective HF interaction operator. For simple functionals, the adjustment of the `FRACTION` parameter is sufficient. For more elaborate functionals consider the parameters in the `&INTERACTION_PO-TENTIAL` subsection.
- `&WF_CORRELATION`: the keywords `SCALE_S` and `SCALE_T` determine the weight of singlet and triplet contributions. The `&INTEGRALS%INTERAC-TION_POTENTIAL` subsection allows the adjustment of the operator.

Beware that CP2K does not read any HF related information from LibXC such that all parameters have to be looked up and set up manually. We recommend testing the setup against reference values for the respective double-hybrid functional as the setup is error-prone. Some examples can be found at https://github.com/cp2k/cp2k/tree/master/data/xc_section. Nuclear forces and stress tensors are analytically available throughout for double-hybrid functionals.[92,95]

2.3.2. Random Phase Approximation and Laplace-Transformed Scaled Opposite-Spin Second-Order Møller−Plesset Perturbation Theory. The RI approach to the direct RPA (dRPA) method given by its correlation energy contribution

$$E^{dRPA} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} d\omega \langle \log(1 + Q(\omega)) - Q(\omega) \rangle$$

$$Q_{RS}(\omega) = 2 \sum_{ia} B_{iaR} \frac{\epsilon_a - \epsilon_i}{(\epsilon_a - \epsilon_i)^2 + \omega^2} B_{iaS}$$

(26)

and the RI-based Laplace-transformed scaled opposite-spin MP2 (SOS-MP2)

$$E^{\text{SOS-MP2}} = -\int_0^\infty d\tau \, \text{Tr}(\overline{Q}(\tau)^2)$$

$$\overline{Q}_{RS}(\tau) = \sum_{ia} B_{iaR} e^{-(\epsilon_a - \epsilon_i)\tau} B_{iaS} \tag{27}$$

are implemented similarly due to the integration of a matrix-valued function and the contraction of the RI-tensor **B** to a frequency-dependent or time-dependent intermediate matrix $\mathbf{Q}(\omega)$. Their energy-only implementation is discussed in ref 84. Although the integration may be performed analytically via the Plasmon equation,[96] or the MP2 energy equation (see eq 24), we pursue the path of numerical integration due to its reduced scaling behavior of fourth power with respect to the number of atoms, as compared to the alternative with fifth order scaling.

CP2K/QUICKSTEP implements two quadrature schemes: Clenshaw−Curtis[97] and Minimax.[98] The Clenshaw−Curtis quadrature is easily accessible and converges for all subsequent methods, but requires a larger number of quadrature points. In contrast to that, the Minimax quadrature rules are difficult to optimize and are thus pretabulated. The main advantage of Minimax quadrature rules lies in their higher accuracy than the Clenshaw−Curtis quadrature rules and are the recommended choice for RPA calculations and the only available option for SOS-MP2 calculations. The setup of these methods is similar to the setup of an (RI-)MP2 calculation. The main difference is that the `&RI_RPA` section

```
&RI_RPA
    # Switch to the Minimax quadrature
    # rules (default: turned off, use
    # Clenshaw-Curtis-quadrature)
    MINIMAX .TRUE.
    # Choose it as large as necessary
    # and as small as possible
    QUADRATURE_POINTS 6
    # Fine-tuning of the parallelization
    NUM_INTEG_GROUPS -1
&END
```

or the `RI_SOS_MP2` section

```
&RI_SOS_MP2
    # Both parameters work similarly to RPA
    QUADRATURE_POINTS 6
    NUM_INTEG_GROUPS -1
&END
```

replaces the `&MP2` or `&RI_MP2`-sections, respectively. Note that the Minimax quadrature has to be turned on manually for RPA calculations. This difference is related to RPA calculations being the precursor of *GW* bandstructure calculations, which are described in Section 3.2, requiring quadratures suitable for different kinds of integration kernels.

The RPA correlation energies are usually combined with exact exchange energies. In CP2K, the `&RI_RPA` section includes a `&HF` section, which is set up similarly to the ordinary `&HF` section used in HF and hybrid functional calculations. Because the ERIs are needed only once here, the `MAX_MEMORY` keyword in the `&MEMORY` should be set to zero in case of a simple RPA energy calculation to save memory. The ADMM acceleration can also be applied for the calculation of exact exchange energies by activating the `ADMM` keyword in the `&RI_RPA` section. We recommend the ADMM approximation for calculations with diffuse basis

functions, which would otherwise increase the computation time tremendously due to their larger extent.

On top of the exact exchange and the dRPA correlation energy, one can also apply a beyond-RPA scheme. The `RSE` keyword activates the renormalized single excitations (RSE) correction to account for nondiagonal Fock-matrix elements.[99] In addition to that, the `&RI_RPA%EXCHANGE_CORRECTION` section turns on either the approximate exchange kernel (AXK),[100] or the second-order screened exchange (SOSEX) corrections to reduce the self-correlation errors of the dRPA method.[101] The corrections are set up as shown below.

```
&RI_RPA
    ...
    # The RSE correction is only relevant
    # for non-HF reference states
    RSE .TRUE.
    # Exchange corrections can be costly
    &EXCHANGE_CORRECTION [NONE|AXK|SOSEX]
        # The Hartree-Fock-based implemen-
        # tation scales better for larger
        # systems but introduces more noise
        USE_HFX_IMPLEMENTATION F
        # This parameter is ignored if
        # USE_HFX_IMPLEMENTATION is set to T
        # Larger values improve performance
        # but increase memory usage
        BLOCK_SIZE 16
    &END EXCHANGE_CORRECTION
    # This HF section activates the exact
    # exchange correction and is also used
    # by the RSE and the AXK/SOSEX methods
    &HF
        ...
        &MEMORY
            # Set to zero if RSE/AXK/SOSEX
            # are not requested
            MAX_MEMORY 500
        &END MEMORY
    &END HF
&END
```

As is customary in CP2K, it is important to use the lone section name as a keyword to specify the actual method. The `USE_HFX_IMPLEMENTATION` keyword in the `&RI_RPA%EXCHANGE_CORRECTION` section switches between a better scaling HF-based implementation and an RI-MP2-based implementation. The first one should be preferred for larger systems, and the latter one should be preferred for smaller systems. Note that the HF-based implementation recalculates the integrals using the HF code. In the periodic case, this requires a truncated Coulomb potential, whereas the integrals for the RI methods can be calculated using the ordinary Coulomb potential. In case of the SOSEX-correction, this makes the implementation not self-correlation-free although the SOSEX method theoretically shows this property.

For the dRPA and SOS-MP2 implementations, CP2K implements analytical nuclear forces and stress tensors.[102] The general setup is similar to RI-MP2 gradient calculations, but adds the `MAX_PARALLEL_COMM` keyword to the `&WF_CORRELATION%CANONICAL_GRADIENTS` section for RI-RPA gradient calculations. It determines the number of parallel communication channels for the nonblocking communication scheme. Larger numbers improve the performance, but increase the memory requirements. We find that a

maximum value of 3 is sufficient. Note that analytical gradients are not available for the beyond-RPA methods.

The computationally most costly part of dRPA and SOS-MP2 calculations is the contraction of **B** to $\mathbf{Q}(\omega)$. Because this step is necessary for each quadrature point, the number of quadrature points should be chosen as low as possible. Therefore, we apply the Minimax quadrature rule. It is also a convenient option for the beyond-RPA schemes. Because CP2K relies heavily on parallel matrix−matrix multiplications using ScaLAPACK, a well-optimized (and GPU accelerated) implementation of this library should be employed if possible. However, please consult Section 11 for more details on the support of various libraries within CP2K. Alternatively, CP2K can employ the communication-optimal COSMA library for even higher performance in exchange for increased memory usage.[103] If CP2K was compiled with COSMA support, it is used automatically. All of this holds for all kinds of RPA- or SOS-MP2-based methods.

As with RI-MP2, both RI-RPA and SOS-MP2 rely on partial replication of the data via a `NUM_INTEG_GROUPS` keyword inside the corresponding sections. It is by default chosen automatically based on the available memory, but can be tuned manually. It must be a divisor of the number of MPI ranks and the number of quadrature points. A larger number of integration groups reduces the communication, but increases the memory demands. This feature is also available for analytic gradient calculations.

2.3.3. Low-Scaling Post-Hartree−Fock. As already mentioned, post-HF methods suffer from the steep scaling of their computational cost, thus limiting their application to small and medium-size systems. Reworking the fundamental equations, using the RI approximation and exploiting the inherent sparsity of AOs, it is possible to reduce that scaling. In CP2K, low-scaling implementations of RI-dRPA and RI-SOS-MP2 are available, for both energies[104] and nuclear forces.[78] It is straightforward to transform a canonical RI-dRPA or RI-SOS-MP2 input file into a low-scaling one, simply by adding the following to the `&WF_CORRELATION` section:

```
&WF_CORRELATION
  # Usual RI-RPA/RI_SOS_MP2 input here
  &LOW_SCALING
    MEMORY_CUT 3
    EPS_FILTER 1.0E-9
  &END LOW_SCALING
  &RI
    &RI_METRIC
      POTENTIAL_TYPE TRUNCATED
      CUTOFF_RADIUS 1.5
    &END RI_METRIC
  &END RI
&END WF_CORRELATION
```

The input parameters specified in `&RI_RPA` or `&RI_-SOS_MP2` also apply in the low-scaling case. Note that the two subsections used in the above example are already discussed in Section 2.2.2 on the RI-HFX, since both methods used the same machinery and a similar theory. All performance aspects, such as the use of `SORT_BASIS EXP` or the choice of `EPS_PGF_ORB` also apply here. The `&INTEGRALS` subsection of a canonical post-HF calculation is also used in the low-scaling case, but only for 2-center integrals. The rest is calculated analytically with the Libint library.[105]

The nominal scaling of these methods is cubic. The exploitation of sparsity allows for a subcubic scaling in most cases. Note that energy-only calculations typically have a more advantageous scaling than nuclear force calculations, which still benefit from a cubic scaling upper bound. A lot of overhead is necessary to run such calculations, making them more expensive than their quartic scaling counterparts for small systems. The choice of the RI metric also adds an additional level of approximation. It is therefore always recommended to use the canonical versions of dRPA or SOS-MP2 whenever possible, and only use the low-scaling implementations for very large systems with hundreds of atoms. If in doubt, it might be worth simply trying both approaches.

2.4. Density Functional Theory plus Hubbard U

In spite of its success, standard local and semilocal DFT shows shortcomings in describing some physical phenomena such as London dispersion interactions or strong electron−electron correlation. The latter phenomenon is dominant in transition metals (TM), lanthanide and actinide oxides, such as NiO, CoO, CeO_2 or UO_2, resulting in a poor description of the electronic structure of these materials only using standard DFT. Anisimov et al. proposed to add a Hubbard U term as a correction to the standard DFT to improve the description of these materials.[106] Since then, many variants of this kind of correction have been proposed. One of the most popular corrections, called DFT + U, was introduced by Dudarev et al.[107,108] It is rotationally invariant, computationally inexpensive, and introduces only an effective Hubbard parameter $U_{eff} = U − J$ as an atomic parameter for the atomic sites concerned. The Dudarev approach is implemented in CP2K/QUICKSTEP.[58] It is based on orbital occupations derived from a Mulliken,[109] or Löwdin population analysis.[110] DFT + U is globally activated with the keyword `PLUS_U_METHOD` in the `&FORCE_EVAL%DFT` section:

```
&FORCE_EVAL
  METHOD Quickstep
  &DFT
    PLUS_U_METHOD Mulliken
    ...
    &PRINT
      &PLUS_U on
      &END PLUS_U
    &END PRINT
  &END DFT
  ...
&END FORCE_EVAL
```

and the DFT + U specific printout is controlled via the corresponding `&PRINT` section within `&PLUS_U`. The U_{eff} parameter `U_MINUS_J` and the angular momentum number `L` must be defined in an atomic `&KIND` section. Here is an example that applies an effective Hubbard U_{eff} of 2 eV to the 5f orbitals ($l = 3$) of the uranium atoms assigned to the atomic kind U_a:

```
&KIND U_a
    ...
    &DFT_PLUS_U on
      L 3
      U_MINUS_J [eV] 2.0
      &ENFORCE_OCCUPATION on
        EPS_SCF 1.0E-3
        MAX_SCF 20
        ORBITALS -3 -1
        SMEAR on
      &END ENFORCE_OCCUPATION
    &END DFT_PLUS_U
&END KIND
```

It is possible to enforce certain initial (desired) orbital occupations for a number of `MAX_SCF` SCF iteration steps or until a certain SCF convergence threshold `EPS_SCF` is reached as in the example above for the 5f orbitals of the atomic kind U_a in which the actual 5f electron density is smeared equally among the $5f_{-3}$ and $5f_{-1}$ orbitals. In this way, the desired orbital occupation patterns can be "enforced" when a straightforward convergence to the electronic ground state does not take place or is not expected due to the presence of metastable states,[57,58] in which case an initial occupation matrix control (OMC) is required. After an initial SCF run (via `RUN_TYPE energy`) has been converged with `&ENFORCE_OCCUPATION on`, a SCF run without any OMC constraint (i.e., `&ENFORCE_OCCUPATION off`) can be restarted using the WF restart file of the initial constraint SCF run.

The "enforcement" is performed based on the initial orbital occupation of an atomic kind, which can be changed via the `&KIND%BS` section

```
&KIND U_a
    ...
    &BS
      # U(+4): 5f3 6d1 7s2 -> 5f2 6d0 7s0
      &ALPHA
        N    5  6  7 # 5       6       7
        L    3  2  0 # f       d       s
        NEL +1 -1 -2 # (3+1)/2 (1-1)/2 (2-2)/2
        # U(alpha)  : 2       0       0
      &END ALPHA
      &BETA
        N    5  6  7 # 5       6       7
        L    3  2  0 # f       d       s
        NEL -3 -1 -2 # (3-3)/2 (1-1)/2 (2-2)/2
        # U(beta)   : 0       0       0
      &END BETA
    &END BS
&END KIND
```

to create a spin-up on-site $5f^2$ triplet for a U^{4+} cation. For the corresponding spin-down $5f^2$ cation, the definitions in the sections `&ALPHA` and `&BETA` for spin-up and spin-down electrons have to be exchanged. The `MULTIPLICITY` in the `&DFT` section must be adapted if the number of spin-up and spin-down U atoms is not equal, which is needed for ferri- and ferromagnetic materials. The corresponding initial orbital occupation for the O^{2-} anion in UO_2 can be specified with

```
&KIND O
    &BS
      # O(-2): 2s2 2p4 -> 2s2 2p6
      &ALPHA
        N    2    # 2
        L    1    # p
        NEL +2    # (4+2)/2
      # O(alpha): 3
      &END ALPHA
      &BETA
        N    2    # 2
        L    1    # p
        NEL +2    # (4+2)/2
      # O(beta) : 3
      &END BETA
    &END BS
```

Atomic forces are only implemented for the `PLUS_U_-METHOD Mulliken` and **k**-points are not available with DFT + U, yet.

3. ELECTRONIC BAND STRUCTURE

Beside the CP2K/QUICKSTEP code, more accurate electronic band structure calculations can be performed using the domain-specific electronic structure library SIRIUS, which implements the FP-LAPW and PP-PW methods,[18,20] or the GW method.[21]

3.1. SIRIUS

With the interface to SIRIUS, CP2K has the ability to run FP-LAPW and PP ground state calculations, similar to Wien2k,[111] FLEUR,[112] ELK[113] and exciting,[114] to mention just a few.

SIRIUS supports:

- norm-conserving,[115] ultrasoft,[116] as well as PAW PPs

- APW[117] and LAPW[20] basis sets with an arbitrary number of local orbitals composed of up to third-order energy derivatives of radial functions

- evaluation of stress tensor and nuclear forces

- collinear and noncollinear magnetism

- symmetrization of lattice-periodic functions and on-site matrices

- generation of irreducible **k**-point meshes

- local-density approximation (LDA)[118] and GGA flavours of XC potentials via the LibXC library[94]

Let us start with the example of PP-PW calculation setup using the second version of unified PP format (UPF) files, which are stored in XML format and can be read by SIRIUS directly. The example below demonstrates the input file for diamond:

```
&FORCE_EVAL
   METHOD SIRIUS
   &PW_DFT
      &MIXER
         BETA 0.5
         USE_HARTREE .TRUE.
      &END MIXER
      &PARAMETERS
      ELECTRONIC_STRUCTURE_METHOD \
         PSEUDOPOTENTIAL
      GK_CUTOFF 5
      PW_CUTOFF 20
      NGRIDK 4 4 4
      SMEARING GAUSSIAN
      SMEARING_WIDTH 0.0004
      &END PARAMETERS
   &END PW_DFT
   &SUBSYS
      &KIND C
         POTENTIAL UPF \
            C.pbe-n-kjpaw_psl.1.0.0.UPF
      &END KIND
   &END SUBSYS
&END FORCE_EVAL
```

With a small change in parameters and atomic kind input sections, it is straightforward to switch to the FP-LAPW mode and compute the same structure with a high-accuracy full-potential method:

```
&FORCE_EVAL
   METHOD SIRIUS
   &PW_DFT
      &PARAMETERS
         ELECTRONIC_STRUCTURE_METHOD \
            FULL_POTENTIAL_LAPWLO
         DENSITY_TOL 1e-06
         ENERGY_TOL 1e-10
         GK_CUTOFF 4.0
         NGRIDK 4 4 4
         PW_CUTOFF 20
         SMEARING GAUSSIAN
         SMEARING_WIDTH 0.0004
         AUTO_RMT 1
         CORE_RELATIVITY DIRAC
         VALENCE_RELATIVITY IORA
         LMAX_APW 12
         LMAX_POT 12
         LMAX_RHO 12
      &END PARAMETERS
   &END PW_DFT
   &SUBSYS
      &KIND C
         POTENTIAL UPF C.json
      &END KIND
   &END SUBSYS
&END FORCE_EVAL
```

Full-potential atomic species JSON files for SIRIUS are available at https://github.com/electronic-structure/species/tree/master/FP/v1. They are similar to UPF files and contain information about the radial grid and LAPW+lo basis set. In the following, we provide an overview of the most relevant variables in the `&PW_DFT%PARAMETERS` section:

- ELECTRONIC_STRUCTURE_METHOD: type of calculation, i.e. FULL_POTENTIAL_LAPWLO, or PSEUDOPOTENTIAL.
- CORE_RELATIVITY: specifies if the core states in FP-LAPW are treated relativistically via the DIRAC equation or not.

- VALENCE_RELATIVITY: type of relativistic treatment of valence FP-LAPW radial basis functions and local orbitals. Possible values are NONE, KOELLING_-HARMON, ZORA, or IORA.
- NUM_BANDS: number of "bands" (KS states) to compute during Hamiltonian diagonalization.
- SMEARING: type of smearing used to compute the Fermi level, such as GAUSSIAN, COLD, FERMI_-DIRAC, GAUSSIAN_SPLINE and METHFESEL_-PAXTON.
- SMEARING_WIDTH: Width (in Ha) of the smearing function.
- PW_CUTOFF: PW cutoff (in a.u.$^{-1}$) for expansion of electron density and potential.
- GK_CUTOFF: PW cutoff (in a.u.$^{-1}$) for $|\mathbf{G} + \mathbf{k}|$.
- NUM_MAG_DIMS: number of magnetic dimensions in the system, i.e. 0—nonmagnetic calculations, 1—spin-collinear case, 3—noncollinear magnetic calculation.
- LMAX_APW: maximum orbital quantum number for LAPW basis functions.
- LMAX_RHO: maximum orbital quantum number for charge density expansion in muffin-tins (LAPW only).
- LMAX_POT: maximum orbital quantum number for KS potential expansion in muffin-tins (LAPW only).
- NGRIDK: dimensions of the $\mathbf{k}$-point mesh. In other words, this is a division of the first Brillouin zone into microcells for $\mathbf{k}$-point integration.
- ENERGY_TOL: SCF convergence tolerance of total energy.
- DENSITY_TOL: SCF convergence tolerance of charge density.
- GAMMA_POINT: Specifies if this is a Γ-point calculation (PP case only). NGRIDK must be set to 1 1 1 in that case.

Similarly, in section `&PW_DFT%MIXER`:

- TYPE: denotes the type of the density mixer, i.e. LINEAR, ANDERSON, ANDERSON_STABLE, or BROYDEN2.
- BETA: determines mixing parameter, which is a real value in the interval $[0, 1]$.
- USE_HARTREE: logical variable that controls what is used as estimation of charge density difference. TRUE: Hartree energy of density residual $\frac{1}{2} \iint \frac{\Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$ is used as a measure of density convergence (works only for PP calculation). FALSE: normalized inner product of density residuals $\frac{1}{\Omega} \int \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}) d\mathbf{r}$ is used as a measure of density convergence.

Computational aspects of CP2K/SIRIUS are configured in section `&PW_DFT%CONTROL`:

- MPI_GRID_DIMS: A 2-dimensional array that defines how the band parallelization is performed. The first dimension of the grid specifies the size of the FFT communicator used in the transformation of individual WFs. The second dimension defines the number of independent FFT communicator groups used to parallelize FFTs across different WFs. The product of the two dimensions defines the total number of MPI ranks used for the band parallelization. An orthogonal communicator will be built from the total available number of MPI ranks to perform $\mathbf{k}$-point parallelization. Additionally, if the MPI grid is square, for example {3,

3}, a parallel eigensolver will be used to diagonalize the subspace Hamiltonian matrix.

3.2. The *GW* Method

The *GW* method is particularly well suited for high-accuracy computations of the electronic band structure of solids and MO energies of molecules.[119] We first describe the band gap problem of DFT to motivate the usage of *GW*, before briefly discussing the theoretical framework of *GW* and our *GW* implementation within CP2K. We also provide input sections of *GW* calculations performed with CP2K.

3.2.1. The Band Gap Problem of Density Functional Theory.
When considering a solid under PBCs, the KS equations read as

$$\left(-\frac{\nabla^2}{2m} + v_H(\mathbf{r}) + v_{XC}(\mathbf{r}) \right) \psi_{n\mathbf{k}}(\mathbf{r}) = \varepsilon_{n\mathbf{k}}^{DFT} \psi_{n\mathbf{k}}(\mathbf{r}) \tag{28}$$

where $v_H(\mathbf{r})$ is the Hartree potential and $v_{XC}(\mathbf{r})$ is the XC potential. These KS equations are solved to obtain the KS orbitals $\psi_{n\mathbf{k}}(\mathbf{r})$ and the KS eigenvalues $\varepsilon_{n\mathbf{k}}^{DFT}$ with band index n and crystal momentum $\mathbf{k}$. For a molecule, the KS orbitals $\psi_n(\mathbf{r})$ and the KS eigenvalues ε_n^{DFT} only carry a single quantum number, the MO index n. These KS eigenvalues $\varepsilon_{n\mathbf{k}}^{DFT}$ are often used to approximate the electronic band structure of a solid. But, this approximation comes with limitations.

First, when using one of the common GGA XC functionals, the band gap in the KS-DFT band structure $\varepsilon_{n\mathbf{k}}^{DFT}$ is much too small compared to experimental band gaps.[119] Even with the exact XC functional, KS-DFT underestimates the fundamental gap due to the derivative discontinuity.[120]

Second, the GGA band structure $\varepsilon_{n\mathbf{k}}^{DFT}$ is insensitive to screening by the environment. As an example, the GGA eigenvalue gap $\varepsilon_{LUMO}^{DFT} - \varepsilon_{HOMO}^{DFT}$ between the lowest unoccupied MO (LUMO) and the highest occupied MO (HOMO) is almost identical for a molecule in the gas phase and a molecule on a surface. In experiment, however, the surface induces an image-charge effect, which can reduce this HOMO − LUMO gap of the molecule by several eV compared to the gas phase. In more general terms, this is a nonlocal screening effect by the surface that is absent in common approximate GGA XC functionals. A similar band gap reduction due to nonlocal screening is present when two materials are brought close to each other, for example, when two sheets of atomically thin materials are stacked on top of each other.

Third, one might use hybrid XC functionals to obtain band gaps from KS eigenvalues that align more closely with experimental values than GGA band gaps.[121−124] However, the issue remains that nonlocal screening effects by the environment are not included in hybrid functionals. The above issues are known as the band gap problem of DFT.

3.2.2. Theory of *GW* Band Structure Calculations.
Green's function theory offers a framework for calculating electron removal and addition energies, known as quasiparticle energies. Hedin's equations provide an exact method for computing these quasiparticle energies within Green's function theory.[51] The *GW* approximation simplifies Hedin's equations by approximating the self-energy Σ as the product of the Green's function G and the screened Coulomb interaction W, i.e.

$$\Sigma^{GW}(\mathbf{r}_1, \mathbf{r}_2, t) = iG(\mathbf{r}_1, \mathbf{r}_2, t) W(\mathbf{r}_1, \mathbf{r}_2, t) \tag{29}$$

The advantage of *GW* is that it captures nonlocal screening effects on the electronic band structure, as previously discussed, and that band gaps computed by *GW* are often in excellent agreement with experiments.

GW calculations in CP2K start from a KS-DFT calculation, i.e. we assume that the above KS equations have been solved. In the G_0W_0 approach, we use KS orbitals and their KS eigenvalues to compute the Green's function G_0 of non-interacting electrons and the screened Coulomb interaction W_0 in the RPA. The G_0W_0 self-energy $\Sigma^{G_0W_0}(t)$ is then obtained by replacing $G \rightarrow G_0$ and $W \rightarrow W_0$ in eq 29, followed by a Fourier transform from time t to frequency (or energy) ε, yielding $\Sigma^{G_0W_0}(\varepsilon)$. In the G_0W_0 method we further approximate that KS orbitals are the quasiparticle WFs. Then, the G_0W_0 quasiparticle energies are

$$\varepsilon_{n\mathbf{k}}^{G_0W_0} = \varepsilon_{n\mathbf{k}}^{DFT} + \left\langle \psi_{n\mathbf{k}} \left| \sum^{G_0W_0} (\varepsilon_{n\mathbf{k}}^{G_0W_0}) - v_{XC} \right| \psi_{n\mathbf{k}} \right\rangle \tag{30}$$

We might interpret this as removing the spurious XC contribution of DFT, i.e. $\langle \psi_{n\mathbf{k}} | v_{XC} | \psi_{n\mathbf{k}} \rangle$, from the DFT eigenvalue $\varepsilon_{n\mathbf{k}}^{DFT}$ and adding the XC contribution from G_0W_0, i.e. $\langle \psi_{n\mathbf{k}} | \Sigma^{G_0W_0}(\varepsilon_{n\mathbf{k}}^{G_0W_0}) | \psi_{n\mathbf{k}} \rangle$.

CP2K also allows performing eigenvalue self-consistency in G (evGW_0) and eigenvalue self-consistency in G and in W (evGW), where especially the G_0W_0 quasiparticle energies can be strongly influenced by the DFT XC functional. In general, we recommend using evGW_0, starting from the PBE functional,[35] for both molecules and solids, motivated by the discussion in ref 125.

CP2K contains three different *GW* implementations: *GW* for molecules,[104] *GW* for computing the band structure of a solid with a small unit cell with $\mathbf{k}$-point sampling in DFT,[126] and *GW* for computing the band structure of a large cell in a Γ-only approach.[127] In the following, we will discuss the details and usage of all of these *GW* implementations.

3.2.3. *GW* for Molecules.
To start a G_0W_0, evGW_0 or evGW calculation for a molecule, one needs to set the keyword RUN_TYPE ENERGY in the &GLOBAL section and the following segment:

```
&XC
    &XC_FUNCTIONAL PBE
    &END XC_FUNCTIONAL
    &WF_CORRELATION
        &RI_RPA
            QUADRATURE_POINTS 100
            &GW
                ! Also possible: EV_GW0 or EV_GW
                SELF_CONSISTENCY G0W0
            &END GW
        &END RI_RPA
    &END WF_CORRELATION
&END XC
```

Therein, the following keywords have been used:

- QUADRATURE_POINTS: number of imaginary-frequency points for computing the self-energy, see eq 21 in ref 104. Usually, 100 points converge the quasiparticle energies within 10 meV.
- SELF_CONSISTENCY: determines which *GW* self-consistency variant (G_0W_0, evGW_0 or evGW) is used to calculate the *GW* quasiparticle energies.

The numerical precision of the *GW* implementation is 10 meV compared to reference calculations, for example, on the *GW*100 test set.[125,128] Furthermore, the following DFT

settings will also have an influence on *GW* quasiparticle energies:

- XC_FUNCTIONAL: starting XC functional for the G_0W_0, evGW_0 or evGW calculation, respectively. We recommend to use evGW_0@PBE, as discussed in ref 125. For further guidance on selecting an appropriate DFT starting functional and self-consistency scheme for your system, you may consult ref 119.

- BASIS_SET: the basis set is of Gaussian type and can have a strong influence on the quasiparticle energies. For computing quasiparticle energies, a basis set extrapolation is necessary,[104] and we recommend all-electron GAPW calculations with correlation-consistent basis sets cc-pVDZ, cc-pVTZ, cc-pVQZ from the EMSL Basis Set Exchange.[44] For computing the HOMO − LUMO gap from *GW*, we recommend the usage of augmented basis sets, for example aug-cc-pVDZ and aug-cc-pVTZ. As RI_AUX basis set, we recommend the RIFIT basis sets from the EMSL database, for example aug-cc-pVDZ-RIFIT.

The computational effort of the *GW* calculation increases with $O(N^4)$, where N is the system size. The memory requirement increases with $O(N^3)$. For large-scale calculations, we recommend starting with a small molecule. After successfully completing the *GW* calculation for the small molecule, you can gradually increase the size of the molecule. The computational resources needed for larger molecules can then be estimated using the $O(N^4)$ scaling for computation time and $O(N^3)$ scaling for memory. The output provides a useful lower limit of the required memory, e.g.

```
Total memory for (ia|K) integrals:   90556 MiB
Total memory for GW-(nm|K) integrals: 4626 MiB
```

When facing an out-of-memory error, please increase the number of nodes for your calculation.

3.2.4. *GW* for Small Unit Cells with k-Point Sampling.

For a periodic *GW* calculation, **k**-point sampling is required. In the following example of a 2D-periodic cell, **k**-point sampling is included in the DFT section:

```
&DFT
   ...
   &KPOINTS
      SCHEME MONKHORST-PACK 32 32 1
      PARALLEL_GROUP_SIZE -1
   &END KPOINTS
&END DFT
```

The **k**-point mesh size is a convergence parameter, and 32 × 32 is expected to reach convergence of the *GW* band gap within 10 meV for a 2D material.

A periodic *GW* calculation is activated via the &BAND-STRUCTURE section, e.g.

```
&PROPERTIES
   &BANDSTRUCTURE
      &GW
         NUM_TIME_FREQ_POINTS      30
         MEMORY_PER_PROC           10
         EPS_FILTER           1.0E-11
         REGULARIZATION_RI    1.0E-2
         CUTOFF_RADIUS_RI         7.0
      &END GW
      &SOC
      &END SOC
      &BANDSTRUCTURE_PATH
         NPOINTS 19
         SPECIAL_POINT K     1/3 1/3 0.0
         SPECIAL_POINT GAMMA 0.0 0.0 0.0
         SPECIAL_POINT M     0.0 0.5 0.0
      &END BANDSTRUCTURE_PATH
   &END BANDSTRUCTURE
&END PROPERTIES
```

All parameters from above have been chosen to converge the *GW* band gap within 10 meV, see also convergence tests in ref 126:

- NUM_TIME_FREQ_POINTS: number of imaginary-time and imaginary-frequency points used for computing the self-energy. Between 20 and 30 points are usually enough for converging quasiparticle energies within 10 meV. Grids up to 34 points are available.
- MEMORY_PER_PROC: specifies the available memory per MPI process. A larger MEMORY_PER_PROC can increase performance.
- EPS_FILTER: filter for three-center integrals, 10^{-11} should be well-converged.
- REGULARIZATION_RI: regularization parameter for RI basis set. For a big RI basis set (>50 RI function per atom), we recommend 10^{-2} to prevent linear dependencies. For a small RI basis set, one can turn RI regularization off by setting 0.0.
- CUTOFF_RADIUS_RI: cutoff radius of truncated Coulomb metric in Å. A larger cutoff leads to faster RI basis set convergence, but also the computational cost increases. A cutoff of 7 Å is an accurate choice.
- &SOC: activates spin−orbit coupling (SOC) from GTH PPs.[32] But, the usage of SOC also needs POTENTIAL_FILE_NAME GTH_SOC_POTENTIALS.
- &BANDSTRUCTURE_PATH: specify the **k**-path in the Brillouin zone for computing the band structure. Relative **k**-coordinates are needed, which you can retrieve for your crystal structure from ref 129.

We recommend the TZVP-MOLOPT basis sets together with GTH PPs.[126] At present, 2D PBCs are supported, while 1D- and 3D PBCs are work in progress.

The *GW* band structure is written to the files bandstructure_SCF_and_G0W0 and bandstructure_SCF_and_G0W0_plus_SOC, respectively. The direct and indirect band gaps are also listed in the CP2K output file. When facing an out-of-memory crash, please increase MEMORY_PER_PROC.

3.2.5. *GW* for Large Cells in Γ-Only Approach.
For a large unit cell, a Γ-only *GW* algorithm is available in CP2K. The requirement on the cell is that elements of the density matrix decay by several orders of magnitude when the two basis functions of the matrix element have a distance similar to the cell size. As a rule of thumb, for a 2D material, a 9 × 9 unit cell is large enough for the Γ-only algorithm.[127]

The input file for a Γ-only *GW* calculation is identical to that of *GW* for small cells with **k**-point sampling except that the `&KPOINTS` section in DFT needs to be removed. The computational parameters from such an input file reach a numerical convergence of the band gap within ∼50 meV (TZVP basis set, 10 time and frequency points).[127] The code prints restart files with ending `.matrix` that can be used to restart an interrupted calculation.

4. EMBEDDING METHODS

In this section, multiscale methods that permit embedding part of a system, which is described using a quantum mechanical electronic structure method, into a surrounding environment at a lower level of theory, are presented and discussed in detail. However, instead of the simple mechanical coupling approach, which can be easily implemented using the `&FORCE_EVAL%MIXED` framework[130] and adaptive resolution simulation methods,[131−133] the focus here will be on more sophisticated methods including nontrivial interactions with the environment.

4.1. Implicit Solvation Methods

The simulation of a solute like a molecule or a cluster with an explicit solvent captures all the details of the solute−solvent interaction, as well as the dynamics and fluctuations within the solvent. However, often embedding the solute within an implicit solvent is sufficient, especially when only an approximate (averaged) inclusion of polarization effects by dielectric screening is required. Representing the solvent as a continuous dielectric medium significantly reduces the number of degrees of freedom and avoids sampling of potentially uninteresting fluctuations within the solvent, which can also lead to great computational time savings. Moreover, the preparation and equilibration of a solvated system in a simulation box becomes also unnecessary. A large number of continuum solvation models can be found in the literature,[134] but these models usually employ a rigid cavity for the solute, introducing a discontinuity in the dielectric function at the vacuum−cavity interface. Such discontinuities cause kinks in the atomic forces which are detrimental in structure relaxations and AIMD simulations. Fattebert and Gygi proposed a smooth dielectric function based on the electron density to palliate the problem.[135,136] In their approach, the response to the electronic density change is self-consistently taken into account in each SCF iteration step, introducing a nested SCF loop within each SCF WF optimization step, which is the price to pay for getting rid of the explicit solvent molecules. This self-consistent continuum solvation (SCCS) model of Fattebert−Gygi and its revised form of Andreussi et al.,[137] are both implemented in CP2K.[138,139] The SCCS input parameters are defined in the `&FORCE_EVAL%DFT%SCCS` section. An SCCS input snippet for a molecule immersed in water ($\varepsilon_0 = 78.36$) looks like:

```
&DFT
  &SCCS on
    ALPHA [mN/m] 50.0
    BETA [GPa] -0.35
    DELTA_RHO 2.0E-5
    DERIVATIVE_METHOD CD5 # or FFT
    DIELECTRIC_CONSTANT 78.36
    EPS_SCCS 1.0E-9
    GAMMA [mN/m] 0.0
    EPS_SCF 1.0E-3
    MAX_ITER 100
    METHOD Andreussi # or Fatteberg-Gygi
    MIXING 0.6
    &ANDREUSSI
      RHO_MAX 0.005
      RHO_MIN 0.0001
    &END ANDREUSSI
    &FATTEBERT-GYGI
      BETA 1.7
      RHO_ZERO 6.0E-4
    &END FATTEBERT-GYGI
  &END SCCS
  ...
&END DFT
```

The values for the keywords `RHO_MAX` and `RHO_MIN` determine the smoothing of the dielectric function for the revised SCCS method of Andreussi et al.[137] Likewise, `BETA` and `RHO_ZERO` are the smoothing parameters for the original SCCS method of Fattebert and Gygi.[135] They are solute dependent, and especially charged solutes require different values for the revised SCCS model.[140] While for neutral solutes 0D, 2D, 1D, and 3D PBCs can be applied, charged systems should be run with 0D PBC using a Poisson solver like MT:[141]

```
&DFT
  &POISSON
    PERIODIC none
    POISSON_SOLVER MT
  &END POISSON
  ...
&END DFT
```

while keeping the periodicity defined in the `&CELL` section identical, i.e.

```
&SUBSYS
  &CELL
    PERIODIC none
    ...
  &END CELL
  ...
&END SUBSYS
```

The first WF optimization steps are performed without SCCS. The SCCS input parameter `EPS_SCF` determines the SCF convergence threshold for the onset of the nested (inner) self-consistent iteration of the polarization potential, and `EPS_SCCS` defines the corresponding convergence threshold for the requested numerical accuracy, performing a maximum of `MAX_ITER` SCCS iteration steps.

The keywords `ALPHA`, `BETA`, and `GAMMA` are optional parameters for additional solvation model terms besides the electrostatic contribution ΔG^{el} defining the solvation free energy

$$\Delta G^{sol} = \Delta G^{el} + G^{cav} + G^{rep} + G^{dis} \tag{31}$$

with a cavitation term

$$G^{\text{cav}} = \gamma S \tag{32}$$

which can be computed based on the "quantum surface" introduced by Cococcioni et al.[142] The latter reads as

$$S = \int \{\vartheta_{\rho_0 - \Delta/2}[\rho^{\text{elec}}(\mathbf{r})] - \vartheta_{\rho_0 + \Delta/2}[\rho^{\text{elec}}(\mathbf{r})]\} \times \frac{|\nabla \rho^{\text{elec}}|}{\Delta} d\mathbf{r} \tag{33}$$

where Δ and γ are defined by the input parameters `DELTA_RHO` and `GAMMA`, respectively.

The terms accounting for the Pauli repulsion G^{rep} and the dispersion interactions G^{dis} can be computed based on the surface S and the volume V of the solute cavity, i.e.

$$G^{\text{rep}} + G^{\text{dis}} = \alpha S + \beta V \tag{34}$$

The prefactors `ALPHA` and `BETA` are solvent-specific tunable parameters. This ultimately allows the calculation of a solvation free energy

$$\Delta G^{\text{sol}} = \Delta G^{\text{el}} + (\alpha + \gamma)S + \beta V \tag{35}$$

Finally, the SCCS printout is controlled by the input section `FORCE_EVAL%DFT%PRINT%SCCS`:

```
&PRINT
    &SCCS ON
        &EACH
            QS_SCF 0
        &END EACH
        &POLARISATION_POTENTIAL off
            &EACH
                QS_SCF 0
            &END EACH
        &END POLARISATION_POTENTIAL
        ...
    &END SCCS
    ...
&END PRINT
```

allowing for printing the polarization potential and the dielectric function in cube file format.

4.2. Quantum Mechanics/Molecular Mechanics Methods

In quantum mechanics/molecular mechanics (QM/MM) hybrid schemes,[143] the system is divided into two subsystems: one is treated at the quantum mechanical (QM) level of theory and the other at the molecular mechanics (MM) level. In CP2K, an additive QM/MM scheme is implemented, which partitions the QM/MM total energy $E_{\text{totQM}/}^{\text{MM}}$ into the following three terms

$$E_{\text{tot}}(\mathbf{R}_\alpha, \mathbf{R}_a) = E_{\text{QM}}(\mathbf{R}_\alpha) + E_{\text{MM}}(\mathbf{R}_a) + E_{\text{QM/MM}}(\mathbf{R}_\alpha, \mathbf{R}_a) \tag{36}$$

where $\mathbf{R}_\alpha$ and $\mathbf{R}_a$ label the coordinates of the QM and MM nuclei, respectively. The QM contribution is evaluated using the QUICKSTEP module,[23] typically employing KS-DFT. Alternatively, SQC methods, such as PM6,[144] self-consistent-charge density-functional tight-binding (SCC-DFTB),[145] or extended tight-binding (xTB) approaches can be used.[146] MM contributions are obtained from classical force fields, either using the built-in FIST program, or by an interface to GROMACS.[147,148]

The QM/MM energy includes nonbonded terms and, if the QM system is connected to the MM region by covalent bonds, bonded terms as well. Nonbonded terms comprise van der Waals (vdW) and electrostatic (elec) interactions, i.e.

$$\begin{aligned} E_{\text{QM/MM}} &= E_{\text{QM/MM}}^{\text{bonded}} + E_{\text{QM/MM}}^{\text{non-bonded}} \\ &= E_{\text{QM/MM}}^{\text{bonded}} + E_{\text{QM/MM}}^{\text{vdW}} + E_{\text{QM/MM}}^{\text{elec}}. \end{aligned} \tag{37}$$

Speaking about bonded terms, special care has to be taken with respect to unsaturated bonds that are occurring when cutting bonds of covalently coupled subsystems. For that purpose, CP2K offers two general options. The simplest approach is to saturate the dangling bond of the QM atom with a monovalent atom (link atom), typically a hydrogen atom. The QM calculations are then performed on the QM atoms plus the link atom. The interaction between the covalently bound QM and MM atom is modeled with classical MM potentials. This approach is often termed the link-atom technique in the literature and referred to as the integrated MO molecular mechanics (IMOMM) method in CP2K.[149] Alternatively, a linking atom with a monovalent PP can replace the MM frontier atom at the QM/MM boundary.[149−152]

Turning to the nonbonded terms, the vdW interactions between QM and MM atoms are often described using a simple Lennard-Jones potential.[149] The treatment of electro-static interactions, however, is the most challenging aspect of the QM/MM Hamiltonian and can be approached with varying levels of complexity.[149] Generally, we distinguish between three different approaches, which are all available in CP2K. The simplest approach is a mechanical embedding scheme, in which precalculated partial atomic charges are assigned to both QM and MM atoms. The charge−charge interactions are then computed using classical expressions, neglecting the response of the QM charge density to the electrostatic potential generated by the MM charges. The most popular scheme, and the default treatment in CP2K, is electrostatic embedding.[153,154] In this approach, an additional term is added to the QM Hamiltonian, allowing the MM charges to polarize the QM region. For QM calculations at the KS-DFT level with GPW and GAPW, CP2K implements a particularly efficient variant of electrostatic embedding called Gaussian expansion of the electrostatic potential (GEEP), which is explained in more detail below. The next level of sophistication is polarized embedding schemes, where the QM atoms can polarize the QM part nonself-consistently or fully self-consistently. CP2K incorporates polarized embedding schemes to some extent. For example, it offers a fully self-consistent image-charge augmented QM/MM (IC-QM/MM) scheme,[155] specifically suited for adsorbate/metal systems. A nonself-consistent core−shell model, which can be applied to a broad range of systems, is also available.[156]

A minimal example of a QM/MM calculation in a periodic cubic simulation box of length 25 Å, where atoms with indices 1 to 6 are described at the KS-DFT level, looks as follows:

```
&FORCE_EVAL
  METHOD QMMM
  ...
  &QMMM
    ECOUPL Coulomb
    &CELL
      ABC 25.0 25.0 25.0
      PERIODIC XYZ
    &END CELL
    &QM_KIND C
      MM_INDEX 1..6
    &END QM_KIND
    &PERIODIC
      &MULTIPOLE OFF
      &END MULTIPOLE
    &END PERIODIC
  &END QMMM
&END FORCE_EVAL
```

In addition to that, the DFT section must be defined for the QM part of the calculation, as well as the MM section with all classical force field parameters, including the classical interactions between the QM and MM subsystems, such as the vdW interactions. The relevant keywords and subsections within the &QMMM section are

- ECOUPL: defines the type of electrostatic treatment. On the one hand GAUSS enables the electrostatic embedding for KS-DFT/MM simulations based on GEEP (see next section). On the other hand COULOMB enables electrostatic embedding for SQC methods, or SCC-DFTB/xTB. Mechanical embedding schemes can be used by setting the keyword to NONE.

- CELL: defines the cell for the QM calculation, which must be orthorhombic.

- QM_KIND: defines the QM system and the corresponding indices are given for each atomic type separately.

- PERIODIC: applies the periodic potential. The MULTIPOLE section turns on the coupling/recoupling of the QM periodic images, as described later. In the previous example, it was turned off because the QM box and the MM box are identical in size.

In our simple example, no covalent bonds are cut through the subsystem boundaries. For cases where covalent bonds are cut, the corresponding techniques and link atoms can be set through the &LINK section.

4.2.1. Electrostatic Embedding by the Gaussian Expansion of the Electrostatic Potential Method. The calculation of electrostatics terms in QM-MM poses problems related to both short-range and long-range behavior. The problems of short-range behavior are related to the electron spill-out. Classical atoms are normally represented by a simple point charge. If a QM atom comes close, the electrons can be trapped into the point-like classical potential energy source. The use of diffuse basis sets (or PWs) can enhance this behavior. The ad-hoc generated pseudopotential-like approach has often been applied to remedy the unphysical problem of overpolarization that arises when MM atoms are in close contact with the QM region. This approach requires the definition of frontier MM and frontier QM atoms, which could result in being cumbersome and not sufficiently adaptable to the dynamic evolution of the system. In CP2K, the spill-out is avoided by assigning to all MM atoms a finite-width charge density in the form of a Gaussian distribution

$$n(\mathbf{r}, \mathbf{R}_{\mathrm{MM}}) = \left(\frac{r_{c,\mathrm{MM}}}{\sqrt{\pi}}\right)^3 e^{-(|\mathbf{r}-\mathbf{R}_{\mathrm{MM}}|/r_{c,\mathrm{MM}})^2} \tag{38}$$

The width of the charge density depends on the atom type and can be expected to be similar to the covalent radius. The exact potential originating from the Gaussian charge distribution, as commonly employed in the Ewald method, is

$$v_{\mathrm{MM}}(\mathbf{r}, \mathbf{R}_{\mathrm{MM}}) = \frac{\mathrm{Erf}\left(\frac{|\mathbf{r}-\mathbf{R}_{\mathrm{MM}}|}{r_{c,\mathrm{MM}}}\right)}{|\mathbf{r}-\mathbf{R}_{\mathrm{MM}}|} \tag{39}$$

This potential tends to $1/r$ at large distances and converges smoothly to a constant at zero. The electrostatic interaction between the QM and the MM atoms can then be computed onto the QM grid by multiplying the QM charge and this potential. For a more efficient calculation, though, the MM electrostatic potential is decomposed in a series of N_g Gaussians of different cutoffs plus a residual, which is why we call this algorithm GEEP.[153] Hence

$$\frac{\mathrm{Erf}\left(\frac{r}{r_c}\right)}{r} = \sum_{N_g} A_g e^{-(r/G_g)^2} + R_{\mathrm{low}}(r) \tag{40}$$

In this expression, A_g is the Gaussian amplitude, G_g its width, and R_{low} is the smooth residual function that can be mapped onto a grid with spacing 1 order of magnitude larger than the one needed for the potential. The advantage of adopting the Gaussian expansion is that commensurate grids with different spacing can be used for the different contributions, i.e. sharper Gaussians to finer grids and coarser grids for smoother components. All contributions are finally interpolated onto the finest QM grid, by means of real-space splines, and summed up. The interpolation depends only on the number of grid points and not on the number of MM atoms, so

```
&QMMM
  ECOUPL GAUSS
  USE_GEEP_LIB 9
  &MM_KIND H
    RADIUS 0.44
  &END MM_KIND
  &MM_KIND O
    RADIUS 0.78
  &END MM_KIND
  &QM_KIND C
    MM_INDEX 1..6
  &END QM_KIND
&END QMMM
```

The treatment of long-range forces in conjunction with PBC is much less well established for QM/MM and most of the implementations use a truncation scheme, neglecting interactions beyond a certain cutoff. The reaction field approach combines truncation and polarizable continuum beyond a given cutoff. Ewald techniques are usually implemented only for QM-MM interactions and not for QM−QM, though long-range QM−QM may play a significant role, in particular for materials science applications. The electrostatic energy in CP2K is then composed of three terms, i.e. $E_{\mathrm{ES}} = E^{\mathrm{MM}} + E^{\mathrm{QM}} + E^{\mathrm{QM/MM}}$. The first term is evaluated using standard techniques, such as particle−particle or particle-mesh schemes. The second

term is the evaluation of the energy of the QM subsystem. Since the total energy of the QM subsystem is usually evaluated using a smaller cell, care needs to be taken to include the correct electrostatics. The last term is the evaluation of the periodic electrostatic potential $v_{MM}(\mathbf{r}, \mathbf{R}_{MM})$ discussed above, divided into a real-space contribution and a periodic correction. The real-space term contains the interactions due to the short-range part of the electrostatic potential of the MM charges. Only MM atoms close to the QM region will contribute to this term. Since we use the Gaussian expansion plus the residual term, the radius of the Gaussian is such that only a few terms in the lattice sum are needed. The effect of the periodic replicas is only in the long-range term and it comes entirely from the residual function, thus

```
&QMMM
  ECOUPL GAUSS
  USE_GEEP_LIB 6
  &PERIODIC
    GMAX 0.5
    NGRIDS  20 20 20
    REPLICA  2
    &MULTIPOLE
      EWALD_PRECISION 0.00000001
      RCUT 8.0
      ANALYTICAL_GTERM
    &END MULTIPOLE
  &END PERIODIC
&END QMMM
```

The residual function is represented in the Fourier space and can be evaluated analytically. Furthermore, since the long-range contribution is very smooth, the Fourier transform is zero for all G vectors larger than a well-defined maximum. In this example, `G_MAX` specifies the maximum G vector in reciprocal space for the Ewald sum and depends strongly on the number of Gaussian functions used in the GEEP scheme. The number of grid points used for the interpolation of the G-space term is given by `NGRIDS`, whereas `REPLICA` is the number of cell replicas to take into consideration for the real-space sum. Usually, 3−4 grid levels are used corresponding to a speed-up of 10^2 times faster than the simple collocation algorithm (interpolations and restrictions account for a negligible amount of computing time). Since the residual function is different from zero only for a few vectors, the sum in reciprocal space is restrained to a few points. Possible sources of error are the cutoff of the finest grid level to properly map the sharpest Gaussian functions, the cutoff of the coarse grid level related to the cutoff of the long-range residual function, and the error in cubic spline interpolation.

When computing the QM electrostatic term, unless the quantum box and the MM box have the same dimensions, the QM images, which are interacting by PBC implicitly in the evaluation of the Hartree potential, have the wrong periodicity. To correct for this error, CP2K implements the so-called Blöchl scheme for decoupling and recoupling according to the correct periodicity.[157,158] Spherical Gaussians, which are QM atom centered and reproduce the correct multiple expansion, are derived by a density fitting scheme in G-space. They can be more than just one per atom site with different exponents. These charges reproduce the correct long-range electrostatics and are used for the decoupling and recoupling procedure. Indeed, since the electrostatic interaction of separated charge distributions (the array of periodic QM charge densities) depends only on its multipole moments, the model charge density is used to modify the Hartree potential and cancel the electrostatic interactions between the periodic images. The `MULTIPOLE` section sets up this scheme and is activated by default for periodic calculations. It should be switched off when QM and MM boxes are the same size, to avoid unnecessary computational costs.

4.2.2. Image-Charge Augmented Quantum Mechanics/Molecular Mechanics Method. The IC-QM/MM approach is intended to simulate adsorbed molecules on metallic surfaces.[155] The molecular adsorbates are described at the QM level, specifically using KS-DFT, while the metal is treated at the MM level. In the IC-QM/MM scheme, as implemented in CP2K, the electrostatic response of the metal to the presence of the adsorbate is described by an image charge distribution

$$\rho_{m}(\mathbf{r}) = \sum_{a} c_{a}g_{a}(\mathbf{r}, \mathbf{R}_{a}) \tag{41}$$

where $\mathbf{R}_a$ is the position of the metal atom a and g_a represents a Gaussian function centered at a. The charge distribution ρ_m generates the potential V_m, while the charge distribution of the adsorbates produces the electrostatic potential V_e. The coefficients c_a are determined self-consistently by enforcing the constant-potential condition, where $V_m(\mathbf{r})$ screens $V_e(\mathbf{r})$ within the metal such that $V_e(\mathbf{r}) + V_m(\mathbf{r}) = V_0$. Therein, V_0 denotes a constant potential, which is typically zero unless an external potential is explicitly applied.

To run an IC-QM/MM calculation, the following `&IMAGE_CHARGE` subsection must be added:

```
&QMMM
  ...
  &IMAGE_CHARGE
    MM_ATOM_LIST 1..576
    EXT_POTENTIAL 0.0
  &END IMAGE_CHARGE
&END QMMM
```

In an IC-QMM/MM calculation, the MM system is typically the whole metallic slab. Hence, QM and MM cells are thus required to have the same size. The relevant keywords are

- `MM_ATOM_LIST`: defines the list of MM atoms that carry the Gaussian charge g_a. These should be the atoms of the metallic slab.

- `EXT_POTENTIAL`: sets the external potential V_0.

There are several other keywords related to print options and to acceleration techniques of an IC-QM/MM calculation for specific run types, see online tutorial for more details.[159] Note that the `IMAGE_CHARGE` section only accounts for the electrostatic interaction between molecules and metal atoms. Other contributions, such as vdW interactions, must be defined separately.

4.2.3. Partial Atomic Charges from Restrained Electrostatic Potential Fitting. Restrained electrostatic potential (RESP) fitting is a widely used method to determine partial atomic point charges q_a.[160] In QM/MM simulations, these charges are required for the MM atoms when using the default electrostatic embedding of CP2K, and for both QM and MM atoms when using the less common mechanical embedding.

These charges are determined such that the potential V_{RESP} generated by q_a reproduces a given QM potential V_{QM} within a specified region of space. In practice, this is achieved by minimizing the residual R_{esp} through a least-squares fitting

procedure for a set of predefined real-space grid points $\mathbf{r}_k$, where R_{esp} is given by

$$R_{esp} = \frac{1}{N} \sum_k^N \left(V_{QM}(\mathbf{r}_k) - V_{RESP}(\mathbf{r}_k) \right)^2 \tag{42}$$

where N denotes the total number of real-space grid points $\mathbf{r}_k$. The resulting q_a are treated as point charges in nonperiodic RESP fittings. For periodic systems, however, the charges are represented by Gaussian functions g_a of fixed width, centered on atom a, resulting in the charge distribution

$$\rho_{RESP} = \sum_a q_a g_a \tag{43}$$

For details on the GPW-based periodic RESP implementation in CP2K, see ref 161. The choice between nonperiodic and periodic RESP fitting is made automatically, depending on the periodicity of the reference potential V_{QM}.

The QM potential V_{QM} is obtained from a previous DFT or HF calculation. The RESP fit starts as a postprocessing step and is enabled by

```
&PROPERTIES
    &RESP
        &SPHERE_SAMPLING
        &END SPHERE_SAMPLING
    &END RESP
&END PROPERTIES
```

The choice of sampling points $\mathbf{r}_k$ is crucial to obtain meaningful charges that accurately reproduce the electrostatic potential in spatial regions relevant to interatomic interactions. As a general guideline, regions where the QM potential V_{QM} varies rapidly (e.g., within the vdW radii of atoms) should be avoided. The specific choice of sampling regions is system-dependent. The following options are available:

- SPHERE_SAMPLING: the real-space points $\mathbf{r}_k$ are sampled in spherical shells around each atom. The shells are defined by a minimal and maximal radius, which can be set by the RMIN and RMAX keywords in this section. This option should be used for molecules, molecular liquids, or porous periodic systems like metal−organic frameworks.

- SLAB_SAMPLING: the option should be used for slab-like systems, where it is important to reproduce the potential well above the surface, e.g. to study adsorption processes. The $\mathbf{r}_k$ grid is then sampled as a thin slice above the surface. In that case, keywords defining the surface atoms, the direction, and the thickness of the slice need to be set.

A set of constraints and restraints is typically employed to avoid unphysical values for q_a and to stabilize the fit. The total residual R, which is minimized, is $R = R_{esp} + R_{rest} + R_{const}$. For the restraint, CP2K uses a harmonic penalty function

$$R_{rest} = \beta \sum_j \left(q_j - t_j \right)^2 \tag{44}$$

where β denotes the strength of the restraint and t_j is the anticipated target charge. Restraints can be explicitly set by

```
&RESP
    ...
    &RESTRAINT
        ATOM_LIST 1..3
        TARGET -0.18
        STRENGTH 0.0001
    &END RESTRAINT
    RESTRAIN_HEAVIES_TO_ZERO .FALSE.
&END RESP
```

In this example, the target charges t_j in eq 44 are set to -0.18 for atoms with the index 1 to 3. The strength β of the restraint is defined by the keyword STRENGTH. When explicitly defining restraints, the default restraint RE-STRAIN_HEAVIES_TO_ZERO, which sets t_j to zero for all elements but hydrogen, should be turned off. Different constraints are possible, yet by default, CP2K constrains the sum of all fitted charges to the total charge q_{tot} of the system, i.e.

$$R_{const} = \lambda \sum_j \left(q_j - q_{tot} \right) \tag{45}$$

Further explicit constraints can be given by adding

```
&CONSTRAINT
    EQUAL_CHARGES
    ATOM_LIST 1 2 3
&END
```

to the &RESP section. This is to say that the atoms with indices 1, 2, and 3 carry the same charge. More details on how to set constraints can be found in the online tutorial.[162]

CP2K also features a GPW implementation of the so-called repeating electrostatic potential extracted atomic (REPEAT) method,[163] which is an adaptation of the RESP approach for periodic systems. When using the REPEAT method, eq 42 is modified so that the variance of the potentials is fitted rather than the absolute difference. Fitting the variance is generally easier and helps stabilize the fit for periodic systems. We note that the REPEAT method was originally introduced to handle the arbitrary offset of the electrostatic potentials V_{QM} and V_{RESP} in infinite systems.[163] In CP2K, both potentials are computed with the GPW method, sharing thus the same offset. The REPEAT method is activated by adding the keyword USE_REPEAT_METHOD to the &RESP section. More details on the REPEAT implementation in CP2K can be found in ref 14 and in the online tutorial.[162]

4.3. Density Functional Embedding Theory

Density functional embedding theory (DFET) separates the systems into the relevant cluster and an environment. The cluster is described with the high-level electronic structure method, often employing a correlated WF (CW) ansatz, while the environment and the interaction between the cluster and the environment are described with DFT via the uniquely defined local embedding potential $v_{emb}(\mathbf{r})$.[164] The total energy of the system is calculated within first-order perturbation theory, i.e.

$$E_{total}^{DFET} = E_{total}^{DFT} + \left(E_{cluster,emb}^{CW} - E_{cluster,emb}^{DFT} \right) \tag{46}$$

where E_{total}^{DFT} and $E_{cluster,emb}^{DFT}$ are the DFT energies of the entire system and the embedded subsystem, respectively, while $E_{cluster,emb}^{CW}$ is the energy of the embedded cluster at the CW level of theory. All these entities are computed with an additional one-electron embedding term $\int d\mathbf{r} \, V_{emb} \rho(\mathbf{r})$ in the Hamiltonian.

The embedding potential is obtained in a top-down approach from the condition that the sum of embedded subsystem densities should reconstruct the DFT density of the total system. This can be achieved by maximizing the Wu—Yang functional with respect to $v_{\mathrm{emb}}(\mathbf{r})$,[165] i.e.

$$W[V_{\mathrm{emb}}] = E_{\mathrm{cluster}}[\rho_{\mathrm{cluster}}] + E_{\mathrm{env}}[\rho_{\mathrm{env}}]$$
$$+ \int d\mathbf{r}\, V_{\mathrm{emb}}(\rho_{\mathrm{total}} - \rho_{\mathrm{cluster}} - \rho_{\mathrm{env}}) \tag{47}$$

with the functional derivative being identical to the density difference such that $\frac{\delta W}{\delta V_{\mathrm{emb}}} = \rho_{\mathrm{total}} - \rho_{\mathrm{cluster}} - \rho_{\mathrm{env}}$.

4.3.1. General Procedure. The DFET workflow starts with obtaining the embedding potential $v_{\mathrm{emb}}(\mathbf{r})$. First, one calculates the total density of the system before DFT calculations on the various subsystems with the current updated embedding potential are performed. Then, the potential is updated and the step is repeated until the total DFT density is matched by a sum of the densities of the embedded subsystems. When this condition is fulfilled, the embedded higher-level theory calculation is performed on the isolated cluster.

4.3.2. Implementation. The DFET implementation is available for closed and open shell systems, in terms of unrestricted and restricted open-shell formalisms, respectively. It is limited to GPW calculations only with PPs describing the core electrons. All electronic structure methods implemented within CP2K/QUICKSTEP are available as a higher-level method, including hybrid DFT, MP2, and RPA. It is possible to perform property calculations on the embedded cluster using an externally provided $v_{\mathrm{emb}}(\mathbf{r})$. The subsystems can employ different basis sets, although they must share the same PW grid.

DFET calculations are activated via the `&MULTIPLE_FORCE_EVALS` section in the root:

```
&MULTIPLE_FORCE_EVALS
    FORCE_EVAL_ORDER 2 3 4 5
    MULTIPLE_SUBSYS T
&END MULTIPLE_FORCE_EVALS
```

The order of force evaluations is arbitrary, but it is recommended to leave it as is. Here, `&FORCE_EVAL 2` refers to the environment, 3 to the embedded cluster, and 4 to the total system, all of which are calculated at the DFT level. The isolated cluster computed at a higher-level of theory is denoted as `&FORCE_EVAL 5`. The first force evaluation in `&FORCE_EVAL` defines the general embedding framework:

```
&FORCE_EVAL
  METHOD EMBED
  &EMBED
    &MAPPING
      &FORCE_EVAL 1
        &FRAGMENT 1
          1 3
          MAP 1
        &END FRAGMENT
      &END FORCE_EVAL
      &FORCE_EVAL 2
        &FRAGMENT 1
          1 2
          MAP 2
        &END FRAGMENT
      &END FORCE_EVAL
      &FORCE_EVAL 3
        &FRAGMENT 1
          1 5
          MAP 3
        &END FRAGMENT
      &END FORCE_EVAL
      &FORCE_EVAL 4
        &FRAGMENT 1
          1 2
          MAP 2
        &END FRAGMENT
      &END FORCE_EVAL
      &FORCE_EVAL_EMBED
        &FRAGMENT 1
          1 3
        &END FRAGMENT
        &FRAGMENT 2
          4 5
        &END FRAGMENT
        &FRAGMENT 3
          1 5
        &END FRAGMENT
      &END FORCE_EVAL_EMBED
    &END MAPPING
  &END EMBED
  ...
&END FORCE_EVAL
```

The mappings define the correspondence of atoms in the parts and in the total system, i.e. atoms 1 and 2 (`MAP 2` of `&FORCE_EVAL 2`) correspond to atoms 4 and 5 in the total system (`&FRAGMENT 2` in the `&FORCE_EVAL_EMBED` section). In addition, this `&FORCE_EVAL` section must represent the total system as defined in the `&SUBSYS` section. The next two `&FORCE_EVAL` sections specify the environment and the cluster via standard DFT input sections. The latter one has the following additional keyword in the `&QS` section:

```
CLUSTER_EMBED_SUBSYS .TRUE.
```

The next `&FORCE_EVAL` section to be defined is the total system, marked by the keyword

```
REF_EMBED_SUBSYS .TRUE.
```

in the `&QS` section. In this force evaluation, the options for optimizing the embedding potential are to be specified. This is done by inserting a `&OPT_EMBED` section within the `&QS` section:

```
&OPT_EMBED
...
&END OPT_EMBED
```

The final `&FORCE_EVAL` section represents the higher-level calculation of the embedded cluster. It can be either a

standard hybrid/double-hybrid DFT or a CW calculation with the following additional keyword in the &QS section:

```
HIGH_LEVEL_EMBED_SUBSYS .TRUE.
```

Several options for representing and optimizing the embedding potential are implemented in CP2K. Most importantly is the &OPT_EMBED%GRID_OPT keyword, which defines whether the potential is represented on the real-space grid, or expanded in the auxiliary basis set. The first option is more accurate and computationally cheaper, and is therefore recommended for all practical purposes.

In addition, several techniques for optimizing $v_{emb}(\mathbf{r})$ are available, including standard gradient-based optimization schemes (e.g., SD, quasi-Newton, and level-shifting methods), as well as an iterative van Leeuwen−Baerends update,[166] which is an alternative to the previously mentioned Wu−Yang functional.[165] Despite its simplicity, the SD approach is typically relatively robust for potential optimizations and is therefore the default. Although several options for the initial guess of the embedding potential are available, starting with the zero potential is a rather reliable procedure, which is typically quickly converging.

The embedding potential is saved as a volumetric Gaussian cube file that can be used for visualization purposes and can be read in for restarting the optimization. For open-shell calculations, CP2K defines two potentials: one interacts with the electron density, whereas the other interacts with the spin density. Both are saved and should be specified for restarting the calculation. In addition, the embedding potential can be used for standalone embedded calculations, as in the following example for an open-shell system:

```
&QS
    DFET_EMBEDDED .TRUE.
    EMBED_CUBE_FILE_NAME potential.cube
    EMBED_SPIN_CUBE_FILE_NAME
    spin_potential.cube
    ...
&END QS
```

5. NUCLEAR MAGNETIC AND ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY

The energy levels probed in magnetic spectroscopy correspond to transitions among nuclear and electronic spin eigenstates in the presence of an external magnetic field. The calculation of magnetic response properties in CP2K is based on variational density functional perturbation theory (DFPT).[167−169]

5.1. Density Functional Perturbation Theory

Generally speaking, CP2K/QUICKSTEP uses DFPT to compute the perturbative corrections to the zeroth-order orbitals $\psi_k^{(0)}$. In case of a perturbing magnetic field B

$$\psi_k = \psi_k^{(0)} + B\psi_k^{(1)} + \dots \tag{48}$$

where $\psi_k^{(1)}$ is the first-order orbital correction, which provides energy corrections up to third order according to Wigner's ($2n + 1$) rule.

The first-order orbitals can be obtained via the inhomogeneous set of coupled (Sternheimer) equations

$$-\sum_i^{N_{occ}} (H^{(0)}\delta_{ij} - \langle\psi_i^{(0)}|H^{(0)}|\psi_j^{(0)}\rangle)|\psi_i^{(1)}\rangle = H^{(1)}|\psi_j^{(0)}\rangle \tag{49}$$

Therein, the indices i, j run over the occupied orbital manifold $\{\psi\}$. This expression already anticipates the possibility that the employed zeroth-order orbitals might not be the canonical KS orbitals. Expanding the first-order orbitals in terms of atomic basis functions ϕ_l

$$\psi_l^{(1)} = \sum_l c_{li}\phi_l^{(1)} \tag{50}$$

and projecting eq 49 on ϕ_k

$$-\sum_i^{N_{occ}} \sum_l^{N_{basis}} (H_{kl}^{(0)}\delta_{ij} - S_{kl}\langle\psi_i^{(0)}|H^{(0)}|\psi_j^{(0)}\rangle)c_{li}^{(1)} = \sum_l H_{kl}^{(1)}c_{lj}^{(0)} \tag{51}$$

we obtain a set of $N_{basis} \times N_{occ}$ simultaneous equations. In case of magnetic responses, the first-order orbitals are imaginary, and one explicitly includes $i = \sqrt{-1}$ on the left-hand side, so that the expansion coefficients $c^{(1)}$ are real. It is to be noted that with atom-centered basis functions and vibrational perturbations (dipole derivatives, Born charges), the Sternheimer equation acquires an additional contribution on the right-hand side, due to the derivatives of the overlap matrix.

In CP2K, eq 51 are solved self-consistently using a CG-based minimizer.[167,170] In the employed parallel-transport gauge, the first-order orbitals are orthogonal to the occupied orbital manifold.[170] This can be imposed by projecting the right-hand sides of the equations onto the unoccupied orbital manifold, which amounts to solving the equations while imposing the orthogonality between the trial first-order orbitals and the occupied orbitals.

A prototypical input for such a linear response calculation looks like:

```
&GLOBAL
    RUN_TYPE LINEAR_RESPONSE
&END GLOBAL
&FORCE_EVAL
    &PROPERTIES
        &LINRES
            ...
        &END LINRES
    &END PROPERTIES
&END FORCE_EVAL
```

In the relevant &LINRES input section, which is located inside the &PROPERTIES section, one can set the parameters of the preconditioned CG optimizer. The general principles on how to choose the specific preconditioner, maximum number of iterations, and convergence threshold are similar to Section 2.1.4. Yet, the default values typically provide a robust starting point. Nevertheless, the convergence behavior of different response properties can differ, and it is advisable to check whether the desired response property is converged with respect to the chosen CG convergence threshold.

Also inside the &LINRES section, different subsections give access to various response properties. As will be explained later, when calculating magnetic field-induced current densities (required for several magnetic response properties) maximally localized Wannier functions (MLWF) are employed as the zeroth-order orbitals. The corresponding &LOCALIZE input section is also found within the &LINRES section, as are

- &CURRENT: the induced current density in response to an external homogeneous magnetic field.[168]

- &LOCALIZE: computes localized orbitals, which are required if the perturbation includes the position operator.
- &NMR: the nuclear magnetic shielding tensors, or more generally, the nucleus-independent chemical shifts (NICS).[171]
- &EPR: the electronic g-tensor.[172]
- &POLAR: computes the polarizability.[173]
- &VCD: enables the calculation of vibrational circular dichroism (VCD).[174]
- &DCDR: analytical gradients of the dipole moments, e.g. for Born effective charges and atomic polar tensors.

Note that the calculation of some response properties might require the explicit inclusion of more than one of the above, e.g. &CURRENT requires &LOCALIZE. These two sections are in turn prerequisites for &NMR and &EPR sections, respectively. Detailed information relating to individual input sections for magnetic response properties will be given in the following.

5.2. The Magnetic Shielding Tensor

The interaction of electrons with an external magnetic field $\mathbf{B}$ leads to an induced current density $\mathbf{j}(\mathbf{r})$, which in turn leads to an induced magnetic field

$$\mathbf{B}^{\text{ind}}(\mathbf{r}) = \frac{1}{c} \int d\mathbf{r} \frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} \times \mathbf{j}(\mathbf{r}') \tag{52}$$

where c is the speed of light, and Gaussian atomic units are being used. The perturbing field can be an external magnetic field $\mathbf{B}^{\text{ext}}$ (taken to be spatially homogeneous), or the magnetic moments associated with spins. We generally assume in this section a closed-shell system with vanishing total electron spin. In this case, the only spins in the system are the nuclear ones $\mathbf{m}^N$

$$\mathbf{B}^{\text{ind}}(\mathbf{r}, \mathbf{B}^{\text{ext}}, \{\mathbf{m}^N\}) = -\boldsymbol{\sigma}(\mathbf{r}) \mathbf{B}^{\text{ext}} - \sum_i^{\text{Nuclei}} \mathbf{K}_i(\mathbf{r}) \mathbf{m}_i^N(\mathbf{r}_i^N) + \dots \tag{53}$$

The reduced indirect spin–spin coupling is denoted as $\mathbf{K}(\mathbf{r})$ in eq 53, while the tensor field

$$\boldsymbol{\sigma}(\mathbf{r}) = -\frac{\partial \mathbf{B}^{\text{ind}}(\mathbf{r})}{\partial \mathbf{B}^{\text{ext}}} \tag{54}$$

is known as the magnetic shielding tensor. Experimentally, the induced magnetic field can only be probed at the position of a nuclear spin $\mathbf{r}^N$. Being a ratio, $\boldsymbol{\sigma}(\mathbf{r})$ is conventionally reported in units of parts per million. The experimental value is almost always referenced to the isotropic average $\left(\frac{1}{3}\text{Tr}(\boldsymbol{\sigma}(\mathbf{r}))\right)$ of the magnetic shielding tensor of some chosen standard for the respective nucleus, i.e.

$$\boldsymbol{\delta}(\mathbf{r}_i^N) = \sigma_{i,\text{ref}}^{\text{iso}} \mathbf{I} - \boldsymbol{\sigma}(\mathbf{r}_i^N) \tag{55}$$

which is known as the chemical shift tensor, and its isotropic average δ is called the chemical shift. Moreover, $\mathbf{I}$ is the unit matrix.

At lowest order, the energy of the system is linear in the total magnetic field, thus one can also define the shielding tensor as the second derivative of the energy with respect to the nuclear magnetic moment $\mathbf{m}^N$ and the external magnetic field

$$\boldsymbol{\sigma}(\mathbf{r}^N) = \frac{\partial^2 E}{\partial \mathbf{B}^{\text{ext}} \partial \mathbf{m}^N}\Bigg|_{|\mathbf{B}^{\text{ext}}| = |\mathbf{m}^N| = 0} \tag{56}$$

Given that the magnetic shielding is a tensor field, it can be computed in any arbitrarily chosen spatial position, even though it is experimentally measurable only at the nuclear positions. These are known as nucleus-independent chemical shifts (NICS), which are also available in CP2K.[171]

From eqs 52 and 54, the elements of the shielding tensor are given by

$$\sigma_{xy}(\mathbf{r}) = \frac{1}{c} \int_\Omega d\mathbf{r} \left[\frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} \times \mathbf{j}_x(\mathbf{r}') \right]_y \tag{57}$$

where the integration is over the whole material, including all periodic replicas in the case of extended systems. It is obvious that the induced current density is the key ingredient in computing magnetic shieldings. Once $\mathbf{j}(\mathbf{r})$ is known, the induced field, and hence the shielding tensor as well, is available as a simple three-dimensional integral.

Using the approach developed by Sebastiani and Parrinello,[168] CP2K evaluates the magnetic perturbation using MLWFs in combination with DFPT.[172] As explained later, the localized nature of the Wannier functions allows the use of the position operator in the perturbation Hamiltonian. An underlying core assumption here is that each Wannier function is contained entirely within the simulation cell once it is centered. For insulators, it is known that the Gaussian functions decay exponentially, so that for sufficiently large simulation cells, this assumption is not irrelevant.[175] In comparison to other approaches developed for condensed phase systems,[168,176] CP2K differs in its use of local atom-centered Gaussian functions (see Section 2), allowing for reduced complexity algorithms and hence large scale calculations of magnetic resonance parameters. With the GAPW method, CP2K/QUICKSTEP is also able to compute all-electron magnetic resonance parameters. One can also use embedding techniques such as QM/MM methods, and even further decompose the QM part into GAPW/GPW regions. The reliance on the localization properties of Wannier functions means, however, that this implementation is not suitable for conductors, where the Wannier functions decay only algebraically.

In order to compute the shielding tensors one first needs to use the GAPW method with an all-electron basis set (the calculation will also be possible using GPW, but then the contributions of the core electrons to the induced current density are missing). Inside the &LINRES section one then also needs to include the &LOCALIZE section to compute the MLWFs, which are then provided to the linear response module as the zero-order orbitals. Otherwise, CP2K will exit with an error if the user requests the calculation of the current density or the shielding tensors, without activating the &LOCALIZE section. Second, the &CURRENT section should be included for computing the induced current density using DFPT.[167] Finally, the &NMR section requests the printing of the shielding tensors or NICS maps by performing the integration in eq 57. However, before discussing the various input options in these sections, we provide a quick overview of the specifics of the DFPT-based implementation for computing the induced current density in CP2K.[172]

For a homogeneous magnetic field, the vector potential $\mathbf{A}$ can be chosen as

$$\mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{B} \times (\mathbf{r} - \mathbf{r}_0) \tag{58}$$

where $\mathbf{r}_0$ is an arbitrary gauge origin. Using the explicit form of the vector potential given in eq 58 leads to perturbations involving the position operator, which is problematic under PBC, as it jumps discontinuously at the edges of the cell. The solution suggested by Sebastiani and Parrinello is to use maximally localized Wannier functions as the zeroth-order orbitals $\psi^{(0)}(\mathbf{r})$.[168] If the simulation cell is chosen to be larger than the decay length of the Wannier functions (which is exponentially decaying for insulators) and choosing for each Wannier function ψ_i a coordinate system such that its center is at the origin (via a translation vector $\mathbf{d}_i$), one avoids the problematic behavior of the position operator at the cell boundaries. It should be noted that this choice of a different origin for the entire coordinate system for individual Wannier functions is not a gauge transformation, as both $\mathbf{r}$ and $\mathbf{r}_0$ in eq 58 are being simultaneously shifted. The current is invariant under arbitrary orbital-specific translations.

With the introduction of orbital-specific translations $\mathbf{d}_i$, one ends up with three perturbation operators.[168] The linear momentum operator

$$\hat{H}^P = \hat{\mathbf{p}} \tag{59}$$

the orbital angular momentum operator

$$\hat{H}^{Li} = (\hat{\mathbf{r}} - \mathbf{d}_i) \times \hat{\mathbf{p}} \tag{60}$$

and the full correction operator

$$\hat{H}^{\Delta i} = (\mathbf{d}_i - \mathbf{d}_j) \times \hat{\mathbf{p}} \tag{61}$$

where in the last equation $\mathbf{d}_i - \mathbf{d}_j$ is computed using the minimum image convention. The last two operators are labeled with the orbital index i to denote their orbital dependence via the chosen orbital centers $\mathbf{d}_i$. All three operators are vector operators, leading to nine contributions per orbital. Each contribution from the first two operators can be obtained essentially at the cost of one total energy calculation. The full correction operator requires one such calculation per orbital, making it by far the computationally most expensive term for any system with more than a handful of orbitals.

Once these perturbation operators are used in the inhomogeneous set of equations shown in eq 49, and all the contributions to the linear response orbitals have been obtained, the linear response current density vector is computed for the three directions of the magnetic field, allowing the computation of all the components of the shielding tensor. The x-component of the linear current density response induced by an external magnetic field applied along the y-axis is given by a sum of a paramagnetic $j_{xy}^p(\mathbf{r})$ and a diamagnetic $j_{xy}^d(\mathbf{r})$ contribution,[172] i.e.

$$j_{xy}^p(\mathbf{r}) = -\frac{1}{2c}\sum_{ikl}[C_{ki}^{(0)}(C_{li}^{L_y} + (\mathbf{r}_0 - \mathbf{d}_i)_x C_{li}^{P_z}$$
$$- (\mathbf{r}_0 - \mathbf{d}_i)_z C_{li}^{P_x} - C_{li}^{\Delta i_y})$$
$$\times \{(\nabla_x \chi_k(\mathbf{r}))\chi_l(\mathbf{r}) - \chi_k(\mathbf{r})\nabla_x \chi_l(\mathbf{r})\}] \tag{62}$$

$$j_{xy}^d(\mathbf{r}) = (\mathbf{r} - \mathbf{r}_0)_z \rho(\mathbf{r}) \tag{63}$$

and the other components are obtained analogously.

With our choice of the vector potential in eq 58, each of these two oppositely signed contributions to $\mathbf{j}(\mathbf{r})$ is linear in the gauge origin $\mathbf{r}_0$; their sum, however, should be invariant to it. With the local atomic basis functions commonly used in quantum chemistry, it is extremely difficult to converge the sum of the two terms at a reasonable computational cost, if one chooses a single fixed gauge origin, and the computed value of the current density becomes parametrically dependent on $\mathbf{r}_0$. This is the well-known gauge origin problem. To address this issue, a distributed gauge origin is used during the computation, with different methods making different choices regarding how this distribution is done. In CP2K the gauge origin is tackled using the individual gauge for atoms in molecules (IGAIM), where the gauge origin is the position of the nearest atom,[177] or alternatively, with the continuous set of gauge transformations (CSGT), where the gauge origin is the current position $\mathbf{r}$.[178] It is obvious that the CSGT approach makes the diamagnetic contribution in eq 63 vanish.

The choice of the gauge origin is set within the &CURRENT input section:

```
&LINRES
    &CURRENT
        GAUGE # the gauge origin r₀ in
              # Eqs. 58 & 62
        GAUGE_ATOM_RADIUS  # atomic radius
                       # in IGAIM gauge
        ORBITAL_CENTER # the choice of dᵢ
                       # in Eqs. 60-61
    &END CURRENT
&END LINRES
```

For the choice GAUGE R, one has the option to use CSGT for both the soft and local contributions (see eq 64 below) to the current density. GAUGE ATOM uses IGAIM centered on the nearest atom to the current grid point, and a combination of IGAIM for the hard part and CSGT for the soft part can be chosen with GAUGE R_AND_STEP_FUNCTION. For the IGAIM method, the radius of the atomic domain is set by the keyword GAUGE_ATOM_RADIUS. Although the choice of CSGT $\mathbf{r} = \mathbf{r}_0$ is appealing, as it makes the diamagnetic term in eq 63 identically vanish, both empirical evidence and theoretical considerations indicate that other choices provide better results.[179]

With the keyword ORBITAL_CENTER, one can also control the choice of the orbital centers $\mathbf{d}_i$ used in the perturbation operators eqs 60 and 61. The choice WANNIER uses the position of the respective Wannier center and is the most expensive, requiring $3M + 6$ response calculations, where M is the number of MOs. All other choices offer means to cluster orbital centers together, potentially leading to a substantial reduction in the required computational effort. At variance, ATOM uses the atomic positions as centers, hence the Wannier functions are clustered on the nearest atom, whereas BOX divides the simulation cell into sub-boxes and clusters the centers within each sub-box (the keyword NBOX sets the number of boxes along each dimension). Finally, COMMON uses a common center, whose position is specified by the keyword COMMON_CENTER, which only works for isolated molecules.

With the keywords

```
SELECTED_STATES_ON_ATOM_LIST
SELECTED_STATES_ATOM_RADIUS
```

one can further choose to perform the response calculation using only the Wannier functions that are within a certain distance from some chosen list of atoms. If only the magnetic

shieldings of some particular atoms are of interest, this can lead to enormous reductions in computational time. Naturally, one needs to ascertain that the selected radius provides acceptable results for the atoms of interest, which depends on the respective chemical environments.

Finally, the calculation of the magnetic shielding tensor requires the evaluation of the integral in eq 57. For a periodic system, the current density is also periodic and the computation can be done efficiently using the GAPW method (Section 2). The current density is decomposed in a manner analogous to the GAPW electron density, thus

$$
\mathbf{j}(\mathbf{r}) = \tilde{\mathbf{j}}(\mathbf{r}) + \sum_A^{\text{atoms}} (\mathbf{j}(\mathbf{r}) - \widetilde{\mathbf{j}_A}(\mathbf{r}))
\tag{64}
$$

where $\tilde{\mathbf{j}}$ is the soft contribution to the current density, $\mathbf{j}_A$ is the local hard contribution at atom A, and $\widetilde{\mathbf{j}_A}$ is the local soft contribution to prevent double counting.

The soft contribution $\tilde{\mathbf{j}}$ is computed in reciprocal space on the PW grid.[168] One has to note here, however, that the $\mathbf{G} = 0$ component cannot be computed under PBC. This term is commonly called the susceptibility correction and depends on the bulk magnetic susceptibility. It is determined by macroscopic magnetostatics from the macroscopic shape of the sample. In CP2K this term is calculated from the susceptibility due to $\tilde{\mathbf{j}}$, assuming a spherical sample shape, which is also the experimental convention[180]

$$
\chi_{xy} = \frac{2\pi}{\Omega_c c} \int d\mathbf{r} \, [\mathbf{r} \times \tilde{\mathbf{j}}_x(\mathbf{r})]_y
\tag{65}
$$

For the local part of the current density, the contribution to σ_{xy} of the nucleus at position $\mathbf{r}^N$, arising from the induced local current densities, is evaluated as

$$
\sigma_{xy}(\mathbf{r}^N) = \frac{1}{c} \sum_B \int_{\Omega_B} d\mathbf{r} \left[\frac{\mathbf{r} - \mathbf{r}^N}{|\mathbf{r} - \mathbf{r}^N|^3} \times (j_{x,B}(\mathbf{r}) - \tilde{j}_{x,B}(\mathbf{r})) \right]_y
\tag{66}
$$

where the sum over atoms B is restricted to the nuclei that are within a radius R_c from $\mathbf{r}^N$ (see keyword &NMR%SHIFT_-GAPW_RADIUS below). The integration over the atomic domain Ω_B is performed numerically on a spherical grid with a logarithmic radial and a Lebedev angular grid. The numerical integration converges rapidly with respect to the number of grid points, and about 10000 grid points per atom are enough to converge the chemical shift below 0.1 ppm.[172] The size of the atomic integration grid is set inside the &KIND subsection, and for the computation of the shielding tensors it can be beneficial to use values that are higher than the defaults, e.g.

```
&SUBSYS
    &KIND H
        BASIS_SET aug-pcSeg-2
        POTENTIAL ALL
        LEBEDEV_GRID 100
        RADIAL_GRID 200
    &END KIND
&END SUBSYS
```

In the &NMR input section, one can set the value of R_c, request a NICS calculation, and specify the spatial points for the latter:

```
&LOCALIZE
    &NMR
        NICS .TRUE.
        NICS_FILE_NAME nics.xyz
        SHIFT_GAPW_RADIUS # Cutoff current
                          # integration
    &END NMR
&END LOCALIZE
```

5.3. The Electron Paramagnetic Resonance g-Tensor

CP2K also offers an implementation of the g-tensor under PBC, which employs the same DFPT approach to compute the induced current densities. This makes it feasible to perform large-scale calculations, including with QM/MM (e.g., paramagnetic active sites in enzymes), to investigate paramagnetic defects in solids under PBC, and also to perform large-scale computations of paramagnetic NMR shifts.[181]

Effectively, the g-tensor plays in EPR spectroscopy a role similar to that played by the shielding tensor in NMR spectroscopy. This is clear in an effective spin-Hamiltonian framework, which gives the coupling between an effective electronic spin $\mathbf{S}$ and an external magnetic field $\mathbf{B}^{\text{ext}}$ as the bilinear term

$$
\hat{H}_{\text{eff}} = \frac{1}{2c} \mathbf{B}^{\text{ext}} \cdot \mathbf{g} \cdot \mathbf{S}
\tag{67}
$$

and like the shielding tensor, the g-tensor is also a second partial derivative of the energy with respect to the external field and a spin magnetic moment, in this case the electronic spin. The corresponding ab initio expression is obtained from the minimal coupling electronic Hamiltonian including relativistic corrections up to $O(\alpha^3)$, where α is the fine structure constant. Identification of the relevant Hamiltonian terms produces the expression for the g-tensor, in the form of spatially anisotropic relativistic correction terms $\Delta \boldsymbol{g}$ to the free electron g-value g_e. The implementation in CP2K is based on the Schreckenbach–Ziegler DFT-based approach,[182] and its extension by Pickard and Mauri.[183,184] One can identify three contributions to each component of $\Delta \boldsymbol{g}$, i.e.

$$
g_{xy} = g_e \delta_{xy} + \Delta g_{xy}^{\text{ZKE}} + \Delta g_{xy}^{\text{SO}} + \Delta g_{xy}^{\text{SOO}}
\tag{68}
$$

The correction terms are the Zeeman kinetic energy (ZKE) term

$$
\Delta g_{xy}^{\text{ZKE}} = -\frac{g_e}{c^2}(T^\alpha - T^\beta)\delta_{xy}
\tag{69}
$$

which is a purely kinematic scalar relativistic correction, the spin–orbit (SO) term (usually the dominant term)

$$
\Delta g_{xy}^{\text{SO}} = \frac{(g_e - 1)}{c} \int_{\Omega_c} d\mathbf{r} [\mathbf{j}_x^\alpha(\mathbf{r}) \times \nabla V_{\text{eff}}^\alpha(\mathbf{r})
$$
$$
- \mathbf{j}_x^\beta(\mathbf{r}) \times \nabla V_{\text{eff}}^\beta(\mathbf{r})]_y
\tag{70}
$$

and the spin-other-orbit (SOO) term

$$
\Delta g_{xy}^{\text{SOO}} = \frac{1}{S} \int_{\Omega_c} d\mathbf{r} \, B_{y,B_x}^{\text{corr}}(\mathbf{r})[\rho^\alpha(\mathbf{r}) - \rho^\beta(\mathbf{r})]
\tag{71}
$$

where α and β denote the spin channels, S the total spin, and T is the kinetic energy. Moreover, B_{y,B_x}^{corr} is the y-component of the magnetic field due to the total induced current density by a field (see eq 52) in the x-direction minus a self-interaction correction

$$B_{y,B_x}^{corr}(\mathbf{r}) = \frac{1}{c} \int_\Omega d\mathbf{r}' \left[\frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} \times (\mathbf{j}_x(\mathbf{r}') - \mathbf{j}_x^{\alpha-\beta}(\mathbf{r}')) \right]_y \tag{72}$$

with $\mathbf{j}^{\alpha-\beta} = \mathbf{j}^\alpha - \mathbf{j}^\beta$. The effective potential V_{eff} in the SO contribution is the sum of the external potential, the Hartree potential, and the XC potential.[182] Currently, the CP2K implementation only supports LDA and GGA functionals; hybrid XC functionals are not yet supported.

The ZKE contribution is computed from the kinetic energy of the spin-polarized KS orbitals in the Gaussian basis set. The SO and SOO terms are similar to the shielding tensor in that they require the (spin-dependent) current density. For the latter, the same techniques based on the aforementioned Sebastiani−Parrinello approach,[168] discussed in the context of the magnetic shielding tensor section, are also employed here. One subtlety with respect to the SOO term is that it requires the induced field over all space, not only at the positions of the nuclei. The nonlocal nature of the dipolar term in eq 52 makes it very taxing to create a GAPW representation of the induced magnetic field. Hence, the implementation employs an approximation of the SOO term, whereby the contributions from atom-centered (local) current densities to the $\mathbf{G} \neq 0$ components of the induced field are ignored.

The g-tensor is generally much less sensitive to the choice of gauge than nuclear magnetic shieldings, and even the computationally convenient CSGT (`&CURRENT%GAUGE R`) can be used. Tables 1 and 2 in ref 181 offer a comparison of g-tensors calculated by CP2K with the IGAIM gauge versus those calculated by ORCA[185] and MAG-ReSpect[186] using a variety of gauge origins and spin−orbit treatments.

It is important to note that if the simulation cell contains more than one paramagnetic center, the individual center g-tensors erroneously add up. This is a problem in existing implementations of g-tensors in condensed phases.[181,187] Hence, in this case one needs to normalize the g-tensor of the simulation cell

$$\mathbf{g} = g_e \mathbf{1} + \frac{1}{n} \Delta \mathbf{g} \tag{73}$$

with n being the number of paramagnetic centers in the cell.

To activate the calculation of the g-tensor in the input file, one has to request the spin density from an unrestricted KS calculation (keyword `&DFT%UKS`). One also needs to include the `&CURRENT`, `&LOCALIZE`, and `&EPR` subsections within the `&LINRES` section. The various control parameters related to the calculation of the induced current density, in the input section `&CURRENT`, were already discussed for the magnetic shielding tensors. The input subsection

`&EPR%PRINT%G_TENSOR%XC%XC_FUNCTIONAL`

within the `&LINRES` section sets the functional for the effective XC potential in eq 70. Other XC potentials are activated in a similar fashion, for example `XALPHA` corresponds to the Dirac−Slater potential, whereas `BECKE88` together with `LYP` will use the BLYP XC potential.

5.4. Hyperfine Couplings

The effective spin-Hamiltonian term for the hyperfine coupling is bilinear in the nuclear and electron spin

$$\hat{H} = \mathbf{S} \cdot \mathbf{A} \cdot \mathbf{I} \tag{74}$$

The ab initio expression for the hyperfine coupling tensor, or **A**-tensor, is obtained from relativistic quantum mechanics, as

in the case of the g-tensor. For any nucleus N, the dominant terms are an isotropic (Fermi contact) term

$$A_{iso,N} = \frac{4\pi}{3} \frac{g_e \mu_e g_N \mu_N}{\langle S_z \rangle} \int d\mathbf{r}\, \rho^{\alpha-\beta}(\mathbf{r}) \delta_T(\mathbf{r}) \tag{75}$$

and an anisotropic dipolar term

$$A_{ani,N} = \frac{1}{2} \frac{g_e \mu_e g_N \mu_N}{\langle S_z \rangle} \times \int d\mathbf{r}\, \rho^{\alpha-\beta}(\mathbf{r}) \frac{3 r_i r_j - \delta_{ij} r^2}{r^5} \tag{76}$$

where μ_e is the Bohr magneton, μ_N the nuclear gyromagnetic ratio, μ_N the nuclear magneton, $\langle S_z \rangle$ is the expectation value of the z-component of total electronic spin, and $\mathbf{r}$ is taken relative to the position of the nucleus N. Moreover, $\rho^{\alpha-\beta}(\mathbf{r})$ and g_e have already been introduced, whereas δ_T is a smeared-out delta function which results from scalar relativistic corrections (in the nonrelativistic limit, it collapses to a Dirac delta function), i.e.

$$\delta_T(\mathbf{r}) = \frac{1}{4\pi r^2} \frac{2}{Z\alpha^2} \frac{1}{\left(1 + \frac{2r}{Z\alpha^2}\right)^2} \tag{77}$$

where Z is the nuclear charge and α is the fine structure constant. Equations 75−77 are the basis for computing hyperfine couplings in CP2K.[188] In addition to these two first-order terms, there is also a second-order spin−orbit contribution to the hyperfine coupling, which can become important for heavy atoms.

It is obvious that the isotropic term particularly requires an accurate description of the electron density at the position of the nucleus; hence, hyperfine couplings are implemented within CP2K's GAPW method.[188] The Fermi contact term at atom N is simply computed by integrating eq 75 inside the local atomic domain U_N using the hard spin density (for detailed definitions of these quantities, see ref 25). The anisotropic hyperfine coupling is again split into a soft contribution to be computed in the PW basis and local spin-density contributions. The contribution of the local spin density of atom N itself is integrated inside the domain U_N, since the total local density of any atom outside its domain is, by definition, zero. The local densities of other atoms $M \neq N$ make a small contribution to the hyperfine coupling of atom N (see Table 4 in ref 188). These "cross terms" needs to be integrated into the respective domains U_M. For this small contribution, only near-neighboring atoms around N need to be included, using a cutoff distance with a default value of 10 Bohr, which the user can modify.

Unlike the magnetic properties discussed so far, the hyperfine coupling in terms of eqs 75−77 is a first-order property, obtained as an expectation value over the unperturbed spin density. Therefore, the input section for requesting the printing of the hyperfine coupling tensors is found directly under the DFT section via

`&DFT%PRINT%HYPERFINE_COUPLING_TENSOR`

In addition to various print control options, the `INTER-ACTION_RADIUS` can be set inside this input section, which specifies the cutoff radius mentioned above.

6. OPTICAL SPECTROSCOPY

Optical spectroscopy is a technique that is used to study the interaction between light and matter. It involves measuring the absorption, emission, or scattering of light by molecules,

atoms, or materials. The resulting spectra provide valuable information about the electronic structure, energy levels, and dynamics of the system under investigation. One of the key quantities of interest is the excitation energy $\Omega^{(n)}$, which is the energy required to excite a molecule from its ground state to an excited state. These excitation energies are directly related to the positions and intensities of spectral lines observed in absorption and emission spectra.

6.1. Linear-Response Time-Dependent Density Functional Theory

Optical properties are calculated by solving a generalized eigenvalue problem that involves the block matrix $ABBA$,[189] hence

$$
\begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(n)} \\ \mathbf{Y}^{(n)} \end{pmatrix} = \Omega^{(n)} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(n)} \\ \mathbf{Y}^{(n)} \end{pmatrix} \tag{78}
$$

Therein, A and B are matrices with indices $A_{ia,jb}$, i.e. they have $N_{occ}N_{empty}$ rows and $N_{occ}N_{empty}$ columns. The eigenvector $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$ has elements $X_{ia}^{(n)}$ and $Y_{ia}^{(n)}$, which quantify the contribution of the transition from occupied orbital i to an empty orbital a in the excitation n. The matrix elements of $\mathbf{A}$ and $\mathbf{B}$ are

$$
A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) + (ia|f_{xc}|jb)
$$
$$
B_{ia,jb} = (ia|bj) + (ia|f_{xc}|bj) \tag{79}
$$

respectively.

The Hermitian matrix $\mathbf{A}$ includes, as a zeroth-order term, the differences within the KS orbital energies. At first order, it incorporates kernel contributions, which vary depending on the chosen density functional approximation. These contributions consist of Coulomb and exact exchange terms, as well as components arising from the XC potential and its associated kernel f_{xc}. The implementation in CP2K/QUICKSTEP is based on the Tamm–Dancoff approximation (TDA),[190,191] which involves constraining $\mathbf{B} = 0$ and $\mathbf{Y} = 0$ in eq 79, thus resulting in the Hermitian eigenvalue problem

$$
A\mathbf{X} = \Omega\mathbf{X} \tag{80}
$$

The TDDFT module facilitates the calculation of excitation energies and excited state properties within the TDA, supporting both GGA and hybrid XC functionals, as well as simplified semiempirical TDA kernels. The optical spectrum calculation with TDDFT requires setting `&GLOBAL%RUN_TYPE ENERGY` and specifying the subsequent `&TDDFPT` subsection within `&FORCE_EVAL%PROPERTIES` as follows:

```
&TDDFPT
    KERNEL FULL
    NSTATES 10
    MAX_ITER 50
    CONVERGENCE [eV] 1.0E-6
    RKS_TRIPLETS F
&END TDDFPT
```

The kernel matrix is controlled by the `KERNEL` keyword, where one can select between the full kernel that is appropriate for GGA or hybrid functional computations, or the semiempirical `sTDA` kernel.[192,193] If the `FULL` kernel is selected, the underlying functional can be specified via the `&XC` section. The choice of XC functional can differ from the ground state functional unless the `ADMM` method is employed. The `sTDA`

kernel omits XC contributions and approximates both Coulomb and exchange terms using pairwise semiempirical operators γ^J and γ^K, which are dependent on the interatomic distance R_{MN} between atoms M and N. Hence

$$
\gamma^J(M, N) = \left(\frac{1}{R_{MN}{}^\alpha + \eta^{-\alpha}} \right)^{1/\alpha}
$$
$$
\gamma^K(M, N) = \left(\frac{1}{R_{MN}{}^\beta + (a_{XC}\eta)^{-\beta}} \right)^{1/\beta} \tag{81}
$$

are relying on the four global parameters, the chemical hardness η, the Fock-exchange mixing parameter a_{XC} and the powers α for Coulomb, and β for the exchange interactions. The proportion of exact exchange, controlled by adjusting the parameter a_{XC}, is critical for a well-balanced treatment of exact-exchange in the ground and excited state potential surfaces. This is controlled by the keyword `FRACTION` in the `&sTDA` section, and it is recommended to use a relatively small amount, e.g. $a_{XC} = 0.1-0.2$. The parameters α and β are controlled by the `MATAGA_NISHIMOTO_CEXP` and `MATAGA_NISHIMOTO_XEXP` keywords, respectively.

```
KERNEL sTDA
&sTDA
    FRACTION 0.2
    DO_EWALD TRUE
&END sTDA
```

It should be noted that the definition of the operators in eq 81 differs from the original method for compatibility with the ADMM definition. For periodic systems, the Ewald summation for Coulomb contributions should be enabled with the `DO_EWALD` keyword, which also activates the `&DFT%POISSON` section.

Additional key parameters include `NSTATES`, which defines the number of excitation energies to compute, and `CONVERGENCE` that sets the convergence threshold for the Davidson algorithm. The `RKS_TRIPLETS` option allows switching from the default singlet excitation energy calculations to triplet excitations. The `RESTART` keyword enables restarting a TDDFT calculation if a valid restart file (`.tdwfn`) is available. Furthermore, a separate grid can be defined for real-space integration in the TDDFT calculation via the `&TDDFPT%MGRID` subsection, though a consistent setup for ground and excited-state calculations is generally recommended.

Upon convergence of the Davidson algorithm, CP2K outputs the excitation energies in eV, the transition dipole moments, and the oscillator strengths for each calculated excited state. The `&DIPOLE_MOMENTS%DIPOLE_FORM` keyword allows modifying the form of the dipole transition integrals, with options such as `BERRY` (for fully periodic systems), `LENGTH` (for molecular systems), and `VELOCITY` (for both). When using the length form, the reference point for calculating dipole moments can be adjusted through the `REFERENCE` keyword, allowing for selection between `COM` (center of mass), `COAC` (center of atomic charges), `USER_DEFINED` (user-specified coordinates), or `ZERO` (origin of the coordinate system). If `USER_DEFINED` is selected, the keyword `REFERENCE_POINT` must be defined in terms of the atomic coordinates.

Natural transition orbitals (NTOs) can be printed by setting `PRINT_LEVEL` to medium in the `&GLOBAL` section or by

enabling it in `&PRINT%NTO_ANALYSIS`. For an NTO analysis, unoccupied orbitals must be generated, whose number included in the analysis is adjusted using the `LUMO` keyword. One can specify a list of states for which NTOs should be printed using the `STATE_LIST` keyword or apply a threshold for the screening of the states based on their oscillator strengths with the `INTENSITY_THRESHOLD` keyword. NTOs can be outputted as `CUBE_FILES` or in Molden format, the latter being controlled by the `MOS_-MOLDEN` keyword.

The excited-state gradients can be calculated by setting `RUN_TYPE ENERGY_FORCE` in the `&GLOBAL` section. The specific excited state for which the gradients will be computed must be defined by adding the `&EXCITED_-STATES` subsection within the `&DFT` section, e.g.:

```
&DFT
    &EXCITED_STATES
        STATE 1
    &END EXCITED_STATES
&END DFT
```

The parameters for the gradient calculation can be adjusted in the `&TDDFPT%LINRES` subsection, allowing to specify the `PRECONDITIONER` for the optimization and `MAX_ITER` for the maximum number of iterations. Similarly, the fluorescence energy can be calculated by optimizing the first excited state. This is achieved by setting `&GLOBAL%RUN_TYPE GEO_OPT` and defining the `&MOTION%GEO_-OPT` subsection, and specifying the reference excited state as outlined above. Additionally, support for excited-state gradient calculations using the GAPW method has recently been introduced.[194]

Furthermore, SOC can be taken into account by adding a perturbative correction via the `&PROPERTIES%BAND-STRUCTURE%SOC` subsection.[195] As triggered by the keyword `RKS_TRIPLETS`, the TDA eigenvalue problem is thereby solved for both singlet and triplet closed-shell references, yielding the corresponding excited state manifolds with excitation amplitudes $\mathbf{X}_{TDA}^{(n)}$, and thereon-based, auxiliary many-electron WFs are generated, with the manifolds $|\Psi_{S,m_S}^{(n)}\rangle$ being here labeled according to the total spin of the system S and the corresponding spin angular momentum quantum number m_S. Using quasi-degenerate perturbation theory, the unperturbed excitation energies $\Omega_{S,m_S}^{(n)}$, as obtained from eq 78, are corrected by adding an additive correction based on a one-electron SOC Hamiltonian $\hat{h}^{SO}$, so that

$$H_{S,m_S,S',m_S'}^{(nm)} = \delta_{(nm)}\delta_{SS'}\delta_{m_Sm_S'}\Omega_{S,m_S}^{(n)} \quad (82)$$

$$+ \langle\Psi_{S,m_S}^{(n)}|\hat{h}^{SO}|\Psi_{S',m_S'}^{(m)}\rangle \quad (83)$$

Two choices are available for $\hat{h}^{SO}$, however, for all-electron GAPW computations, van Wüllen's model potential based on the zeroth-order regular approximation (ZORA) is chosen by default, requiring no change in the CP2K input.[196,197] Restricting the description to valence electrons, a SOC-corrected PP can be used,[32] as described in Subsection 3, by specifying `GTH_SOC_POTENTIALS` as the `&DFT%POTEN-TIAL_FILE_NAME` in the CP2K input file. To print the resulting SOC-corrected energies, spin−orbit matrix elements as defined in eq 83, and oscillator strengths, the following

`&SOC_PRINT` subsection must be added to the `&TDDFPT%PRINT` section of TDDFT:

```
&PRINT
    &SOC_PRINT
        SOME
        SPLITTING
    &END SOC_PRINT
&END PRINT
```

Adding the keywords `UNIT_EV` or `UNIT_WN` furthermore determines whether the chosen unit of the printout is in eV or cm^{-1}, respectively.

6.2. Bethe−Salpeter Equation

The BSE is a method for computing electronic excitation energies $\Omega^{(n)}$, $n = 1, 2, ...$ and optical absorption spectra.[198−203]

6.2.1. Electronic Excitation Energies. For a closed-shell system, the entries of A and B of eq 78 are

$$A_{ia,jb} = (\varepsilon_a^{GW} - \varepsilon_i^{GW})\delta_{ij}\delta_{ab} + \alpha^{(S/T)}v_{ia,jb} - W_{ij,ab}$$

$$B_{ia,jb} = \alpha^{(S/T)}v_{ia,bj} - W_{ib,aj} \quad (84)$$

where δ_{ij} is the Kronecker delta, $v_{pq,rs}$ the bare Coulomb interaction and $W_{pq,rs}$ the statically screened Coulomb interaction, with p, q, r, s being KS orbital indices. The user needs to set $\alpha^S = 2$ for computing singlet excitations and $\alpha^T = 0$ for computing triplet excitations.[198,204]

A BSE calculation requires occupied KS orbitals $\varphi_i(\mathbf{r})$ and empty KS orbitals $\varphi_a(\mathbf{r})$ from a DFT calculation, where $i = 1, ..., N_{occ}$ and $a = N_{occ} + 1, ..., N_{occ} + N_{empty}$, as well as the GW eigenvalues ε_i^{GW} and ε_a^{GW}, respectively. In CP2K/QUICKSTEP, it is possible to use G_0W_0, evGW_0 or evGW eigenvalues, see details in the GW section and in ref 119. Thus, also GW and DFT input parameters might have an impact on BSE excitation energies $\Omega^{(n)}$. We recommend calculating the BSE based on evGW_0@PBE eigenvalues,[35] as discussed in ref 125.

The TDA constrains $B = 0$, such that $\Omega_{TDA}^{(n)}$ and $\mathbf{X}_{TDA}^{(n)}$ can be computed from the Hermitian eigenvalue problem

$$A \mathbf{X}_{TDA}^{(n)} = \Omega_{TDA}^{(n)}\mathbf{X}_{TDA}^{(n)} \quad (85)$$

Diagonalizing A in the TDA of eq 85, or the full block-matrix $ABBA$ of eq 78 takes on the order of $(N_{occ}N_{empty})^3$ floating-point operations. This translates into a computational scaling of $O(N^6)$ for a BSE calculation of system size N.

The eigenvectors $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$ of eq 78 with elements $X_{ia}^{(n)}$ and $Y_{ia}^{(n)}$ enter the WF of the electronic excitation[203,205]

$$\Psi_{excitation}^{(n)}(\mathbf{r}_e, \mathbf{r}_h) = \sum_{ia} X_{ia}^{(n)}\varphi_i(\mathbf{r}_h)\varphi_a(\mathbf{r}_e) + Y_{ia}^{(n)}\varphi_i(\mathbf{r}_e)\varphi_a(\mathbf{r}_h) \quad (86)$$

i.e. $X_{ia}^{(n)}$ and $Y_{ia}^{(n)}$ describe the transition amplitude between an occupied orbital φ_i and an empty orbital φ_a for the n-th excitation. In TDA, we can interpret the excitation in terms of electrons and holes as follows: an electron leaves a hole behind in the (formerly) occupied orbital $\varphi_i(\mathbf{r}_h)$ and is afterward located in the empty orbital $\varphi_a(\mathbf{r}_e)$.

6.2.2. Optical Absorption Spectrum. Within the BSE framework, we can compute the photoabsorption cross-section tensor $\sigma_{\mu,\mu'}(\omega)$ with $(\mu, \mu' \in \{x, y, z\})$ as[206,207]

$$\sigma_{\mu,\mu'}(\omega) = -\frac{4\pi\omega}{c}\text{Im}\left[\sum_n \frac{2\Omega^{(n)}d_\mu^{(n)}d_{\mu'}^{(n)}}{(\omega + i\eta)^2 - (\Omega^{(n)})^2}\right] \quad (87)$$

with broadening η and the transition moments of the Singlet solution given by[202]

$$d_\mu^{(n)} = \sqrt{2} \sum_{i,a} \langle \varphi_i | \hat{\mu} | \varphi_a \rangle (X_{ia}^{(n)} + Y_{ia}^{(n)}) \tag{88}$$

When the molecules are not aligned, e.g. for gas phase and liquids, we can compute the photoabsorption cross section from the spatial average of eq 87

$$\overline{\sigma}(\omega) = \frac{1}{3} \sum_{\mu \in \{x,y,z\}} \sigma_{\mu,\mu}(\omega)$$

$$= -\frac{4\pi\omega}{c} \text{Im} \left[\sum_n \frac{f^{(n)}}{(\omega + i\eta)^2 - (\Omega^{(n)})^2} \right] \tag{89}$$

where we have introduced the oscillator strengths[202]

$$f^{(n)} = \frac{2}{3} \Omega^{(n)} \sum_{\mu \in \{x,y,z\}} |d_\mu^{(n)}|^2 \tag{90}$$

6.2.3. Measures for the Size of an Excited State. The WF of an excited state $\Psi_{\text{excitation}}^{(n)}(\mathbf{r}_e, \mathbf{r}_h)$ of eq 86 and its spatial properties can be analyzed by introducing measures for its spatial extent following ref 205. To that end, we define the corresponding expectation value for a generic operator

$$\langle \hat{O} \rangle_{\text{exc}}^{(n)} = \frac{\langle \Psi_{\text{excitation}}^{(n)} | \hat{O} | \Psi_{\text{excitation}}^{(n)} \rangle}{\langle \Psi_{\text{excitation}}^{(n)} | \Psi_{\text{excitation}}^{(n)} \rangle} \tag{91}$$

For example, we can compute the distance between the electron and the hole as

$$d_{\text{h} \to \text{e}} = |\langle \mathbf{r}_h - \mathbf{r}_e \rangle_{\text{exc}}| \tag{92}$$

This can be used to characterize a charge-transfer state, where the electron and hole sit on different parts of the molecule and therefore $d_{\text{h} \to \text{e}}$ is nonvanishing.

Further, we can compute the exciton size

$$d_{\text{exc}} = \sqrt{\langle |\mathbf{r}_h - \mathbf{r}_e|^2 \rangle_{\text{exc}}} \tag{93}$$

which quantifies the spatial extent of the combined electron–hole pair. It increases when the distance between the electron and hole $d_{\text{h} \to \text{e}}$, or their respective sizes, increases.

A typical BSE calculation starts with RUN_TYPE ENERGY and the following prototypical *GW*-BSE section:

```
&GW
    SELF_CONSISTENCY        evGW0
    &BSE
        TDA                 TDA+ABBA
        SPIN_CONFIG         SINGLET
        NUM_PRINT_EXC       15
        ENERGY_CUTOFF_OCC   -1
        ENERGY_CUTOFF_EMPTY -1
        NUM_PRINT_EXC_DESCR -1
        &BSE_SPECTRUM
            ETA_LIST 0.01 0.02
        &END BSE_SPECTRUM
    &END BSE
&END GW
```

Therein, the following keywords are available:

- SELF_CONSISTENCY: determines which *GW* self-consistency (G_0W_0, $evGW_0$ or $evGW$) is used to calculate the single-particle *GW* energies ε_p^{GW} necessary

in the BSE calculation of eq 84. We recommend using evGW0 for BSE runs.[125]

- TDA: specifies if the TDA and/or diagonalization of the full ABBA-matrix is employed. OFF: Generalized diagonalization of *ABBA* of eq 78. ON: use the TDA of eq 85 and diagonalize only *A*. TDA+ABBA: compute excitation energies $\Omega^{(n)}$ and $\Omega_{\text{TDA}}^{(n)}$ from eqs 78 and 85, respectively.

- SPIN_CONFIG: choose between SINGLET for computing singlet excitation energies ($\alpha^S = 2$) and TRIPLET for computing triplet excitation energies ($\alpha^T = 0$). The standard is SINGLET as an electronic excitation directly after photoexcitation is a singlet due to angular momentum conservation; triplet excited states can form by intersystem crossings.

- NUM_PRINT_EXC: number of excitation energies $\Omega^{(n)}$ to be printed.

- ENERGY_CUTOFF_OCC (in eV): only use indices *i* of occupied MOs in the interval $\varepsilon_i^{\text{DFT}} \in [\varepsilon_{\text{HOMO}}^{\text{DFT}} - \text{ENERGY\_CUTOFF\_OCC}, \varepsilon_{\text{HOMO}}^{\text{DFT}}]$ to set up the matrices *A* and *B* in eq 84. A small ENERGY_CUTOFF_OCC reduces computation time and memory consumption, but can affect the computed excitation energies $\Omega^{(n)}$. Usage of ENERGY_CUTOFF_OCC is recommended for molecules with more than 30 atoms. We recommend a convergence test by increasing ENERGY_CUTOFF_OCC and observing the effect on $\Omega^{(n)}$.[202]

- ENERGY_CUTOFF_EMPTY (in eV): analogously to ENERGY_CUTOFF_OCC, but for the empty states, i.e. only empty states in the interval $\varepsilon_i^{\text{DFT}} \in [\varepsilon_{\text{LUMO}}^{\text{DFT}}, \varepsilon_{\text{LUMO}}^{\text{DFT}} + \text{ENERGY\_CUTOFF\_EMPTY}]$ are used in the BSE calculation.

- NUM_PRINT_EXC_DESCR: number of excitations *n*, for which the exciton descriptors are printed, e.g. $d_{\text{exc}}^{(n)}$ of eq 93.

- BSE_SPECTRUM: Activates the computation and printing of the photoabsorption cross section tensor, i.e. its spatial average $\overline{\sigma}(\omega)$ of eq 89 and its elements $\sigma_{\mu,\mu'}(\omega)$ ($\mu, \mu' \in \{x, y, z\}$) of eq 87 are printed.

In addition to the upper *GW*-BSE keywords, the settings within the &DFT section influence the BSE excitation energies $\Omega^{(n)}$:

- XC_FUNCTIONAL: choose between one of the available XC functionals. The starting point can have a profound influence on the excitation energies.[208] Motivated by the discussion in ref 125, we strongly recommend to use BSE@$evGW_0$@PBE, i.e. the PBE functional[35] as the DFT starting point.

- BASIS_SET: the all-electron aug-cc-pVDZ basis set[44,209] should be sufficient for most organic molecules, but needs to be checked with respect to convergence.

The memory consumption of the BSE method is large, i.e. approximately $100 \, N_{\text{occ}}^2 N_{\text{empty}}^2$ Bytes. The estimated memory consumption, N_{occ}, as well as N_{empty} are all printed in the BSE output. The BSE implementation is very well parallelized, i.e. one can use several nodes that can also provide the required memory. We have compared the excitation energies from the BSE implementation in CP2K against the FHI-aims code[202] and found excellent agreement on the order of 5 meV mean absolute error for the first 10 excitations of Thiel's data set.[210]

7. EXCITED STATE DYNAMICS

In RT-TDDFT, instead of solving the static SE, the aim is to solve the time-dependent SE

$$i\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \hat{H}(\mathbf{r}, t)\Psi(\mathbf{r}, t) \tag{94}$$

Contrary to the time-independent case, there is no variational principle for the total energy, and the total energy has to be replaced by the quantum mechanical action

$$A[\Psi] = \int_{t_0}^{t_f} dt \left\langle \Psi(t) \left| i\frac{\partial}{\partial t} - \hat{H}(t) \right| \Psi(t) \right\rangle \tag{95}$$

for which the function $\Psi(t)$ that makes the action stationary will be its solution.

7.1. Real-Time Propagation and Ehrenfest Dynamics

The resulting time-dependent KS equations read as

$$i\frac{\partial}{\partial t}\psi_i(\mathbf{r}, t) = \left[-\frac{\nabla^2}{2} + v_{KS}(\mathbf{r}, t) + \mathbf{F}(t) \right] \psi_i(\mathbf{r}, t) \tag{96}$$

with ψ_i being the KS orbitals

$$\rho(\mathbf{r}, t) = \sum_i f_i |\psi_i(\mathbf{r}, t)|^2 \tag{97}$$

the time-dependent density, and $\mathbf{F}(t)$ an optional time-dependent external field. By applying an external electric field, we intend to simulate the interaction between the electrons and an external radiation. The light is then treated classically using either the length-gauge for nonperiodic systems or the velocity-gauge for periodic systems.

A rather general derivation for (nonadiabatic quantum-classical) Ehrenfest dynamics can be obtained starting from the action of a system.[211] For a situation in which the electrons are treated quantum mechanically, while the nuclei are treated classically, the total action can be written as the sum of the two environments $A = A_c + A_q$, where

$$A_c = \int_{t_0}^{t_f} dt \left[\sum_A \frac{M_A}{2}\dot{\mathbf{R}}_A - U(\mathbf{R}, t) \right] \tag{98}$$

and A_q the quantum mechanical action as defined above. The equations of motion are then derived by making the action stationary

$$\frac{\delta A}{\delta\langle\Psi(\mathbf{r}, t)|} = 0 \tag{99a}$$

$$\frac{\delta A}{\delta\langle\mathbf{R}(t)|} = 0 \tag{99b}$$

Evaluating these expressions in the framework of TDDFT, the equations of motion become

$$M\ddot{\mathbf{R}} = -\frac{\partial}{\partial\mathbf{R}}U(\mathbf{R}, t) - \sum_j \left\langle \Psi^j \left| \frac{\partial}{\partial\mathbf{R}}V_{int}(\mathbf{r}, \mathbf{R}) \right| \Psi^j \right\rangle \tag{100}$$

for the nuclear motion, while the time-dependent Schrödinger equation as given above is used for the electrons. Yet, in a numerical calculation, the WFs are often replaced by a finite basis set representation in terms of the LCAO. For PWs, the equations remain the same, since they do not depend on the nuclear coordinates. Using Gaussian basis functions for the

expansion of the WFs, however, introduces an implicit dependence of the WFs on the nuclear positions. Since in AIMD the nuclear coordinates are a function of time, the time-derivative in the quantum mechanical action has to be replaced by the total time-derivative

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_A \frac{\partial\mathbf{R}_A}{\partial t}\frac{\partial}{\partial\mathbf{R}_A} \tag{101}$$

Due to the introduction of a finite basis, the independent variables for making the action constant become the expansion coefficients $C_{j\alpha}(t)$ of the MO j in the contracted basis function $\phi_\alpha(\mathbf{r}, t)$, i.e.

$$\dot{C}_{j\alpha} = -\sum_{\alpha\beta} S_{\beta\gamma}^{-1}(iH_{\beta\gamma} + B_{\beta\gamma})C_{j\gamma}$$

where

$$S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle \qquad B_{\alpha\beta} = \left\langle \phi_\alpha \left| \frac{d}{dt}\phi_\beta \right. \right\rangle$$

Hence, in Ehrenfest dynamics an additional contribution due to the derivative of the basis functions becomes part of the Hamiltonian used in the exponential operator. Instead of being purely imaginary, the matrix in the exponential of the propagator becomes fully complex.

To run real-time propagation (RTP), where the nuclei are fixed and only the electronic degrees of freedom are propagated, or genuine Ehrenfest dynamics, the corresponding `RUN_TYPE` in the `&GLOBAL` section has to be set to `RT_PROPAGATION` or `EHRENFEST_DYN`, respectively. Furthermore, the `&MOTION%MD` section has to be present to specify the discretized `TIMESTEP` and the desired number of `STEPS`. It is crucial to set an appropriate time step to propagate the electronic degrees of freedom, which is much smaller than for MD and is typically on the order of attoseconds. All other input parameters related to the RT-TDDFT run are specified in the `&DFT%REAL_TIME_PRO-PAGATION` section. The simulation needs to start from the electronic density at t_0 specified by the keyword `INI-TIAL_WFN`. This can be either the ground state obtained by means of an initial SCF optimization denoted by `SCF_WFN`, or by providing the restart file of a previously computed WF via `RESTART_WFN`. Alternatively, the propagation can also be restarted from a RTP or Ehrenfest dynamics run by `RT_RESTART` and providing the correct restart file. Note that the RTP restart file has a format different from the SCF restart file. The path to the restart file is in both cases provided by the usual `&DFT%RESTART_FILE_NAME` keyword.

Three different propagators are available in CP2K, the enforced time-reversible symmetry propagator (ETRS), the exponential midpoint (EM) propagator, and the Crank-Nicholson propagator, which can be seen as a first-order Padé approximation of the EM propagator.[212] The ETRS approach starts with an exponential approximation to the evolution operator $\hat{U}(t, 0) = \exp\{-it\hat{H}(0)\}$, to then compute the final time-reversible and unitary propagator self-consistently. In the RTP scheme (fixed ionic positions), the self-consistent solution involves only the calculation of the new KS matrix for the propagated coefficients. For Ehrenfest dynamics, the iterative procedure is embedded in the integrator of the nuclear equations of motion. Hence, each iteration step for Ehrenfest dynamics involves an RTP step and the evaluation of

the nuclear forces. Due to this, more iterations are needed to reach self-consistency for the propagator. The convergence criterion implemented in CP2K is defined as

$$\|\Delta C^T S \Delta C\|_{\max} < \epsilon \tag{102}$$

with ΔC being the difference of coefficient matrices in two successive steps, and ϵ given by EPS_ITER.

In terms of computational cost, the most expensive part is the evaluation of the matrix exponential in the propagator. Four different methods, among those listed in ref 213, have been implemented in CP2K, i.e. diagonalization, the Taylor expansion, the Padé approximation, and the Arnoldi subspace iteration. The method is selected using the MAT_EXP keyword. The Arnoldi method often provides superior performance. Comparing the theoretical scaling, the Arnoldi method is expected to be about 5 times faster than the Padé or Taylor approaches. However, the Padé approximation can sometimes be a faster and more stable choice than the Arnoldi method (e.g., in the case of large integration timesteps). Since the propagation is based on an iterative procedure, it is convenient to apply an extrapolation scheme to speed up convergence.[11] For Ehrenfest dynamics, an extrapolation based on the product of the density and the overlap matrix turns out to be a good choice. However, the quality of the different extrapolation approaches strongly depends on the system applied and the simulation settings. Therefore, which method to use and the extrapolation order need to be tested on the particular system of interest. For very large systems, the density matrix-based method can be used to achieve linear scaling by activating the keyword DENSITY_PROPAGATION.[214]

One of the most relevant application domains for RT-TDDFT is the study of light−matter interactions, e.g. in the field of spectroscopy, excited state dynamics, and radiation damage. To mimic these phenomena, at any time during the propagation, it is possible to apply a time-dependent electric field $\mathbf{F}(t)$. The applied field is in general modulated by an envelope function $E_{env}(t)$ and is defined as

$$\mathbf{E}(t) = \mathbf{P} E_{env}(t) \cos(\omega_0 t + \phi) \tag{103}$$

where $\mathbf{P}$ is the field polarization, ω_0 is the carrying frequency at which the field oscillates, and ϕ its initial phase. The characteristic of the applied field, as well as its time extension, is provided through the section &DFT%EFIELD. The time-dependent electric field defined within this section can only be used in combination with RT-TDDFT. By default, the coupling between the electric field and the electronic degrees of freedom is described within the length-gauge, by adding to the Hamiltonian the dipole coupling term $e\mathbf{E}(t)\cdot\mathbf{r}$. This approach is only valid for isolated molecular systems, and not when periodic boundary conditions are applied. The velocity-gauge form of the equations suitable for periodic systems is obtained through a gauge transformation involving the vector potential[215−217]

$$\mathbf{A}(t) = c \int^t dt' \, \mathbf{E}(t') \tag{104}$$

In the time-dependent KS equations, the vector potential $\mathbf{A}(t)$ appears in the kinetic energy term and, in the case where nonlocal PPs are used, the gauge field also transforms the electron−ion interaction. To use this representation, the VELOCITY_GAUGE keyword has to be used within the aforementioned REAL_TIME_PROPAGATION section. The total energy varies over time as the applied external field interacts with the system. To monitor the time evolution of the field and the various terms that contribute to the total electronic energy, which can be activated via the &REAL_−TIME_PROPAGATION%PRINT print key.

Once the RTP has started, the evolution of the electronic structure can be monitored by means of several descriptors, such as the time-dependent dipole moment, the time-dependent total electronic density, and the spin density, which can be printed out as a series of cube files for every arbitrary number of propagation steps. As for ground state calculations, the output of these quantities is activated by the corresponding &DFT%PRINT print key.

The electron current density can be obtained as

$$\mathbf{j}(\mathbf{r}, t) = \frac{1}{2} \sum_i f_i \left(\psi_i^*(\mathbf{r}, t) \, \boldsymbol{\pi} \, \psi_i(\mathbf{r}, t) + \text{c. c. } \right) \tag{105}$$

where

$$\boldsymbol{\pi} = [\mathbf{r}, \mathbf{H}] = -i\nabla + \mathbf{A}(t) + [V_{PP}, \mathbf{r}] \tag{106}$$

Integrating $\mathbf{j}(\mathbf{r}, t)$ in the simulation cell yields the macroscopic current

$$\mathbf{J}(t) = -\frac{1}{V} \int d\mathbf{r} \, \mathbf{j}(\mathbf{r}, t) \tag{107}$$

where V is the cell volume. The print key &REAL_TIME_−PROPAGATION%PRINT%CURRENT_INT activates the calculation and the output of the macroscopic current. This quantity can then be used to calculate the frequency-dependent conductivity, which is defined as the ratio of the current's Fourier transform to that of the applied electric field

$$\sigma_{ij}(\omega) = \frac{\mathcal{F}[J_i](\omega)}{\mathcal{F}[E_j](\omega)} \tag{108}$$

The projection of the time-dependent orbitals onto some previously selected reference states is particularly useful for verifying the variation in population of specific states, while monitoring density differences is usually helpful for detecting charge-transfer processes. By projecting time-dependent MOs on some reference states (e.g., the initial MOs) using the print key &PRINT%PROJECTION_MO the overlap between the propagated orbital $\psi_i(\mathbf{r}, t) = \sum_\alpha C_{i\alpha}(t)\phi(\mathbf{r})$ and any reference orbital ψ_m^{ref} is calculated. When the reference orbitals are the ground state virtual orbitals, the quantity

$$N_{exc}(t) = \sum_m^{unocc} \sum_i^{occ} \|\langle \psi_m^{ref} | \psi_i(t) \rangle\|^2$$

is an estimate of the number of electrons that have been excited into the unoccupied space.

7.2. Real-Time Bethe−Salpeter Propagation

Although the exact solution of time-dependent KS equations of eq 96 can describe the nonlinear-response of a system to an external field $\hat{F}(t)$, using approximate XC kernels often leads to inaccuracies. For instance, the adiabatic LDA fails to describe excitonic energies accurately.[203,206,218]

Hence, choosing an approximate self-energy operator is more suitable for the description of nonlocal screening effects present in the excitonic spectra.[218] In the RT-BSE scheme implemented in CP2K,[219] the Coulomb-hole plus screened-exchange (COHSEX) self-energy is employed.[51,220−223] Moreover, instead of state propagation, solution of the equation of motion for the single-particle density matrix $\hat{\rho}(t)$

$$\frac{\partial \hat{\rho}}{\partial t} = \frac{-i}{\hbar}[\hat{H}^{\mathrm{eff}}(t), \hat{\rho}(t)] \quad (109)$$

is implemented, where the effective Hamiltonian is given by[220]

$$\hat{H}^{\mathrm{eff}}(t) = \hat{H}^{G_0 W_0} + \hat{U}(t)$$
$$+ \hat{V}^{\mathrm{Hartree}}[\hat{\rho}(t)] - \hat{V}^{\mathrm{Hartree}}[\hat{\rho}(t_0)]$$
$$+ \hat{\Sigma}^{\mathrm{COHSEX}}[\hat{\rho}(t)] - \hat{\Sigma}^{\mathrm{COHSEX}}[\hat{\rho}(t_0)] \quad (110)$$

Therein, $\hat{H}^{G_0 W_0}$ is the $G_0 W_0$ Hamiltonian, $\hat{U}(t)$ is the applied external field, $\hat{V}^{\mathrm{Hartree}}[\hat{\rho}(t)]$ is the Hartree potential (which depends on the single-particle density matrix) and $\hat{\Sigma}^{\mathrm{COHSEX}}[\hat{\rho}(t)]$ is the COHSEX self-energy. This form of the effective Hamiltonian leads to oscillations of $\hat{\rho}(t)$ formally equivalent to solving the Casida equation of eq 78 with Bethe–Salpeter entries shown in eq 84,[220] hence the name of the method.

In order to run a RT-BSE calculation (so far only implemented for molecules, i.e. without PBC, and for static nuclei), a similar setup as for running an RT-TDDFT propagation calculation is employed, namely

- set RUN_TYPE to RT_PROPAGATION

- include a &MD section to set the size of the TIMESTEP and the total number of STEPS

- specify all options necessary to run an accurate $G_0 W_0$ calculation for molecules (in particular the subsection &PROPERTIES%BANDSTRUCTURE%GW should be specified)

- include the &REAL_TIME_PROPAGATION section with all relevant entries and add subsection &RTBSE.

Note that the RT-BSE method employs the RI approximation for the Hartree and COHSEX terms,[127] so an appropriate RI_AUX basis should be set. Example RT-BSE calculations for organic molecules are also available in repositories of the RT-BSE implementation paper.[219] A typical RT-BSE input with the most important &REAL_TIME_PROPAGATION section will look as follows:

```
&REAL_TIME_PROPAGATION
  &RTBSE ! Start the RTBSE Method
    FT_DAMPING [fs] 5.0 ! FT damping
    FT_START_TIME 0.0 ! FT offset
    &PADE_FT ! Padé interpolation
      E_MIN [eV] 0.0
      E_MAX [eV] 50.0
      E_STEP [eV] 0.001
      FIT_E_MIN [eV] 0.0
      FIT_E_MAX [eV] 300.0
    &END PADE_FT
  &END RTBSE
  EPS_ITER 1.0E-8 ! Check convergence
  MAX_ITER 50
  MAT_EXP BCH
  EXP_ACCURACY 1.0E-14 ! < EPS_ITER
  APPLY_DELTA_PULSE
  DELTA_PULSE_DIRECTION 1 0 0
  DELTA_PULSE_SCALE 0.0001
  &PRINT
    &MOMENTS ! Print moments time trace
      FILENAME MOMENTS
    &END MOMENTS
    &MOMENTS_FT ! Fourier transform
      FILENAME MOMENTS_FT
    &END MOMENTS_FT
    &FIELD ! Print applied field trace
      FILENAME FIELD
    &END FIELD
    &POLARIZABILITY
      FILENAME POLARIZABILITY
      ELEMENT 1 1 ! Tensor element
    &END POLARIZABILITY
  &END PRINT
&END REAL_TIME_PROPAGATION
```

The most relevant keywords and subsections in the above section entail:

- EPS_ITER: the equation of motion of eq 109 is solved by exponential evolution operators applied to the single-particle density matrix $\hat{\rho}$. These are applied self-consistently by an enforced time reversal symmetry scheme described previously,[212,214,224] so that

$$\hat{\rho}(t + \Delta t) \approx e^{-i/\hbar \hat{H}^{\mathrm{eff}}(t + \Delta t) \Delta t / 2} \times e^{-i/\hbar \hat{H}^{\mathrm{eff}}(t) \Delta t / 2} \hat{\rho}(t)$$
$$\times e^{i/\hbar \hat{H}^{\mathrm{eff}}(t) \Delta t / 2} \times e^{i/\hbar \hat{H}^{\mathrm{eff}}(t + \Delta t) \Delta t / 2}$$
$$(111)$$

The self-consistency is present since the effective Hamiltonian $\hat{H}^{\mathrm{eff}}(t + \Delta t)$ depends on the single-particle density matrix at time $t + \Delta t$, to which we need to propagate. The self-consistency threshold is controlled by EPS_ITER. Smaller threshold implies more stable propagation, but at the same time usually requires more iterations. Usually, one can choose EPS_ITER similar to EPS_SCF, although a larger tolerance might be sufficient for consistent results - one should test the convergence of EPS_ITER and TIMESTEP pair. Note that the maximum among the absolute values of elements of $\hat{\rho}(t + \Delta t) - \hat{\rho}(t)$ in the AO basis is used as the metric for the threshold.

- MAX_ITER: limits the maximum number of the self-consistent iterations in the ETRS scheme.
- MAT_EXP: the matrix exponentiation in RT-BSE is implemented by two approaches—the (Baker–Campbell–Hausdorff) BCH[212] and the EXACT diagonalization. The keyword MAT_EXP controls the choice between these two.

- EXP_ACCURACY: for the BCH method, it is also necessary to specify the matrix exponentiation accuracy, which should be stricter than the ETRS threshold EPS_ITER.
- APPLY_DELTA_PULSE: in order to observe any oscillations, the external field $\hat{U}(t)$ needs to be specified. Either the explicit time-dependent field can be supplied in the &DFT%EFIELD subsection, or a delta pulse can be triggered by inclusion of this keyword.[225]
- DELTA_PULSE_DIRECTION: sets the Cartesian direction of the delta pulse. The resulting vector is always normalized by the code.
- DELTA_PULSE_SCALE: sets the magnitude of the delta pulse directly (in atomic units). The change in the metric of the single-particle density matrix after the application of the delta pulse is reported by the code - if the change is larger than 1 in atomic units, then the ETRS loop might struggle to converge—we recommend decreasing the DELTA_PULSE_SCALE in such cases.
- &MOMENTS: activates printing of the time-dependent dipole moment

$$\mu_j(t) = \mathrm{Tr}(\hat{\rho}(t)(\hat{r} - \mathbf{r}_{CC})) \tag{112}$$

where μ_j is the jth component of the dipole moment $\mu(t)$ and $\mathbf{r}_{CC}$ is the reference point—in RT-BSE, this is the center of atomic charges.
- FT_DAMPING: Damping γ applied during Fourier transform to stabilize it.[226] Example is given in the &MOMENTS_FT section. If not given explicitly, γ is determined so that factor of e^{-4} is applied at the end of the time trace of the observables, starting from FT_START_TIME.
- FT_START_TIME: offset along the time axis for the purposes of Fourier transform. Defaults to zero. Useful for real-time pulses defined in &EFIELD section. For example, for GAUSSIAN_ENV with nonzero T0 parameter, setting FT_START_TIME to the same value will result in the envelope being a real function in frequency space.
- &MOMENTS_FT: activates printing of the Fourier transform of the dipole moment $\boldsymbol{\mu}(t)$, which is stabilized by the application of FT_DAMPING γ, so that

$$\mu_j(\omega) = \int \mathrm{d}t\, e^{i(\omega + i\gamma)t} \mu_j(t + t_0) \tag{113}$$

where t_0 is the defined by the START_TIME.
- &FIELD: activates printing of the time-dependent trace of the external field. When no &EFIELD subsection within the &DFT section is present, outputs just zeros (for example when only the delta pulse is applied.)
- &POLARIZABILITY: activates printing of the polarizability tensor, defined in terms of the Fourier transform of the dipole moment and the Fourier transform of the electric field as[227,228]

$$\alpha_{jk}(\omega) = \frac{\mu_j(\omega)}{E_k(\omega)} \tag{114}$$

The polarizability is related to the dipole strength function,[227,228] which characterizes the absorption spectrum of the system. Control over which element is printed is given by the ELEMENT keyword. The polarizability calculation also uses the settings from the &MOMENTS_FT section.

- &PADE_FT: section controlling the Padé interpolation[229−231] of the Fourier transforms. A new frequency grid spanning from E_MIN to E_MAX with E_STEP steps is created, on which the Fourier transforms are reevaluated using the Padé fitting of the original transforms. The original transforms are restricted to energies within FIT_E_MIN and FIT_E_MAX. By default, no restriction is applied—entire frequency range of the original transform is used to construct the Padé parameters.

The code automatically saves the state of the system after each successful time step, so that the propagation can be continued after interruption. Use INITIAL_WFN RT_RESTART in the REAL_TIME_PROPAGATION section to enable the restart. Note that the total propagation time determines the energy resolution in the resulting Fourier transform, while the time step determines the maximum frequency/energy accessible to the Fourier transform. In other words, if a finished RT-BSE calculation does not yield the required energy resolution, it is possible to simply increase the number of STEPS and continue the propagation. Note that the &MOMENTS and &FIELD should not be suppressed (for example by an insufficient PRINT_LEVEL) in order for the program to be able to read the time traces from the previous run.

7.3. Surface Hopping via NEWTON-X

The interface with the NEWTON-X program package enables to perform on-the-fly nonadiabatic molecular dynamics (NAMD) simulations using Tully's fewest switches surface hopping algorithm.[232−234] The total electronic WF

$$|\Psi(\mathbf{R}(t))\rangle = \sum_{(n)} d^{(n)}(t)|\Psi^{(n)}(\mathbf{R}(t))\rangle \tag{115}$$

is thereby relying on the TDDFT module of CP2K, thus the singly excited states $|\Psi^{(n)}(\mathbf{R}(t))\rangle$ are defined based on the excitation amplitudes $X_{ia}^{(n)}$ of the eigenvalue problem of eq 78, i.e.

$$|\Psi^{(n)}\rangle = \sum_{ia} X_{ia}^{(n)}|\Phi_{ia}\rangle \tag{116}$$

with the singly excited Slater determinants $|\Phi_{ia}\rangle$. The Lagrangian of eq 78 furthermore defines the excited state nuclear gradients, as outlined in Section 6. The therein defined force $\mathbf{F}$, given as the negative derivative of the variational Lagrangian with respect to the nuclear coordinates $\mathbf{R}$

$$F_\alpha(t) = -\nabla_\alpha L(\mathbf{R}(t)) \tag{117}$$

determines the acceleration $\mathbf{a}$ of nuclei α as

$$\mathbf{a}_\alpha(t) = \frac{\mathbf{F}_\alpha(t)}{M_\alpha} \tag{118}$$

with M_α indicating the corresponding nuclear mass. Corresponding CP2K inputs thus require choosing RUN_TYPE ENERGY_FORCE with a suitable &TDDFPT section and to print excited state nuclear gradients and amplitudes by adding &PRINT subsections to the &FORCE_EVAL and &TDDFPT sections, respectively:

```
&FORCE_EVAL
    &PRINT
        &FORCES
        &END FORCES
    &END PRINT
    ...
&END FORCE_EVAL
&TDDFPT
    &PRINT
        &NAMD_PRINT
            PRINT_VIRTUALS T
        &END NAMD_PRINT
    &END PRINT
    ...
&END TDDFPT
```

The keyword `PRINT_VIRTUALS` implies that the excited state amplitudes are printed within the MO basis, no longer referring to Sternheimer atomic-orbital based quantities. Real-time propagation of the coefficients $d^{(n)}(t)$ of the total electronic WF, as defined in eq 115, requires furthermore computing nonadiabatic coupling elements $\sigma_{(mn)}(t)$ via

$$i\frac{\mathrm{d}d^{(m)}(t)}{\mathrm{d}t} = \sum_{(n)} d^{(n)}(t)[\delta_{(mn)}E_{(n)}(t) - i\sigma_{(mn)}(t)] \tag{119}$$

with $E_{(n)}(t)$ representing the total energy of excited state n, i.e. $E_{(n)}(t) = E_{\mathrm{GS}}(t) + \Omega^{(n)}(t)$. Input options for efficient time-derivative couplings are set within corresponding NEWTON-X input files, requiring no further specifications within the CP2K input file. To generate initial geometries and velocities based on a Wigner probability distribution, a print statement has to be added to the `&VIBRATIONAL_ANALYSIS` section to enable printing of Cartesian normal modes:

```
&VIBRATIONAL_ANALYSIS
    &PRINT
        &NAMD_PRINT
        &END NAMD_PRINT
    &END PRINT
&END VIBRATIONAL ANALYSIS
```

8. X-RAY SPECTROSCOPY

The prediction of properties that depend on the core electronic density requires the explicit calculation of all-electron WFs and the use of large basis sets, close to the basis set limits. This is the case for inner-shell spectroscopy, where excitation processes of core orbitals involve the rearrangement of the all-electron charge distribution. In CP2K/QUICKSTEP, all-electron calculations require the GAPW approach, as described in Section 2.

The X-ray absorption process is related to the incoming photon being absorbed by core-level electrons, which get resonantly promoted to an excited state. The process has a certain probability depending on the properties of the initial and final states. Within the dipolar approximation, certain selection rules apply, such as total spin conservation and total angular momentum change by ±1. Due to the very local atomic character of the process, for excitation of a 1s level, only unoccupied states with a local p-character are probed. The spectroscopic process is ultrafast (attosecond) so that the nuclear movement during the excitation can be neglected.

8.1. Transition Potential Method

A KS orbital is by definition the WF of a fictitious noninteracting system introduced to reproduce the electron density of the real interacting system. Hence, KS orbitals are considered formal auxiliary quantities without specific physical meaning. Nevertheless, the KS eigenvalues are commonly used to describe physical quantities such as ionization energies and excitation energies, and the orbitals are used to obtain the transition moment. To compute excitation energies in less approximate ways, the ΔSCF KS method has been devised, comparing two systems with one or more excited electrons, taking into account full relaxation.

The incoming radiation is represented by the electromagnetic field

$$\mathbf{A}(\mathbf{r}, t)A_0\mathbf{e}\cos(\mathbf{k}\cdot\mathbf{r} - \omega t) \tag{120}$$

The resulting transition probability between the initial and final state is

$$P_{if} = \frac{\pi e^2}{2\hbar m^2 c^2}A_0^2 |\langle f|e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{e}\cdot\mathbf{p}|i\rangle|^2 \tag{121}$$

which within the dipole approximation simplifies to $P_{if} \approx |\langle f|\mathbf{e}\cdot\mathbf{p}|i\rangle|$, when using the velocity form. This form is equivalent to the length form $P_{if} \approx (E_f - E_i)|\langle f|\boldsymbol{\mu}|i\rangle|$, where $\boldsymbol{\mu}$ is the dipole operator, when the WFs are exact eigenstates. The single-electron excitation approximation makes it possible to rewrite the initial state as a core WF and the final state as an unoccupied/free electron WF, thereby assuming that all other electrons do not participate. The corresponding matrix element can be rewritten as a single-electron matrix element. The transition potential (TP) method adopts this independent particle approach, but the orbital energies and WFs are taken as a solution of the KS equation with a modified core potential on the absorbing atom. The modified potential employed in standard applications is the TP introduced by Slater and is based on the creation of half core-hole (HCH). Other types of core-hole potentials can also be employed with satisfactory results, such as the full core-hole (FCH). The actual final location of the promoted electron is not taken into account for the determination of the spectra, assuming that it is immediately delocalized in the conduction band and its contribution is nearly equivalent, irrespective of the specific final state. In this manner, only a single electronic structure calculation is needed to determine the entire spectrum.

The procedure consists of running an initial ground state calculation, identifying the core state that has to be excited, changing the occupation number of the corresponding KS orbital, and restarting the SCF optimization while keeping the occupation of the excited core state at the modified value. The excitation energies and transition moments are finally evaluated from the KS energies and MOs obtained with the modified occupation

$$\hbar\omega_{if} = \varepsilon_f^{\mathrm{TP}} - \varepsilon_i^{\mathrm{TP}} \qquad P_{if} = |\langle\psi_f^{\mathrm{TP}}|\nabla|\psi_i^{\mathrm{TP}}\rangle|^2 \tag{122}$$

It is good practice, after the first ground state optimization, to localize the core orbitals to facilitate the identification of the target core state. The localization is activated via the `&DFT%XAS%LOCALIZE` section, and is carried out for `STATE_SEARCH` states, starting from the lowest:

```
        &XAS
          &SCF
              ...
          &END SCF
          METHOD        TP_HH
          DIPOLE_FORM   VELOCITY
          STATE_TYPE    1s
          STATE_SEARCH  10
          ATOMS_LIST    1
          ADDED_MOS     1000
          &LOCALIZE
          &END LOCALIZE
          &PRINT
            &PROGRAM_RUN_INFO
          &END PRINT
          &RESTART
            FILENAME ./root
          &END RESTART
          &XAS_SPECTRUM
            FILENAME ./root
          &END XAS_SPECTRUM
          &XES_SPECTRUM
            FILENAME ./root
          &END XES_SPECTRUM
        &END XAS
```

By default, the TP calculation uses the same settings as specified for the ground state. Any different settings given in the `&XAS%SCF` subsection override the original ones. Anyway, TP calculations run only with the diagonalization method because KS orbitals and their occupations are needed. The emptied core state is defined by the atom, as specified in `ATOMS_LIST`, and the orbital character `STATE_TYPE`. An independent TP calculation is performed for each atom of the list. The `METHOD` keyword defines how the occupation of the core orbital is modified, whereas `ADDED_MOS` are the additional virtual orbitals that will be computed to obtain a spectrum over a wider energy range. The TP calculation simultaneously provides XAS, core-to-empty transitions, XES, and valence-to-core transitions. Since each TP calculation entails a full SCF optimization, the usual electronic structure output, such as the projected density of states and the cube files of the MOs, are available. These quantities are generally very useful for the analysis of the transition character and rationalize the electronic properties.

8.2. Linear-Response Time-Dependent Density Functional Theory

Within LR-TDDFT theory, optical absorption energies are calculated as solutions of the generalized eigenvalue problem presented in Section 6.1. A full diagonalization yields the complete spectrum, with electronic transitions from all occupied to all unoccupied states, including X-ray absorption near-edge structure (XANES), also known as near-edge X-ray absorption fine structure (NEXAFS), from core states to low-lying unoccupied states. However, this approach is not suitable for large systems, as its computational cost scales as $O(N^6)$ with system size (diagonalization of a $N_{occ}N_{empty}$ matrix).

For efficiency, the CP2K/QUICKSTEP LR-TDDFT implementation for XAS relies on the following three approximations.[197] The first is the core−valence separation (CVS) approximation.[235] Due to large differences in energy and localization, the core and valence states are only weakly coupled. This reduces the dimensions of the eigenvalue problem from $N_{occ}N_{empty}$ to $N_{core}N_{empty}$, where only occupied core states are considered. In the so-called sudden approximation,[236] excitations from various core states are further

decoupled. Smaller eigenvalue problems of size N_{empty} for s-type core states or $3N_{empty}$ for p-type states can be solved independently, leading to further computational savings. However, this approximation relies on the assumption that the core states are highly localized in space.

The localization of core states is further exploited for the efficient calculation of 4-center 2-electron ERIs, using a core-specific RI scheme. Coulomb integrals involving the core states I, J are computed as

$$(pI|Jq) \approx (pI|\mu)(\mu|\nu)^{-1}(\nu|Jq) \tag{123}$$

where μ and ν correspond to RI basis functions centered on the same atom where I and J are localized. The same holds for exact-exchange ERIs in hybrid DFT calculations, i.e. $(pq|IJ) \approx (pq|\mu)(\mu|\nu)^{-1}(\nu|IJ)$. Finally, ERIs involving the XC kernel are computed as follows

$$(pI|f_{xc}|Jq) \approx (pI|\kappa)(\kappa|\lambda)^{-1}(\lambda|f_{xc}|\mu)(\mu|\nu)^{-1}(\nu|Jq) \tag{124}$$

Yet, since all RI basis elements are centered on the excited atom, $f_{xc}[n]$ can be evaluated by a projection of the total electronic density onto the RI basis, expressed as a linear combination

$$\begin{aligned} \rho(\mathbf{r}) &= \sum_{pq} P_{pq}\varphi_p(\mathbf{r})\varphi_q(\mathbf{r}) \\ &\approx \sum_{pq} \sum_{\mu\nu} P_{pq}(pq\nu)S_{\mu\nu}^{-1}\chi_\nu(\mathbf{r}) \\ &= \sum_\nu d_\nu\chi_\nu(\mathbf{r}), \end{aligned} \tag{125}$$

where P_{pq} is the density matrix in the AO basis, while $(pq\nu)$ represents the overlap integrals between two general AOs and a RI basis element, and $S_{\mu\nu}^{-1}$ is the inverse overlap matrix element of the RI functions centered on the excited atom.

Using LR-TDDFT to compute XAS spectra is based on perturbation theory on top of a ground state. Therefore, the quality of the underlying converged SCF calculation is critical. Since the core electrons of the excited atoms need to be explicitly described, the GAPW method is necessary. Excited atoms must be described with all-electron basis sets, while PPs may be used for all other atoms. A RI basis can be provided for any atomic kind in the system using the `RI_XAS` keyword, in the `&KIND` subsection (e.g., `BASIS_SET RI_XAS def2-TZVP-RIFIT`). If not provided, a default RI basis is generated on-the-fly using the robust method of Stoychev.[81] For hybrid DFT simulations, the ADMM approximation can be used for the SCF calculation, specifically in its ADMM2 flavor. The `RUN_TYPE` keyword in the `&GLOBAL` input section must be set to `ENERGY`. In addition, LR-TDDFT-based XAS calculations can be performed for both molecular and periodic systems.[237]

The input subsection governing XAS LR-TDDFT calculations is `&XAS_TDP`, which is within the `&DFT` section. Some keywords are related to general LR-TDDFT: `TAMM_DANCOFF`, which controls the TDA (enabled by default), `DIPOLE_FORM` to calculate the electronic dipole in either the length or velocity-gauge, and `EXCITATIONS` to enable the calculation of singlet/triplet excitations in closed-shell systems or spin-conserving/spin-flip excitations in open-shell systems. The remaining input is specific to XAS calculations.

Because all approximations used in the LR-TDDFT method of the `&XAS_TDP` section rely on the localization of core states, it is crucial to select them properly and check their properties. This is done via the `&XAS_TDP%DONOR_STATE` subsection, which sets the rules for searching for the most suitable core states among the low-energy MOs. First, excited atoms are defined either `BY_INDEX` or `BY_KIND` with the `DEFINE_EXCITED` keyword. In the former case, an `INDEX_LIST` must be provided with the indices of the excited atoms. In the latter case, a `KIND_LIST` is provided, specifying which atomic kind is to be excited. The `STATE_-TYPES` keyword then defines which core-level to excite from (1s, 2s or 2p), for each atom index/kind. Finally, the `N_SEARCH` keyword specifies the N lowest-energy MOs to screen for the donor core states previously defined. Note that in some cases, excited atoms might be equivalent under symmetry (e.g., carbon atoms in an acetylene molecule C_2H_2). In such a case, the two lowest-energy canonical MOs will be linear combinations of s-type states centered on each atom and effectively delocalized. To ensure that the underlying approximations of the LR-TDDFT XAS method are valid, the N candidate donor states can be localized using the `LOCALIZE` keyword. Localization can be expensive for a large number of MOs: therefore, it is recommended to keep the value of `N_SEARCH` to a minimum. An example `&DONOR_-STATE` subsection is given below for the CO_2 molecule:

```
&DONOR_STATES
    DEFINE_EXCITED BY_KIND
    KIND_LIST O
    LOCALIZE
    N_SEARCH 3
    STATE_TYPES 1s
&END DONOR_STATES
```

In the above example, the three lowest-energy MOs of the system are localized and used as candidates for O 1s donor states. The 2 candidates that best overlap with a minimal STO basis representation of an O 1s core state are selected. The overlap, as well as the Mulliken charge analysis, is reported in the calculation output. These figures should be checked to ensure the suitability of donor states.

Another important input subsection of the LR-TDDFT method is `&XAS_TDP%KERNEL`, which defines the XC kernel. Both the XC functional and exact exchange for hybrid DFT calculations are specified here. This subsection must always be explicitly included, even if the kernel is the same as the ground state functional. All recommendations for periodic calculations with HF apply equally for the `&EXACT_EX-CHANGE` subsection. Note that only global hybrids are currently supported. Below is an example `&KERNEL` sections for a periodic PBE0 hybrid calculation, with the truncated Coulomb potential and a cutoff radius of 6 Å:

```
&KERNEL
    &XC_FUNCTIONAL
        &GGA_X_PBE
        &END GGA_C_PBE
        &GGA_X_PBE
            SCALE 0.75
        &END GGA_X_PBE
    &END XC_FUNCTIONAL
    &EXACT_EXCHANGE
        FRACTION 0.25
        OPERATOR TRUNCATED
        CUTOFF_RADIUS 6.0
    &END EXACT_EXCHANGE
&END KERNEL
```

Also related to the XC kernel is the `&XAS_TDP%GRID` keyword. As shown in eq 125, f_{xc} is evaluated on a projection of the density on the RI basis elements centered on the excited atom. The integral is then performed numerically on an atomic grid. By default, the grid dimensions are the same as the GAPW settings for the SCF calculation. It is, however, recommended to use a finer grid. For example,

```
GRID O 100 250
```

would define grids with 100 angular points (Lebedev scheme) and 250 radial grid points for excited oxygen atoms. Note that the GAPW default is a 50 × 50 grid.

Finally, SOC calculations for L-edge spectroscopy can be enabled by adding the `SPIN_ORBIT_COUPLING` keyword. The SOC is added perturbatively on top of an LR-TDDFT calculation using the auxiliary many-electron WF (AMEW) framework.[238] For spin-restricted systems, both single- and triplet-excitations must be calculated. In the open-shell case, spin-conserving and spin-flip excitations must be calculated. For example,

```
SPIN_ORBIT_COUPLING
EXCITATIONS RCS_SINGLET
EXCITATIONS RCS_TRIPLET
```

would turn on SOC for LR-TDDFT-based XAS calculations on top of a restricted closed-shell ground state.

It is generally recommended to run LR-TDDFT XAS calculations with hybrid functionals containing a large fraction of exact exchange to reduce the self-interaction error (e.g., PBEh (α = 0.45) or BHHLYP). For such simulations, the computational bottleneck is often still the SCF cycle. Yet, large performance gains can be achieved using the ADMM approximation. Diagonalization costs can also become high for large systems with multiple excited atoms (e.g., all oxygen atoms in a liquid-water simulation). Hence, the iterative `&OT_SOLVER` can be used to reduce the time spent diagonalizing multiple matrices. Note that, due to its relatively high initialization cost, the OT solver should not be used for a single excited state. Generally, the automatically generated RI basis is very robust and accurate. However, its size is typically larger than some existing preoptimized RI bases (e.g., def2-TZVP-RIFIT). If such a basis exists, using it might save computing time as well.

If unphysical excitation energies are obtained, there are a few possible culprits. The most likely reason is that the donor core states are not properly prepared. They may not be localized or fall outside the range specified by `&DONOR_STATES%N_SEARCH`. One should check the output file for the overlap and the Mulliken charge analysis (both should be close to 1.0, or even higher for the overlap of 2p states). Another possibility is that the density projection in eq 125 is inaccurate. This sometimes happens when a heavier atom described with an all-electron basis is in the vicinity of the excited atom: because the RI basis functions are centered on the latter, they might fail at describing the core states of the former. The easiest solution is to describe the heavier atom with a PP. If this is not possible, e.g. because of interest in excitations from both atoms, one can use the `RI_REGION` keyword of the `&KERNEL` subsection. This allows for the projection of the density on a larger RI basis, with contributions from all atoms within a sphere around the excited atom. In such a case, the use of a fine radial grid is recommended, in order to describe the sharp features of the neighboring atom. Finally, the issue could be numerical, and

the `GRID`, or various general parameters could need tightening, or the usage of a better basis set.

The LR-TDDFT XAS method and its implementation are extensively discussed in ref 197. Various accuracy and performance benchmarks were performed, and all input files are openly available in ref 239.

8.3. Real-Time Time-Dependent Density Functional Theory

Spectroscopic properties can also be obtained by means of RT-TDDFT. The absorption spectrum ranging from ultraviolet–visible (UV–vis) to X-rays can be computed by studying the time dependence of the induced electric dipole moment, or the induced current density in condensed matter systems, after subjecting the system to an external electric field in the form of a very narrow function in time (delta kick). Such a step-function-like external potential will excite all the modes in the system. A corresponding sample input section looks like this:

```
&REAL_TIME_PROPAGATION
    INITIAL_WFN  SCF_WFN
    PROPAGATOR ETRS
    EXP_ACCURACY 1E-15
    EPS_ITER 1E-9
    MAX_ITER 100
    MAT_EXP TAYLOR
    APPLY_DELTA_PULSE .TRUE.
    DELTA_PULSE_DIRECTION 0 0 1
    DELTA_PULSE_SCALE 0.001
&END REAL_TIME_PROPAGATION
&PRINT
    &MOMENTS
        FILENAME =mom.dat
        PERIODIC F
    &END MOMENTS
&END PRINT
```

This approach has been extensively studied for years to simulate UV–vis absorption spectra. The first completely general implementation of RT-TDDFT for core states was done in 2012 by K. Lopata et al. within the NWChem code.[240] The absorption spectrum of a system can be obtained by performing three simulations employing the RTP procedure, each applying the field polarized along one Cartesian axis. For nonperiodic simulations, one applies a field in the form of a Dirac delta function, i.e. $\mathbf{E}_\delta = \kappa\delta(t)\mathbf{e}$, where κ is the strength and $\mathbf{e}$ the polarization. The time-dependent electric dipole moment is obtained from the propagated electronic charge density

$$\boldsymbol{\mu}(t) = \int d\mathbf{r}\, n(\mathbf{r}, t)\mathbf{r} \tag{126}$$

Assuming the linear regime, i.e. $\boldsymbol{\mu}(\omega) = \boldsymbol{\alpha}(\omega)\cdot\mathbf{E}(\omega)$, the molecular polarizability tensor is then obtained by Fourier transforming the time-dependent dipole moment

$$\alpha_{ij}(\omega) = \frac{1}{\kappa}\mu_{ij}(\omega) \tag{127}$$

where i is the vector component and j the polarization direction. The absorption spectrum is finally obtained by computing the trace of the polarizability tensor

$$S(\omega) = \frac{4\pi\omega}{3c}\mathrm{Tr}[\mathfrak{F}[\alpha(\omega)]] \tag{128}$$

According to time-dependent perturbative approaches, the polarizability is real for nonresonant frequencies and diverges at resonances. Therefore, the resonant frequencies are characterized by a nonzero imaginary value and a change of sign for the real part. By analyzing both components of the polarizability together, it is thus possible to verify whether the peaks in the spectrum correspond to actual electronic transitions. This behavior of the polarizability can be helpful to understand whether the spectra obtained via RTP can be considered converged with respect to different simulation parameters, such as the time step and total propagation time.

The velocity-gauge RT-TDDFT formalism is required to apply a time-dependent field to periodic systems. As discussed in a previous section, the vector potential appears in the kinetic term and, when PPs are used, the gauge field also transforms the electron–ion interaction. Although the length-gauge and velocity-gauge forms yield the same result for finite systems, the velocity-gauge representation is periodic for spatially uniform external electric fields. Integrating the time-dependent current density $\mathbf{j}(\mathbf{r}, t)$ in the simulation cell yields the macroscopic current

$$\mathbf{J}(t) = -\frac{1}{V}\int d\mathbf{r}\, \mathbf{j}(\mathbf{r}, t) \tag{129}$$

In the linear-response regime, by applying a very short pulse, the Fourier transform of the time-dependent current yields the frequency-dependent conductivity $\sigma_{ij}(\omega)$ and the frequency-dependent dielectric function.

$$\epsilon(\omega) = 1 + \frac{4\pi i\, \mathrm{Tr}[\boldsymbol{\sigma}(\omega)]}{\omega} \tag{130}$$

The imaginary part of the frequency-dependent dielectric function corresponds to the optical absorption spectrum in a condensed matter system. An example input looks as follows:

```
&REAL_TIME_PROPAGATION
    INITIAL_WFN SCF_WFN
    PROPAGATOR ETRS
    EXP_ACCURACY 1E-15
    EPS_ITER 1E-9
    MAX_ITER 100
    MAT_EXP TAYLOR
    VELOCITY_GAUGE
    &PRINT
        &FIELD
        &END FIELD
        &E_CONSTITUENTS
        &END E_CONSTITUENTS
        &CURRENT_INT
        &END CURRENT_INT
    &END PRINT
&END REAL_TIME_PROPAGATION
&EFIELD
    INTENSITY 3.5E10
    POLARISATION  0  0  1
    ENVELOP GAUSSIAN
    &GAUSSIAN_ENV
        T0 [fs]    0.006240
        SIGMA [fs] 0.002340
    &END GAUSSIAN_ENV
    VEC_POT_INITIAL  0.00 0.00 ${VECZ}
&END
```

9. ENERGY DECOMPOSITION ANALYSIS

Intermolecular bonding arises from a complex interplay of electrostatic interactions between permanent charges and molecular multipole moments, polarization effects, Pauli repulsion, donor–acceptor orbital interactions (also referred

to as covalent, charge-transfer, or delocalization interactions), and weak dispersion forces. Energy decomposition analysis (EDA) quantifies the contribution of each of these components to the total binding energy, providing deeper insight into the physical origins of intermolecular bonds.

The CP2K/QUICKSTEP implementation of EDA is based on absolutely localized molecular orbitals (ALMOs), which are molecular orbitals confined entirely to individual molecules or ions within a larger system.[241,242] ALMO EDA separates the total interaction energy (TOT) into a frozen density (FRZ), polarization (POL) and charge-transfer (CT) terms, i.e.

$$\Delta E_{\text{TOT}} = \Delta E_{\text{FRZ}} + \Delta E_{\text{POL}} + \Delta E_{\text{CT}} \tag{131}$$

The frozen interactions term is defined as the energy required to bring isolated molecules into the system without any relaxation of their MOs, apart from modifications associated with satisfying the Pauli exclusion principle, hence

$$\Delta E_{\text{FRZ}} \equiv E(R_{\text{FRZ}}) - \sum_x E(R_x) \tag{132}$$

where $E(R_x)$ is the energy of the isolated molecule x and R_{FRZ} is the density matrix of the system constructed from the unrelaxed MOs of the isolated molecules. ALMO EDA is closely related to the block-localized WF EDA,[243] because both approaches use the same variational definition of the polarization term as the energy lowering due to the relaxation of each molecule's ALMOs in the field of all other molecules in the system

$$\Delta E_{\text{POL}} \equiv E(R_{\text{ALMO}}) - E(R_{\text{FRZ}}) \tag{133}$$

The strict locality of ALMOs is utilized to ensure that the relaxation is constrained to include only intramolecular variations. This approach gives an upper limit to the true polarization energy,[244] and its mathematical and algorithmic details have been described by several authors.[245,246] The remaining portion of the total interaction energy, the CT term, is calculated as the difference in the energy of the relaxed ALMO state and the state of fully delocalized and optimized orbitals (R_{SCF}), i.e.

$$\Delta E_{CT} \equiv E(R_{\text{SCF}}) - E(R_{\text{ALMO}}) \tag{134}$$

Therefore, ΔE_{CT} includes the energy lowering due to electron transfer from the occupied ALMOs on one molecule to the virtual orbitals of another molecule $\Delta E_{CT}^{(\text{pair})}$, as well as the further energy change caused by many-body higher-order induction ΔE_{HO} that accompanies such an occupied-virtual orbital mixing and is typically small for intermolecular interactions, thus

$$\Delta E_{CT} = \Delta E_{CT}^{(\text{pair})} + \Delta E_{\text{HO}} \tag{135}$$

A distinctive feature of ALMO EDA is that both the amount of the electron density transferred between a pair of molecules and the corresponding energy lowering can be computed via[247]

$$\Delta Q_{\text{CT}}^{(\text{pair})} = \sum_{x,y>x} \{\Delta Q_{x \to y} + \Delta Q_{y \to x}\} \tag{136a}$$

$$\Delta E_{\text{CT}}^{(\text{pair})} = \sum_{x,y>x} \{\Delta E_{x \to y} + \Delta E_{y \to x}\} \tag{136b}$$

9.1. Implementation

Within CP2K, ALMO EDA is restricted to closed-shell fragments. The linear-scaling optimization of ALMOs serves as its underlying computational engine,[248] which can be applied to both gas-phase and condensed phase systems. It has also been extended to fractionally occupied ALMOs enabling investigation of interactions between metal surfaces and molecular adsorbates.[249] Another unique feature of our implementation is the ability to control the spatial range of CT between molecules using the cutoff radius R_c (see below). Additionally, ALMO EDA in combination with CP2K's efficient AIMD engine allows us to attenuate or switch off CT interactions in AIMD simulations, thus measuring their contribution to the dynamical properties of molecular systems.[250,251] A single-point ALMO EDA calculation can be set up by creating a regular input file for DFT calculation and then modifying it in several steps.

9.1.1. Step 1. Define Fragments. First of all, all atoms in the system must be assigned to nonoverlapping subsets called fragments. In other words, the system must be partitioned into fragments. A fragment can include a single atom, several nonbonded atoms, a single molecule (that is, several bonded atoms), or several molecules. This partitioning tells the code about localization regions of ALMOs when calculating the ΔE_{FRZ} and ΔE_{POL} energy terms.

To define fragments, the ALMO code uses the CP2K subroutines that analyze bonding between atoms using distance thresholds. By default, all atoms that are considered bonded to each other are assigned to the same fragment. This approach enables a very flexible definition of fragments through a combination of (a) the `&GENERATE` subsection of the `&&FORCE_EVAL%SUBSYS%TOPOLOGY` section, (b) the fifth column in the `&SUBSYS%COORD` section, and (c) an optional connectivity file. The information about each fragment is printed at the beginning of output files under the "ALMO SETTINGS" marker.

The `&GENERATE` subsection automatically determines bonding between atoms by comparing interatomic distances to the sum of the atomic radii. This is the most convenient strategy to define bonding between atoms and thus to create fragments. The default behavior of `&GENERATE` can be changed by modifying its keywords. For example, regular molecules can be partitioned into atomic fragments using low values of the `BONDPARM_FACTOR`. Low values prevent atoms from being combined into molecular fragments, as shown in the following input:

```
&SUBSYS
    &TOPOLOGY
        &GENERATE
            BONDLENGTH_MAX 1.0
            BONDPARM COVALENT
            BONDPARM_FACTOR 0.3
        &END GENERATE
    &END TOPOLOGY
&END SUBSYS
```

Partitioning of the system can be refined using the fifth column in the `&COORD` section that assigns fragment labels to each atom. The main principle of this partitioning scheme is that atoms with different fragment labels cannot be combined into a fragment. On the other hand, atoms with the same label may still be separated into fragments as guided by the `&GENERATE` subsection. Finally, the bonding between atoms

Table 1. ALMO EDA Energies in CP2K Output Files

state	CP2K print out	definition	EDA terms
(0)	"single-molecule energy"	$\sum_x E(R_x)$ in eq (132)	$\Delta E_{FRZ} = (1) - (0)$
(1)	"energy of the initial guess"	$E(R_{FRZ})$ in eq (132)	$\Delta E_{POL} = (2) - (1)$
(2)	"ENERGY OF BLOCK-DIAGONAL ALMOs"	$E(R_{ALMO})$ in eq (133)	$\Delta E_{CT}^{(pair)} = (3) - (2)$
(3)	"CORRECTED ENERGY"	eq (135)	$\Delta E_{HO} = (4) - (3)$
(4)[a]	"ENERGY|"	$E(R_{SCF})$ in eq (134)	$\Delta E_{CT} = (4) - (2)$

[a]If FULL_X or XALMO_X methods are used then the "ENERGY\|" line contains the energy of state (3).

can be explicitly specified in a connectivity file. In this case, all bonded atoms are assigned to the same fragment.

9.1.2. Step 2. Assign Electrons to Fragments. All electrons of a neutral atom are assumed to belong to the atom's fragment. This default behavior can be changed by adding (removing) electrons to (from) atoms using the &BS subsection in the &SUBSYS%KIND section. In the following example, one electron is removed from the 1s orbitals of hydrogen atoms to create H^+ fragments. Meanwhile, two electrons are added to the 2p orbitals of oxygen atoms to create O^{2-} fragments. Since both H^+ and O^{2-} are intended to be closed-shell fragments, the same modification has to be performed for the &ALPHA and &BETA subsections of the &BS section. Note that ALMO EDA can even handle zero-electron fragments, such as H^+.

```
&SUBSYS
   &KIND H
      ...
      &BS
         &ALPHA
            NEL  -1
            L     0
            N     1
         &END
         &BETA
            NEL  -1
            L     0
            N     1
         &END
      &END
   &END KIND

   &KIND O
      ...
      &BS
         &ALPHA
            NEL  +2
            L     1
            N     2
         &END
         &BETA
            NEL  +2
            L     1
            N     2
         &END
      &END
   &END KIND
&END SUBSYS
```

After atoms and electrons are assigned to fragments, the Gaussian basis set functions, all of which are centered on the atoms in CP2K/QUICKSTEP, are also assigned to the same fragments as their atoms. This last partitioning defines the structure of the block-diagonal matrix of ALMO coefficients.

9.1.3. Step 3. Turn ALMO SCF On. To run ALMO EDA calculations, set the ALMO_SCF keyword in the &DFT%QS

section to TRUE. This tells CP2K to optimize ALMO coefficients in the ALMO SCF loop instead of optimizing MO coefficients in the conventional SCF loop. That is, the &DFT%SCF section of the input does not control ALMO EDA calculations. Instead, the ALMO SCF loop is controlled by the &DFT%ALMO_SCF section.

9.1.4. Step 4. Compute ΔE_{FRZ}. To request the calculation of the FRZ term, set &ALMO_SCF%ALMO_SCF_GUESS to MOLECULAR. This instructs CP2K to perform a series of conventional SCF calculations for individual fragments in the periodic box of the entire system. These calculations are controlled by the &SCF section of the input file, and, therefore, it is important that all keywords in this section (e.g., EPS_SCF) are set to ensure sufficient accuracy of the energies of the individual fragments. CP2K/QUICKSTEP then combines the MO of individual fragments into the block-diagonal matrix of the ALMO coefficients and evaluates the energy of this state without any optimization of the coefficients. The computed energies are printed as described in Table 1.

9.1.5. Step 5. Compute ΔE_{POL}. Subsequent optimization of the block-diagonal ALMO coefficients can be performed using various algorithms, which are selected using the ALMO_ALGORITHM keyword and controlled by the corresponding subsections of the &ALMO_SCF section. Although DIIS-accelerated diagonalization of the diagonal blocks of the projected KS Hamiltonian matrix typically converges the ALMO SCF in just a few iterations,[245,246] this method does not guarantee convergence and can fail for strongly interacting fragments. In problematic cases, preconditioned conjugate gradient or trust region algorithms can be employed instead of the DIIS-accelerated diagonalization. The polarization energy does not depend on the algorithm and is printed as described in Table 1.

9.1.6. Step 6. Compute ΔE_{CT}. The calculation of the CT term is controlled by the DELOCALIZE_METHOD keyword in the &ALMO_SCF section. In order to compute ΔE_{CT} defined in eq 134, set this keyword to FULL_SCF. In order to separate CT into pair contributions defined in eqs 136a and 136b, respectively, DELOCALIZE_METHOD should be set to FULL_X_THEN_SCF. If only the pair contributions are of interest and the higher-order term is expected to be negligible, set DELOCALIZE_METHOD to FULL_X.

For very large systems, the computation of the CT terms can be significantly speeded up by evaluating them only between fragment neighbors within the user-defined localization radius. This is done by specifying the localization cutoff distance with &ALMO_SCF%XALMO_R_CUTOFF_FACTOR and by setting DELOCALIZE_METHOD to XALMO_SCF, or to XALMO_X. The latter two settings are equivalent to using FULL_SCF and FULL_X for fully delocalized orbitals. Note that pair CT components can be calculated with XALMO_X, but cannot be obtained with XALMO_SCF, respectively.

The computed CT energies are printed as described in [Table 1](#). The printout of the pair CT terms can be activated by adding the &ANALYSIS subsection inside &ALMO_SCF and specifying output file names inside the &PRINT subsection. Each message passing interface process will print its own part of the pair list into a separate file. These output files have three columns. The first column is the index of the fragment that accepts electrons. The second column is the index of the fragment that donates electrons. The third column is the amount of electron density transferred in units of the elementary charge or the corresponding CT energy in Hartrees.

9.1.7. Input File Example. The relevant input sections for an ALMO EDA calculation involve:

```
&QS
   ALMO_SCF T
&END QS
&ALMO_SCF
   ALMO_ALGORITHM DIAG
   ALMO_SCF_GUESS MOLECULAR
   DELOCALIZE_METHOD FULL_X_THEN_SCF
   EPS_FILTER 1.0E-10
   &ALMO_OPTIMIZER_DIIS
      EPS_ERROR 1.0E-5
      MAX_ITER 10
      N_DIIS 5
   &END ALMO_OPTIMIZER_DIIS
   &ANALYSIS T
      &PRINT
         &ALMO_CTA
            FILENAME charge_terms
         &END ALMO_CTA
         &ALMO_EDA_CT
            FILENAME energy_terms
         &END ALMO_EDA_CT
      &END PRINT
   &END ANALYSIS
   &XALMO_OPTIMIZER_PCG
      CONJUGATOR DAI_YUAN
      EPS_ERROR 1.0E-5
      LIN_SEARCH_EPS_ERROR 0.1
      LIN_SEARCH_STEP_SIZE_GUESS 0.5
      MAX_ITER 100
      MAX_ITER_OUTER_LOOP 0
   &END XALMO_OPTIMIZER_PCG
&END ALMO_SCF
```

9.1.8. Illustrative Applications. ALMO EDA has been applied to study intermolecular interactions in a variety of gas and condensed phase molecular systems,[252] as well as their interfaces.[253,254] The implementation of ALMO EDA in CP2K has played a crucial role in advancing our understanding of hydrogen bonding in liquid water,[242,255−257] ice[258] and water confined to low dimensions.[259−261] These studies have systematically shown that the small amount of electron density transferred between molecules has a profound effect on the structure,[257,262] spectroscopic response,[256,262−265] and dynamical properties[250,251,266−268] of liquid water.

10. FINITE TEMPERATURE EFFECTS

For the purpose of explicitly including finite temperature effects, CP2K implements a large variety of MC and MD algorithms that can be activated via the &MOTION%MC and &MOTION%MD sections, respectively.[5,269−273] The most distinctive feature of CP2K, however, is the ability to conduct efficient AIMD simulations,[10,13] and in combination with linear-scaling algorithms, even for rather large-scale sys-

tems.[77,274,275] Yet, in spite of its name, initially CP2K only contained Born−Oppenheimer MD,[276−279] since the original Car−Parrinello algorithm is not implemented.[280] Instead, the second-generation Car−Parrinello method is available,[11,12] but since this has been extensively covered elsewhere,[281−286] the present focus is solely on MLPs and NQE by means of PIMD.

10.1. Neural Network and Machine Learning Interaction Potentials

The rise of MLPs in recent years provides efficient and accurate representations of potential energy surfaces from electronic structure reference data.[287−294] This has enabled the simulation community to reach previously inaccessible spatiotemporal scales through atomistic simulations of complex, reactive systems at near ab initio accuracy as hitherto only known from explicit (on-the-fly) AIMD[9,13] simulations. High-dimensional neural network potentials (NNPs)[288,295] combined with atom-centered symmetry functions to describe atomic environments[296] have been the first methodology to provide a scalable and accurate representation of potential energy surfaces using artificial neural networks. Introduced in 2007 by Behler and Parrinello, this framework has been used to provide fundamental insight into a plethora of challenging systems, from reactive gas-phase clusters,[297,298] to phase behavior of liquids like water,[299,300] chemical reactions at interfaces,[301] high pressure systems,[302−304] complex materials,[305] battery interfaces,[306] and phase-transitions,[307,308] to name but a few.[288,293,309] Although many other methods have been proposed over the years, NNPs remain a mainstay in the toolbox of computational scientists. Their conceptually simple and computationally efficient design and their systematic extensions to treat long-range effects via environment-dependent charges[310] and self-consistent charge equilibration[311,312] to target remaining limitations within the same framework makes them a versatile and robust tool to address problems of ever increasing complexity.[293] For the purpose of this tutorial, we concentrate here on the original formulation, which is deeply implemented in CP2K and very simple to use by newcomers in the field.

To represent a potential energy surface by NNPs, the atomistic structure is first transformed using atom-centered symmetry functions into translationally and rotationally invariant descriptors of atomic environments.[296] These serve as input for atomic neural networks that output auxiliary components of the total potential energy, which is then obtained as a sum of contributions from all atoms in the system. The resulting permutationally invariant structure-energy relation can be analytically differentiated to obtain the nuclear forces, for example, to drive MD, and is scalable to essentially arbitrary system sizes.[295] Model training is achieved by optimizing the parameters (weights and biases) of the atomic neural networks to reproduce reference energies and optionally forces. Besides the original RuNNer code of Behler, other training codes like n2p2 and RubNNet4MD are available and yield models that can be directly used in CP2K. For a light introduction into NNPs and other MLPs we refer the reader to ref [313], as well as to previous reviews on this topic.[314,315]

The implementation of NNPs in CP2K is fully compatible with the original RuNNer code and thus n2p2 format and can be used in the same way as other potential energy models. The NNP module in CP2K is part of the &FORCE_EVAL section and can be used in combination with other methods such as DFT, traditional force fields, or QM/MM methods. Inside the

`&NNP` subsection, the keyword `NNP_INPUT_FILE_NAME` reads the NNP from a file, which can be generated using the RuNNer code or other compatible training codes such as n2p2 or the RubNNet4MD package.[316] It is noted only in passing that the atomic cluster expansion (ACE) MLP technique[317] has already been interfaced with CP2K and that its tree graph extension (grACE)[318] is addressed as this article is written. Similarly, well-developed interfaces to CP2K also exist for DeePMD-kit,[319] NequIP[320] and Allegro,[321] all of which can be activated via their respective specialized subsections in `&FORCE_EVAL%MM%FORCEFIELD%NONBONDED`. Contrary to the native NNP implementation of CP2K, the corresponding libraries must be linked during compile time, as described in detail in Section 11.1.

The NNP is then used to compute the energy and forces of the system, where the analytical stress tensor is also available for constant-pressure simulations. Parallization is achieved using particle decomposition and scales well up to thousands of cores and number of atoms. An example of the input section for the NNP module is shown here:

```
&FORCE_EVAL
   METHOD NNP
   &NNP
      NNP_INPUT_FILE_NAME input.nn
      SCALE_FILE_NAME scaling.data
      &MODEL
         WEIGHTS weigths
      &END MODEL
   &END NNP
&END FORCE_EVAL
```

The original NNP formalism can also be extended to so-called ensemble or committee models, where multiple NNPs are combined for boosted accuracy and error estimation. Individual committee members are trained separately from independent random initial conditions to a subset of the total training set.[322] If the resulting committee shares the same atom-centered symmetry functions as descriptors for the atomic environments,[296] there is only a small, often negligible, computational overhead in production runs. While the committee average provides more accurate predictions than the individual NNPs, the committee disagreement, defined as the standard deviation between the committee members, grants access to an estimate of the error of the model. This committee disagreement provides an objective measure of the error of the underlying model,[323] which can be used in active learning protocols to improve the model systematically.[324−326] This has been utilized extensively in recent times for the automated development of NNPs for various systems,[327] while current research showed that it leads to more diverse configuration sampling, but not necessarily to more accurate models.[328] This committee approach is also implemented in CP2K.[322]

Committee NNPs (C-NNPs) can be used in the same way as single NNPs in CP2K, with the addition of the `&NNP% MODEL` section for each committee member. The committee members are defined by the number of `&MODEL` subsections in the input file, where each subsection contains the weights of the NNP. It is important to ensure that these models have been trained with the same set of descriptors because they are assumed to be shared between all committee members. An example of the input section for a C-NNP setup with three committee members is shown here:

```
&FORCE_EVAL
   METHOD NNP
   &NNP
      NNP_INPUT_FILE_NAME input.nn
      SCALE_FILE_NAME scaling.data
      &MODEL
         WEIGHTS nnp-1/weigths
      &END MODEL
      &MODEL
         WEIGHTS nnp-2/weigths
      &END MODEL
      &MODEL
         WEIGHTS nnp-3/weigths
      &END MODEL
   &END NNP
&END FORCE_EVAL
```

Besides these features, the NNP module in CP2K also provides the possibility to print the energy and force disagreement between committee members, as well as configurations which are in the extrapolation regime of the model. This can be used to identify regions of the potential energy surface where the model is inaccurate in order to improve it systematically. These outputs are controlled using the `&PRINT` subsection in the `&NNP` section. An example of the input section for the printing of the committee energy disagreement is shown in the following:

```
&PRINT
   &ENERGIES
      &EACH
         MD 1
      &END EACH
   &END ENERGIES
&END PRINT
```

Similarly, the printing of forces, force disagreement, and extrapolation points can also be enabled (`FORCES`, `FORCE_SIGMA`, `EXTRAPOLATION`).

In addition, there is an option to bias the committee disagreement to stabilize simulations by restraining the system dynamics to regions where the committee members agree. This can be achieved by adding a harmonic repulsive wall acting on the committee disagreement after a chosen threshold, which adds a bias to the total potential energy of the system. An example of the input section for the biasing of the committee energy disagreement looks like:

```
&BIAS
   SIGMA_0 [hartree] 0.0001
   K_B 5000
&END BIAS
```

Moreover, `&BIAS%ALIGN_NNP_ENERGIES` can be used to subtract an energy shift from individual committee members to reduce misalignment effects, while bias energies can be printed with the `&BIAS%PRINT` section using `&BIAS_ENERGY`; further details of this feature can be found in ref 322.

Finally, we note that the parameters of a C-NNP for water are available within the CP2K data repository and can be accessed via the `NNP/bulkH2O-jcp2020-cnnp/nnp-X/` path in the input file. The model has been trained against revPBE0-D3 hybrid functional reference data and is suitable for water over a wide range of temperatures and pressures. Further details of the development can also be found in ref 322. We use this model to illustrate the easy usability of the NNP implementation in CP2K by showing results for liquid

water at 300 K in Figure 2a–c, respectively. From the comparison of the structural and dynamical properties of liquid
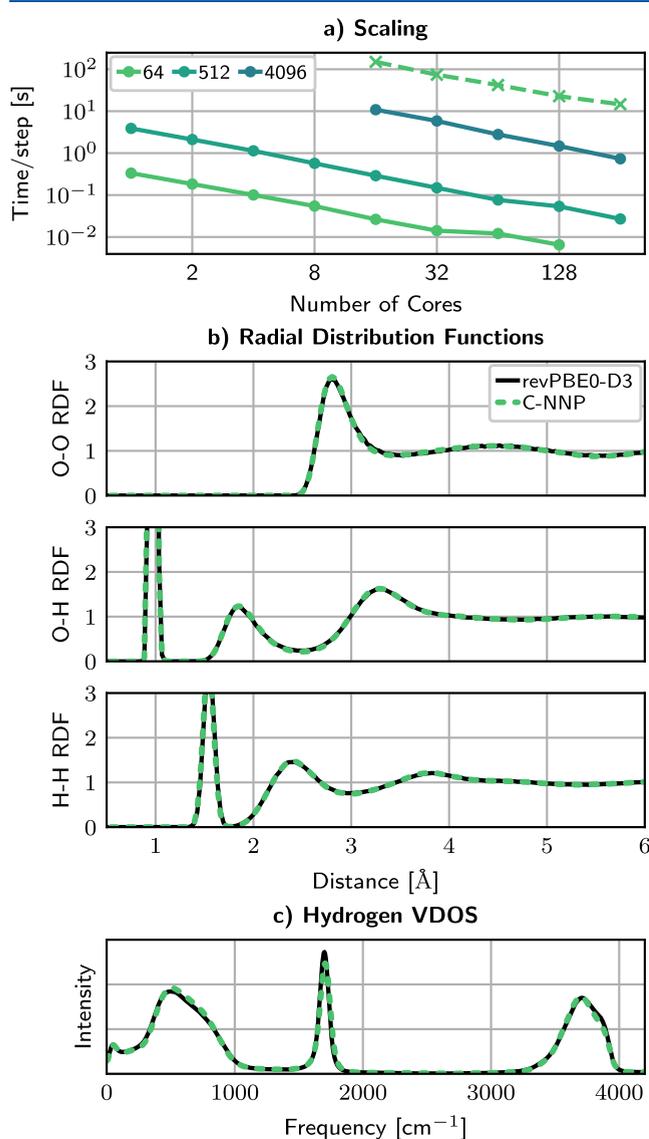
Figure 2. Comparison of NNP and AIMD results for liquid water at 300 K as obtained from CP2K. (a) Time per MD step for 64, 512, and 4096 water molecules with the C-NNP model (solid lines with circles) and 64 water molecules with the revPBE0-D3 hybrid functional (dashed line with crosses) reported on a logarithmic scale. (b) Radial distribution functions and (c) vibrational density of states of the hydrogen atoms. Data in panel (b,c) are partially reused from ref 322. Copyright 2020, American Institute of Physics.

water, we can see that the C-NNP model provides a faithful representation of the system, with the radial distribution functions and vibrational density of states being in perfect agreement with the reference AIMD data. The computational efficiency of the C-NNP model is also demonstrated by the comparison of the time per MD step for different system sizes, which shows that the C-NNP model (solid lines) is orders of magnitude faster than the reference AIMD simulations (dashed line). This makes the C-NNP model a powerful tool for the study of complex systems at finite temperatures.

10.2. Nuclear Quantum Effects

The quantum nature of nuclei can have a significant impact on the properties of materials, especially at low temperatures.[329−343] A very elegant way to include these NQEs in molecular simulations is the imaginary time path-integral (PI) formalism, which is based on the Feynman−Kac formulation of quantum statistical mechanics.[7,9,344−346] In the PI formalism, the quantum mechanical system is mapped onto a classical system of P replica, connected by harmonic springs to adjacent replica under cyclic boundary conditions in the simplest case (leading to the so-called primitive Trotter approximant). Replicas (a.k.a. "beads") of the same index are coupled through a scaled version of the potential energy surface, which is determined by the Trotter number P. This approach allows for the calculation of quantum statistical observables by sampling the classical phase space of the P-fold system. For more details on the PI method, we refer the reader to refs 7,9,347−350.

In CP2K, PI simulations are available via the `&PINT` section, which provides a flexible and modular environment for the inclusion of the quantum nature of the nuclei. In this PIMD framework, the equations of motion resulting from the effective potential are integrated numerically to sample the PI formulation of the canonical (NVT) partition function of the system.[7,9,351,352] This implementation mainly supports the normal mode transformation for the efficient integration of the stiff harmonic ring-polymer modes.[353] In addition to the reversible reference system propagator algorithm (RESPA) multiple time step integrator (`NRESPA` with `HARM_INT NUMERIC`),[354] an exact integrator for normal mode coordinates is also available via `HARM_INT EXACT`.[355] An example input section for a PIMD simulation is shown here for a system of $P = 32$ replica at 250 K and 1000 steps of 0.25 fs:

```
&PINT
    P 32
    NUM_STEPS 1000
    TEMP        250
    DT          0.25
    TRANSFORMATION NORMAL
    HARM_INT EXACT
    PROPAGATOR PIMD
&END PINT
```

For standard PIMD simulations in the canonical ensemble, the temperature can be controlled using two thermostatting techniques, namely the massive Nosé-Hoover chain (mNHC) thermostat,[356,357] applied separately to all degrees of freedom mostly in normal mode coordinates, and the PI Langevin equation (PILE) thermostat.[355] They are enabled via dedicated subsections that change the ensemble from the microcanonical (NVE) to the canonical ensemble, if specified. The Nosé−Hoover chain thermostat is controlled by the `&NOSE` subsection, whereas the PILE thermostat is controlled by the `&PILE` subsection. While the Nosé-Hoover chain thermostat only requires the definition of the chain length, the PILE thermostat needs the definition of the coupling constant λ and the relaxation time τ as demonstrated here:

```
      &NOSE
         NNOS 5
      &END NOSE

      or

      &PILE
         LAMBDA 0.5
         TAU 1000.0
      &END PILE
```

In addition, there are also quantum thermostats available that are specifically designed to improve the convergence with respect to the number of replicas in PI simulations, discussed in more detail in Section 10.3.

Finally, it is also possible to constrain the centroid or all replica degrees of freedom using the CP2K-wide constraint framework defined by the &MOTION%CONSTRAINT section. Within that section, the PIMD_BEADWISE_CONSTRAINT F|T keyword switches between beadwise and centroid constraints. The KT_CORRECTION keyword in the &PINT section can be used to correct for the loss of temperature due to constrained degrees of freedom for Nosé-Hoover chains and numerical integration. Constraints and restraints, which are available through the same framework, can be used to obtain free energy profiles, including NQEs through thermodynamic integration, umbrella sampling and umbrella integration, respectively.[358,359]

10.3. Quantum Convergence

The PI formalism is exact in the limit of infinite discretization, i.e. $P \rightarrow \infty$. However, in practice, some finite P value must be selected while the properties of interest converge with increasing P. Unfortunately, this convergence is system, temperature and mass dependent, making it computationally demanding to treat systems in the deep quantum regime, where unpleasantly large P values are required. For example, the bead convergence of structural properties of hydrogen-bonded systems requires roughly $P \approx 32$ at 300 K, $P \approx 128$ at 100 K, and $P \approx 8192$ at about 1 K, as analyzed in ref 360. Due to these reasons, different techniques have been developed to accelerate the convergence of PI simulations.[361] In particular, colored noise thermostats[362−364] in combination with generalized Langevin dynamics and quantum thermal baths have been shown to significantly reduce the required number of beads for convergence down to the ultracold temperature regime.[360,365] Two such thermostats, the PI generalized Langevin equation thermostat (PIGLET)[363] and the PI quantum thermal bath (PIQTB) thermostat,[364] are available in CP2K. The main idea behind both of these approaches is to impose a frequency-dependent temperature $T(\omega)$, which matches the quantum fluctuations of a harmonic oscillator. Due to zero-point energy leakage, these thermostats require rather harsh coupling constants, rendering them incompatible with approximate quantum dynamics methods like centroid MD (CMD) and ring-polymer MD (RPMD), described in more detail in Section 10.4.

The PIGLET thermostat, based on the generalized Langevin equation, is enabled with the &PINT%PIGLET thermostat subsection. This formalism requires dedicated matrices to impose the correct frequency-dependent thermostatting for each combination of temperature and number of beads. A large variety of matrices can be downloaded from https://gle4md.org/index.html?page=matrix, requiring to use the raw format for data consistency. The matrices file must be specified in the &PIGLET subsection using the MATRICES_FILE_NAME

keyword. The PIQTB formalism, using a quantum thermal bath thermostat for each ring-polymer normal mode, is also based on the one-body density matrix as PIGLET but has the advantage of working continuously for any combination of temperature and number of beads. It is enabled using the &QTB thermostat section after having specified the underlying PIMD simulation, we show an example input section for the PIQTB thermostat below. In addition to the coupling constant λ and the relaxation time τ, the angular cutoff frequency for the normal modes needs to be specified by suitable values for TAUCUT and LAMBCUT, respectively. The example values below have been shown to work well for hydrogen-bonded systems[360] and should also be transferable to other systems. They have been used for the bead convergence study in Figure 3a.

```
      &QTB
         LAMBDA 0.2
         TAU 50.0
         TAUCUT 0.9
         LAMBCUT 1.5
      &END QTB
```

Both of these quantum thermostats have been shown to significantly reduce the number of beads required for convergence in PI simulations from ambient down to ultralow temperatures.[360] We note in passing that they can also be used in the limit of $P = 1$ in CP2K, where they impose the frequency-dependent enhancement of the quantum fluctuations of a harmonic oscillator on the classical system to mimic quasi-classical quantum fluctuations. For PIGLET, this is supported via a dedicated thermostat section &GLE, while for PIQTB, the same thermostat section &QTB is simply used in combination with setting $P = 1$.

10.4. Approximate Quantum Dynamics

Within the standard PI formalism, MD is usually only performed as a sampling tool for the quantum partition function in the spirit of an extended Lagrangian formalism.[7,9,351,352] However, there are several PI formulations that allow for the approximate description of quantum dynamics, such as CMD[366] and RPMD.[367] Both of these techniques can be seen as approximations to Matsubara dynamics,[368] which is the exact quantum dynamics in the PI formalism. For more details on the theoretical footing of these methods, we refer the reader to refs 369−371. Both CMD and RPMD are available in CP2K and are the only practical methods that allow one to approximately include NQEs on dynamics in (chemically complex) condensed phase systems. Unfortunately, both methods have been shown to suffer from fundamental artifacts, namely CMD from the curvature problem and RPMD from the chain resonance problem;[372] thermostated RPMD (TRPMD) has been introduced to reduce the resonance problem by adding friction[373] (see below). We note in passing that more recently, Brownian chain MD (BCMD)[374] has been introduced, which is a short-time step version of the PI hybrid MC (PIHMC) technique[375] applied to all noncentroid modes (whereas the centroids are propagated using Newtonian dynamics). Interestingly, BCMD has been shown[374] to eliminate the resonance problem of RPMD and to alleviate the curvature problem of CMD.

The RPMD formalism requires switching the keyword PROPAGATOR in the &PINT section to RPMD. First, RPMD simulations in the NVT ensemble should be performed to generate initial configurations and velocities for subsequent
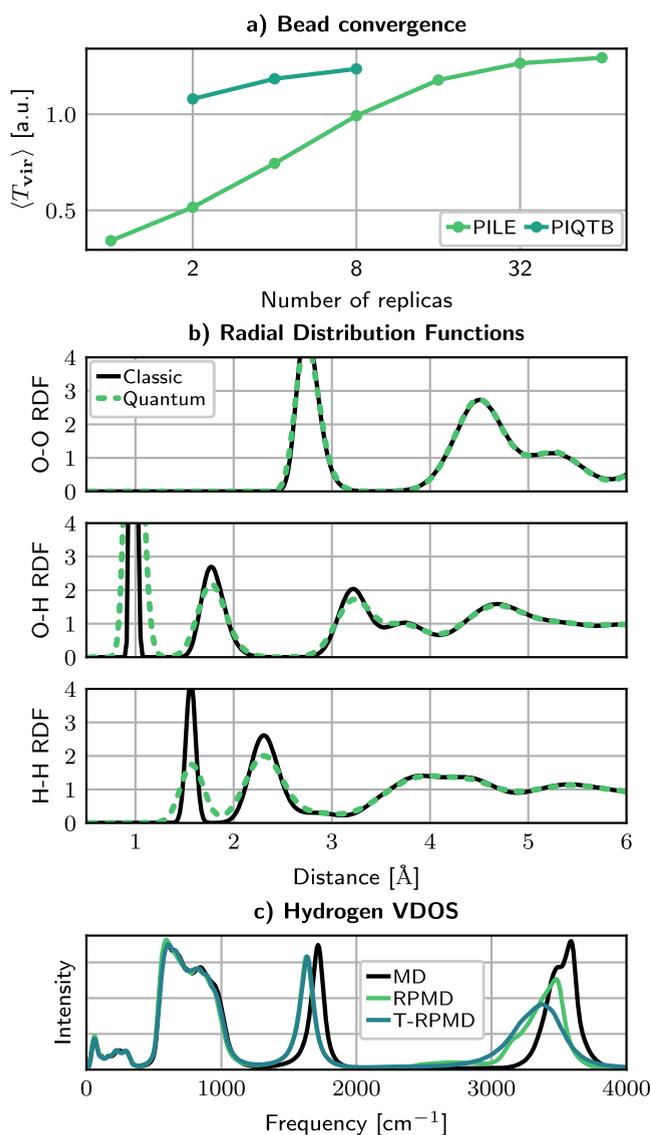
Figure 3. Impact of NQEs on the structure and dynamics of hexagonal ice at 250 K. (a) Bead convergence of the virial kinetic energy estimator using the PILE and PIQTB thermostat. (b) Radial distribution functions with classical and quantum nuclei and (b) vibrational density of states obtained with classical MD, RPMD and TRPMD, respectively. All simulations rely on using the C-NNP trained with revPBE0-D3 reference data,[322] as available in the CP2K data repository (see also Section 10.1).

NVE simulations to generate the actual RPMD dynamics. To speed-up structural thermalization, it is possible to start with standard PIMD *NVT* simulations and then switch to RPMD, resampling the velocities, while keeping the equilibrated configuration. In addition, CP2K also supports the TRPMD method,[373] which is a variant of RPMD that applies a stochastic thermostat to all ring-polymer modes, except for the centroid mode. This idea is an ad hoc approach to remove spurious resonances of the ring-polymer modes with the true modes of the system, first discussed in detail in ref 372, by adding noise. We note that the TRPMD approach usually leads to peak broadenings and line shape distortions in vibrational spectra due to the influence of the thermostat. In CP2K, TRPMD can be simply enabled by setting the TAU parameter in the &PILE subsection to a negative value. An example

input section for TRPMD to reproduce the results of Figure 3c is shown here, where RPMD dynamics would be enabled by removing the &PILE subsection:

```
&PINT
    P 32
    NUM_STEPS 1000
    TEMP        250
    DT         0.25
    HARM_INT   EXACT
    NRESPA 1
    TRANSFORMATION NORMAL
    PROPAGATOR RPMD
    &PILE
        TAU   -1000
        LAMBDA 0.5
    &END PILE
&END PINT
```

To conduct CMD simulations, the keyword &PINT% PROPAGATOR has to be set to CMD. Partially adiabatic CMD (paCMD) shifts the noncentroid degrees of freedom to a high-frequency regime in order to sample the average force of the centroid used for the propagation of the centroid. In CP2K, paCMD is controlled by the GAMMA parameter in the &NORMALMODE subsection which is a mass-scaling factor for the noncentroid degrees of freedom.[372] It is crucial to reduce the time step in CMD simulations given the very high frequency of the noncentroid degrees of freedom. We recall that CMD results in artificial temperature-dependent red-shifts of vibrational frequencies due to the well-known curvature problem.[372,376]

Despite their known limitations and approximations,[370,372] CMD and RPMD are powerful tools to generate approximate quantum dynamics in the PI formalism. They can be used to study a wide range of condensed phase systems, as well as of course gas phase systems, to compute not only vibrational spectra, but also rate constants, relaxation times, and transport properties. In particular, these methods can be easily combined with the above-described NNP module to study the quantum dynamics of complex systems, as recently demonstrated for liquid water at coupled-cluster (CC) accuracy[300] including quantification of H/D isotope effects on translational and orientational dynamics,[377] as well as on hydrogen bond kinetics.[378]

We illustrate the capabilities of the PI formalism in CP2K with a study of hexagonal ice I_h at 250 K using the C-NNP model introduced in Section 10.1. Bead convergence of the virial kinetic energy estimator, which illustrates the quantum nature of the system, is shown in Figure 3a. While the virial kinetic energy converges with roughly $P \approx 32$ at 250 K using the PILE thermostat, we get a significant improvement in convergence using the PIQTB thermostat presented in Section 10.3. We note here that although 32 replica suffice to converge NQEs in liquid water at 300 K (as explicitly demonstrated in Figure S1 of ref 377 using the CCSD(T) C-NNP from ref 300), that number of beads is on the smaller side for crystalline water at 250 K and certain properties require more replica for convergence. The radial distribution functions in Figure 3b reveal the well-known structural differences imprinted by the quantum nature of the system. Finally, the vibrational density of states obtained from standard MD with classical nuclei, as well as from RPMD and TRPMD in Figure 3c shows the impact of NQEs on the vibrational spectrum. While all quantum spectra are red-shifted compared to the classical

spectrum, the RPMD spectrum features some additional peak modulations (namely a red-wing shoulder of the OH stretching peak and artificial vibrational intensity between that stretch and the bending band) due to artificial resonances of the fictitious ring-polymer modes with the physical vibrational modes of the system. The TRPMD spectrum, on the other hand, is seen to reduce the chain resonance problem, but shows an unphysical broadening of the peaks due to the influence of the thermostat on the ring-polymer modes as alluded to in Section 10.4. In summary, the PI formalism in CP2K provides a powerful tool to study the quantum nature of nuclei in a wide range of systems, in particular if combined with NNPs.

10.5. Bosonic Quantum Solvation at Ultralow Temperatures

So far, our discussion of NQEs has been restricted to particles that are assumed to be distinguishable even if they are identical. This implies that all nuclei in PIMD-based simulations of such systems can be numbered akin to classical particles, which results in Maxwell−Boltzmann (MB) quantum statistics. In MB quantum statistics, all nuclei are subject to quantum delocalization and tunneling effects according to their mass, in stark contrast to classical particles. In nature, however, identical particles are fundamentally indistinguishable. Their many-body WF, or finite-temperature density matrix $\rho(R,R',\beta)$, must be totally symmetric or antisymmetric in case of identical bosonic or Fermionic particles, according to the spin-statistics theorem. The density matrix can be formally symmetrized or antisymmetrized by introducing a sum over all possible permutations $\{\mathcal{P}\}$ of identical particles, i.e.

$$\rho^{BE}_{FD}(R, R', \beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \rho(R, \mathcal{P}R', \beta) \quad (137)$$

where the positive and negative sign refer to bosons and Fermions, respectively; R are all particle coordinates and $\beta = 1/k_B T$. The resulting statistics are called Bose−Einstein (BE) and Fermi−Dirac (FD) quantum statistics; we refer to ref 7 in the context of the molecular simulation approach to quantum statistical mechanics.

In the framework of numerical PI simulations of BE and FD statistics, this implies that the discretized density matrix used to compute the partition function and all properties needs to be explicitly symmetrized or antisymmetrized with respect to the exchange of all identical particles.[7,346] Therefore, an additional sum must be sampled in permutation space since the final coordinates of some particle along its quantum path might be some permutation of the initial coordinates, as illustrated here for the partition function that describes three identical particles:

In case of bosonic systems such as ⁴He, all terms contribute with a positive sign to the symmetrized trace,[379] while the odd permutations carry a negative sign in the Fermionic case; we mention in passing that this gives rise to the infamous "sign problem" in equilibrium PI simulations of many-body Fermionic systems such as electrons.[380,381]

The pioneering bisection PIMC method to sample the bosonic permutations, which are required in PI simulations of superfluid ⁴He, has been introduced four decades ago using a Lévy construction;[382,383] we refer to ref 349 for an exhaustive review on this technique. Two decades later, the continuous/space worm algorithm for PIMC simulations of bosons in the grand-canonical ensemble has been devised to greatly improve the sampling efficiency, in particular for large bosonic systems using open path ("worm") sampling that involves off-diagonal elements of the density matrix.[384,385] In CP2K, a canonical worm algorithm with specific sampling extensions, which are relevant for finite bosonic clusters of a given particle number subject to strong interactions of the ⁴He atoms with impurities, has been introduced and implemented, as described in the appendix of ref 386. It therefore represents the default technique to simulate bosonic exchange and thus superfluid properties when using CP2K.

As a technical aside, we recall that sampling bosonic exchange is characterized by nonpolynomial complexity and, therefore, gets increasingly demanding the larger the bosonic system is, even when using efficient worm sampling. For efficiency reasons, not only cubic, but also truncated octahedral boundary conditions have been implemented in CP2K for the solvation supercell. The latter type allows for a smaller overall volume and thus a smaller necessary number of bosonic solvent species to describe a fixed spherical internal volume. For superfluid fraction calculations, the winding number estimator (see below) has been properly generalized for the noncubic supercell shape.[387] In CP2K, the truncated octahedral supercell is activated as follows in conjunction with the use of canonical worm sampling:

```
&HELIUM
    ....
    CELL_SHAPE OCTAHEDRON
    PERIODIC T
    SAMPLING_METHOD WORM
    ....
&END HELIUM
```

Apart from sampling the exchange of all identical particles via path permutations, PI simulations of superfluid helium require one to reach temperatures on the order of 1 K or less. In traditional PIMC or PIMD simulations, some high-temperature density matrix is used that, by applying it many times in a product form, brings the system down to the desired temperature. Due to their simplicity, one-body high-temperature density matrices together with the primitive Trotter approximant are employed in most simulations including NQEs using PIMD-based methods, as discussed in Sections 10.2 and 10.3, in the framework of mNHC, PILE, PIGLET, PIQTB, CMD, RPMD or TRPMD thermostats, all of which are implemented in CP2K.

In the realm of PIMC simulations at ultralow temperatures, say on the order of 1 K or even less, the numerical pair density matrix approach has been introduced decades ago[347] to greatly reduce the number of PI beads or replica that are needed to converge the discretized PI to its continuum limit in the case of two-body interactions,[388] as reviewed earlier.[349] Transcending

primitive approximants, the high-temperature many-body density matrix is represented in this case in terms of one- and two-body contributions to the total PI action, where the latter can be exactly computed for all individual pairs of bosons that interact by a known two-body (pair) potential. In CP2K, we provide a numerical fit[349] of the ^{4}He$\cdots^4$He pair density matrix computed at a "high temperature" of 80 K, using a very accurate two-body interaction potential,[389] which has been computed on a regular grid and subsequently spline-tabulated for its efficient use. Based on this pair density matrix, a PI discretization of only P = 80 replica suffices to perform converged PIMC simulations of bulk superfluid helium at a temperature of 1 K using the CP2K internal implementation:

```
&MOTION
    &PINT
        NUM_STEPS 100
        &HELIUM
            HELIUM_ONLY
            NATOMS 32
            NBEADS 80
            DENSITY 0.0218312
            CELL_SHAPE OCTAHEDRON
            PERIODIC T
            POTENTIAL_FILE_NAME \
                helium_aziz95_80k.potx
            PRESAMPLE T
            SAMPLING_METHOD WORM
            N_OUTER 10000
        &END HELIUM
    &END PINT
&END MOTION
```

Note that for converged observables, many more steps than the 100 selected above need to be simulated, and the necessary pair density file `helium_aziz95_80k.potx` can be found in the `tests/Pimd/untested_inputs` directory of CP2K.

The superfluid fraction f_s is a key property (order parameter) that characterizes the phase transition from the normal liquid state of bosons, where f_s = 0, to its superfluid phase, where $f_s \to$ 1 as $T \to$ 0 K. The CP2K output also allows one to compute, in a postprocessing step, the superfluid fraction of bulk systems and the associated superfluid density ρ_s using the winding number estimator[349,383]

$$f_s = \frac{\rho_s}{\rho_{tot}} = \frac{m_{He}\langle W^2 \rangle}{\beta \hbar^2 N_{He}} \tag{138}$$

where W is the winding number that counts how many times a given path wraps around the periodic supercell along a given direction multiplied by its box length in that direction; all other variables are self-explanatory. The output of this quantity is controlled via the `&WINDING_NUMBER` subsection inside the `&PINT%HELIUM%PRINT` section. The output consists of the numerical value of the prefactor of eq 137 as the initial comment line and the vectorial output of W for each requested time step. Alternatively, one can turn on the output of the mean squared of the simulation so far via the `&WINDING_NUMBER_2_AVG` subsection. Also, this output contains the necessary prefactor value. Note that in both cases, the prefactor in the files is without the $1/N_{He}$ factor. After finite-size extrapolation, eq 137 provides the rigorous expectation value of f_s in bosonic bulk systems subject to periodic boundary conditions.[390]

We illustrate in Figure 4b the behavior of the superfluid fraction, as well as that of the heat capacity in Figure 4a, featuring the characteristic λ-shape as a function of temper-
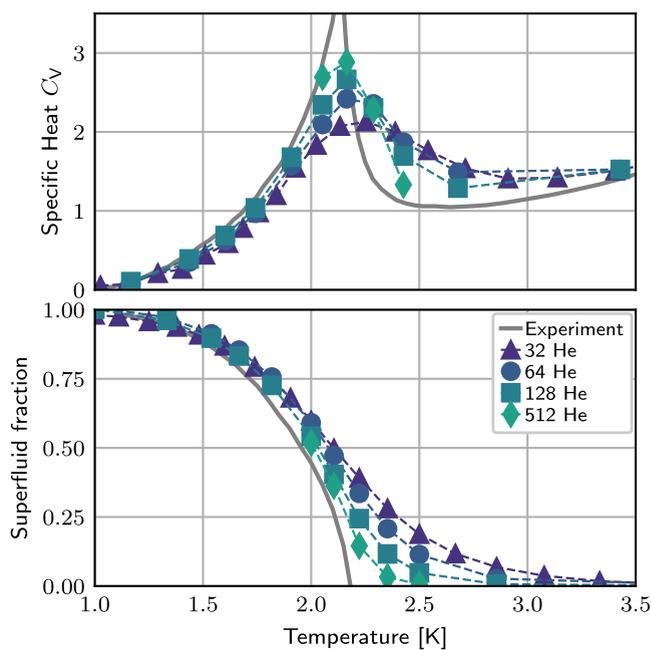
Figure 4. Temperature dependence of (a) the heat capacity and (b) the superfluid fraction of bulk ^{4}He obtained using the canonical worm algorithm to sample bosonic exchange and the winding number estimator to analyze superfluidity (see text). The numerical pair density matrix is the one provided with the CP2K package and has been computed at 80 K using the He$\cdots$He potential from ref 389, thus $P \times T$ = 80 K is kept constant for the temperature scan. The present bosonic PIMC data were obtained from only 32 atoms (in a truncated octahedral supercell according to the sample input provided in the text) and are compared to published results of ref 386 generated from 64, 128, and 512 atoms in the corresponding periodic truncated octahedron using CP2K and the same methodology. For reference, the respective experimental data[391,392] are shown by gray lines; note that deviations of the computed properties from experiment are mainly due to finite-size effects close to the superfluid phase transition given such small system sizes.[390]

ature of bulk ^{4}He once it gets cooled down from the normal liquid state below the λ-transition where superfluidity sets in note that f_s approaches unity at about 1 K. This calculation has been carried out using 32 ^{4}He atoms (the complete input can also be found in the `tests/Pimd/untested_inputs` directory of CP2K) for the point at 1.6 K (20 independent runs of this type enter the plotted data for each temperature).

For finite systems, as particularly relevant in the case of molecular impurities solvated in bosonic environments at very low temperatures to be introduced in the following, there is no rigorous expression available to compute the superfluid fraction. In CP2K, the area estimator[393] is the default for good reasons, i.e.

$$f_s = \frac{\rho_s}{\rho_{tot}} = \frac{4m_{He}^2 \langle (\vec{n}\cdot\vec{A})^2 \rangle}{\beta \hbar^2 I_c} \tag{139}$$

where $\vec{n}\cdot\vec{A}$ is the projected area with

$$\vec{A} = \frac{1}{2} \sum_{i=1}^{N_{He}} \sum_{s=1}^{P} \vec{r}_i^{(s)} \otimes \vec{r}_i^{(s+1)} \tag{140}$$

along some direction $\vec{n}$ and
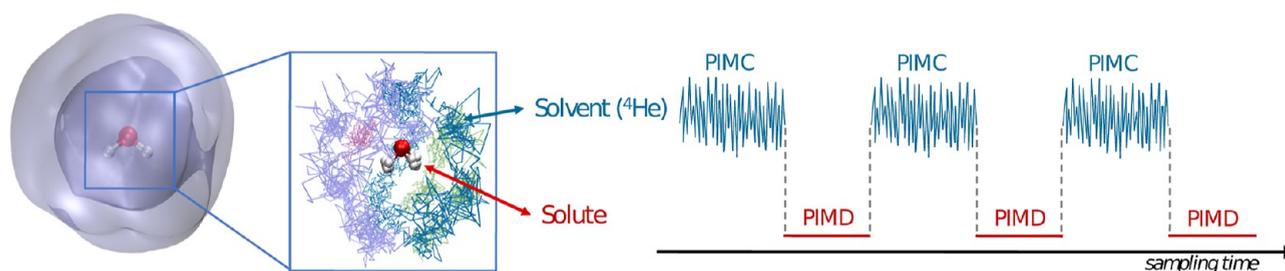
Figure 5. For an illustration of the HPIMD/MC method, as implemented in CP2K.[386] The fully flexible and reactive solute species are propagated using PIMD techniques (such as PIQTB in conjunction with the one-body density matrix), while the bosonic solvent (for instance many ^{4}He atoms either forming a finite cluster as illustrated here, or hosted within a periodic supercell) is sampled using bosonic PIMC (for instance based on the canonical worm algorithm in conjunction with the numerical pair density matrix) at a common temperature T (see text). The interactions can be described using high-dimensional NNPs with CCSD(T) accuracy, as described in Section 10.1.

$$I_c = \left\langle \frac{1}{P} \sum_{i=1}^{N_{He}} \sum_{s=1}^{P} m_i (\vec{n} \otimes \vec{r}_i^{(s)}) \cdot (\vec{n} \otimes \vec{r}_i^{(s+1)}) \right\rangle \tag{141}$$

is the respective classical moment of inertia.

Obtained from linear-response theory of macroscopic systems, the area estimator works convincingly for large systems but, unfortunately, provides a clearly nonzero superfluid fraction in the limit of simulating a single boson, which is obviously an unphysical artifact. The reason is that this estimator is based on computing the mean-squared area of paths in some plane, which is nonvanishing even for one boson where no exchange paths exist. As a remedy, the exchange estimator has been introduced.[394,395] It computes the so-called exchange superfluid fraction by essentially renormalizing the projected area obtained in BE simulations with the one computed from corresponding MB simulations, where no paths exchange and yet result in nonzero $\vec{n} \cdot \vec{A}$ contributions. In CP2K, the exchange estimator can be obtained after performing simulations identical to those sampling BE statistics, while switching off the MC permutation moves in the worm algorithm to provide the respective MB averages necessary to correct the area estimator (see Appendix C.1 in ref 396 for details).

The sampling scheme of the hybrid PIMD/bosonic PIMC (HPIMD/MC) algorithm[386,387,397] to treat quantum solvation of fully flexible and reactive molecular species in bosonic environments (such as those provided by ^{4}He), as available in the CP2K package, is presented in Figure 5.

In this approach, the molecular species and the bosonic solvent are sampled using PIMD and PIMC subject to permutation moves, respectively. After completing one PIMD step of the molecule, where the forces on the molecule are obtained from the intramolecular potential U_{mol}, as well as those coming from the solute−solvent interaction potential U_{int}, i.e.

$$\vec{F}_{PIMD} = -\vec{\nabla}_{\vec{r}_{mol}} U_{mol}(\{\vec{r}_{mol}\}) - \vec{\nabla}_{\vec{r}_{mol}} U_{int}(\{\vec{r}_{mol}\}, \{\vec{r}_{He}\}) \tag{142}$$

a bosonic PIMC simulation is then performed for the solvent species alone, for example ^{4}He. This includes the exchange moves to establish BE statistics of the bosonic solvent environment, while the molecule is kept frozen after the previous PIMD step, as illustrated in Figure 5. It is important to carry out sufficiently many bosonic PIMC steps such that the bosonic environment can adapt to the current configuration of the molecule in both the real and permutation space

before the next PIMD step is performed. The required bosonic PIMC chains are generated based on the change of the potential

$$U_{PIMC} = U_{He}(\{\vec{r}_{He}\}) + U_{int}(\{\vec{r}_{mol}\}, \{\vec{r}_{He}\}) \tag{143}$$

due to the solute−solvent interactions U_{int} together with the solvent−solvent pair potential interactions (such as $U_{He}(\{\vec{r}_{He}\})$ to describe helium[389]) as a evaluated during the Metropolis step. At this point, the bosonic permutations can be sampled in CP2K using either the bisection method,[349] or our canonical worm algorithm,[385,386] the latter being recommended for most cases. At the end of the PIMC chain, the next PIMD step is carried out using the new solvent position to compute $\vec{F}_{PIMD}$, followed by the next bosonic PIMC sweep, and so forth. This particular PIMD−PIMC coupling scheme and propagation algorithm establish the correct quantum NVT ensemble of the total solute-in-solvent system.[386] The HPIMD/MC approach implemented in CP2K has been generalized to also deal with molecular bosons, such as para-H_2 or ortho-D_2 molecules, in the framework of the adiabatic hindered rotor (AHR)[395] averaging technique.[398]

At this point, a decision has to be made with respect to the intramolecular potential energy surface U_{mol} and the solute−solvent interaction potential U_{int} to describe fully flexible and reactive molecular solutes in a quantum solvent, while an accurate solvent−solvent pair potential $U_{He}(\{\vec{r}_{He}\})$ is readily available in parametrized form.[389] In the original HPIMD/MC approach,[387] the ab initio PI philosophy[9,399,400] has been adapted to compute U_{mol} on-the-fly, while U_{int} has been fitted to accurate CC quantum chemistry data in the classic sense of a physics-based interaction potential;[401,402] this approach has been successfully applied to various systems[397,403,404] using CP2K. This technique has been superseded by the use of NNPs, as explained in Section 10.1, which offer two advantages at the same time, namely allowing for highly efficient PI sampling and providing high accuracy of the intra- and intermolecular interactions.[360] The current NNP-based HPIMD/MC approach in CP2K thus allows one to converge the PI even at ultralow temperatures (using techniques summarized in Section 10.3) and achieves essentially basis set converged CCSD(T) accuracy;[360] we note in passing that such CC quality PIMD simulations, dubbed CCMD, have been generalized more recently to condensed phase systems.[300,377]

The generation and use of NNPs in CP2K have been reviewed herein in Section 10.1, see there for general background and input options, which transfers to applications
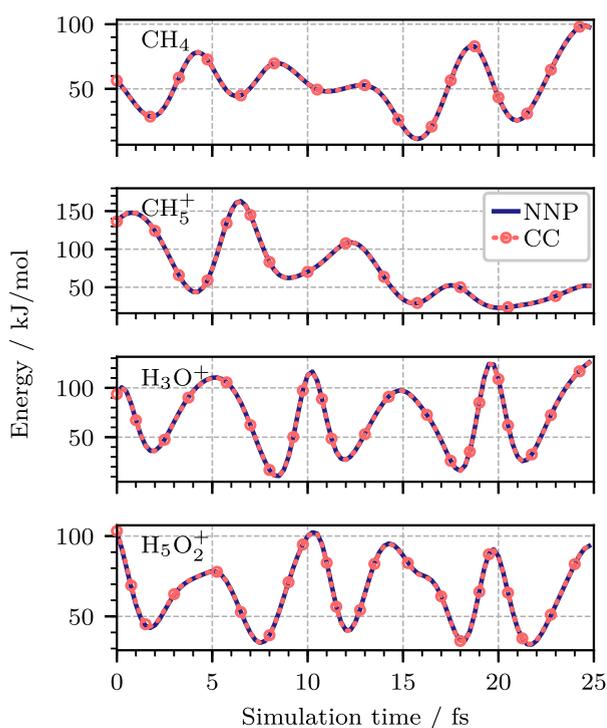
Figure 6. Potential energy U_{mol} along one Trotter replica of quantum PIMD trajectories of (from top to bottom) CH_4, CH_5^+, H_3O^+, $H_5O_2^+$, respectively, all at 1.67 K using NNPs trained with CCSD(T*)-F12a/aug-cc-pVTZ reference energies obtained from Molpro. The corresponding CC single-point data were obtained by recomputing the energies at each and every step of the depicted NNP trajectories and are shown as red dotted lines. All energies are reported relative to the equilibrium structures of the respective global minima. Reproduced from Figure 6 of ref 386. Copyright 2020, American Institute of Physics.

of the HPIMD/MC method. But in this case, additional challenges arise since the intramolecular potential energy surface of finite molecular systems, as well as intermolecular interaction potentials are required at very high ("quantum chemical") accuracy, while analytic gradients and thus forces are not efficiently available in the framework of CC theories. The situation is distinctly different from usual applications of MLPs in general and NNPs in particular to condensed phase systems, such as water and ice presented in Section 10.1, which are based on computationally economic periodic DFT calculations, where analytic forces are readily available.

These challenges posed by HPIMD/MC have been addressed using hierarchical MLP sampling approaches to cope with intermolecular interactions and thus solvation[405] combined with active learning protocols[298] to efficiently parametrize both U_{int} and U_{mol} in the framework of NNPs, which are deeply interfaced with CP2K. In particular, NNP training codes such as the RuNNer, n2p2 or RubNNet4MD packages[316] have been used in conjunction with quantum chemistry packages that offer efficient implementations of CCSD(T) theory, such as Molpro[406] and ORCA[407] to generate on the order of 10 000 to 100 000 total energies of finite molecular systems in the relevant size regime to train, test and validate the MLPs for use in CP2K-based HPIMD/MC and CCMD simulations.[298,300,360,396,398,405,408−414] The accuracy of the NNPs U_{mol} and interaction potentials U_{int} at the CCSD(T) level of theory for the methane molecule (CH_4),

the protonated methane complex (CH_5^+), the hydronium cation (H_3O^+), and the Zundel cation ($H_5O_2^+$) is illustrated in Figures 6 and 7, respectively. This rigorous end-to-end testing of the MLP representations of CCSD(T) references energies of diverse systems in terms of NNPs validates the accuracy of the overall methodology used in CP2K to perform HPIMD/MC and CCMD simulations.[298,300,360,396,398,405,408−414]

We note in passing that dipole moment surfaces of finite molecular systems have also been generated with CCSD(T) accuracy using a very similar systematic NN-based training approach,[415] as implemented in Version 2 of the RubN-Net4MD package.[316] Using CP2K, this enables the calculation of infrared (IR) absorption cross sections based on approximate RPMD (or TRPMD or CMD) quantum dynamics, as described in Section 10.4, in order to include NQEs in such vibrational spectra consistently at CCSD(T) quality.[414]

The HPIMD/MC method implemented in CP2K allows one to investigate fully flexible and reactive solutes, such as molecules, complexes or clusters, in finite and bulk-like bosonic environments, such as superfluid ^{4}He or para-H_2 species, at relevant temperatures on the order of 1 K (Figure 7). Here, a typical CP2K input section is provided to simulate a protonated water molecule embedded in a helium cluster consisting of 8 ^{4}He atoms (i.e., $H_3O^+ \cdot ^4He_8$) at a temperature of 1 K:

```
&MOTION
  &PINT
    DT 0.1
    NUM_STEPS 1000000
    P 160
    PROPAGATOR RPMD
    TEMP 1.0
    &HELIUM T
      CELL_SHAPE OCTAHEDRON

      CELL_SIZE 19.1157
      DROPLET_RADIUS 15.0
      GET_FORCES LAST
      NATOMS 8
      NBEADS 80
      N_OUTER 1000
      PERIODIC F
      POTENTIAL_FILE_NAME ./helium.potx
      PRESAMPLE T
      SAMPLING_METHOD WORM
      SOLUTE_INTERACTION NNP
      &NNP
        NNP_INPUT_FILE_NAME inter-input.nn
        SCALE_FILE_NAME inter-scaling.data
        &MODEL
          WEIGHTS inter-weights
        &END MODEL
```

AX

```
      &SR_CUTOFF
        ELEMENT "H"
        RADIUS   1.25
      &END SR_CUTOFF
      &SR_CUTOFF
        ELEMENT "He"
        RADIUS   0.00
      &END SR_CUTOFF
      &SR_CUTOFF
        ELEMENT "O"
        RADIUS   2.05
      &END SR_CUTOFF
     &END NNP
     &WORM
       ALLOW_OPEN T
       CENTROID_DRMAX  0.25
       MAX_OPEN_CYCLES 100
       OPEN_CLOSE_SCALE   50.0
       STAGING_L 6
     &END WORM
    &END HELIUM
    &PILE
       LAMBDA 0.5
       TAU 200
    &END PILE
   &END PINT
  &END MOTION
&FORCE_EVAL
  METHOD NNP
  &NNP
    NNP_INPUT_FILE_NAME solute-input.nn
    SCALE_FILE_NAME solute-scaling.data
    &MODEL
      WEIGHTS solute-weights
    &END MODEL
  &END NNP
  &SUBSYS
    &COORD
      O  0.00  0.00  0.00
      H  1.05  0.00  0.00
      H -0.49  0.77  0.00
      H -0.55 -0.51 -0.48
    &END COORD
  &END SUBSYS
&END FORCE_EVAL
```

This simulation uses the numerical ^{4}He$\cdots^4$He pair density matrix at 80 K obtained from the very accurate two-body interaction potential of Aziz to describe U_{He};[389] the resulting PIMC discretization of the bosonic environment relies on 80 replica. The employed intramolecular and intermolecular potentials U_{mol} and U_{int}, are NNP-based representations of reference energies at the CCSD(T) level of theory (obtained from the Molpro package[406]) that have been trained using the RubNNet4MD package[316] and are available from the CP2K GitHub repository. For PIMD sampling of the solute using the RPMD approach, a bead number of 160 is used. This relies on the striding approach, which allows one to couple two PIs with different Trotter discretizations, as visualized in Figure 8. In HPIMD/MC, the efficient pair density matrix approach allows one to converge the solvent PI using a rather small replica number (here 80 at 1 K), which is not sufficient to represent the solute PI. Using a stride of two and thus 160 replicas to represent the solute interactions, versus 80 for the solvent−solvent interactions and thus the solute−solvent interactions,
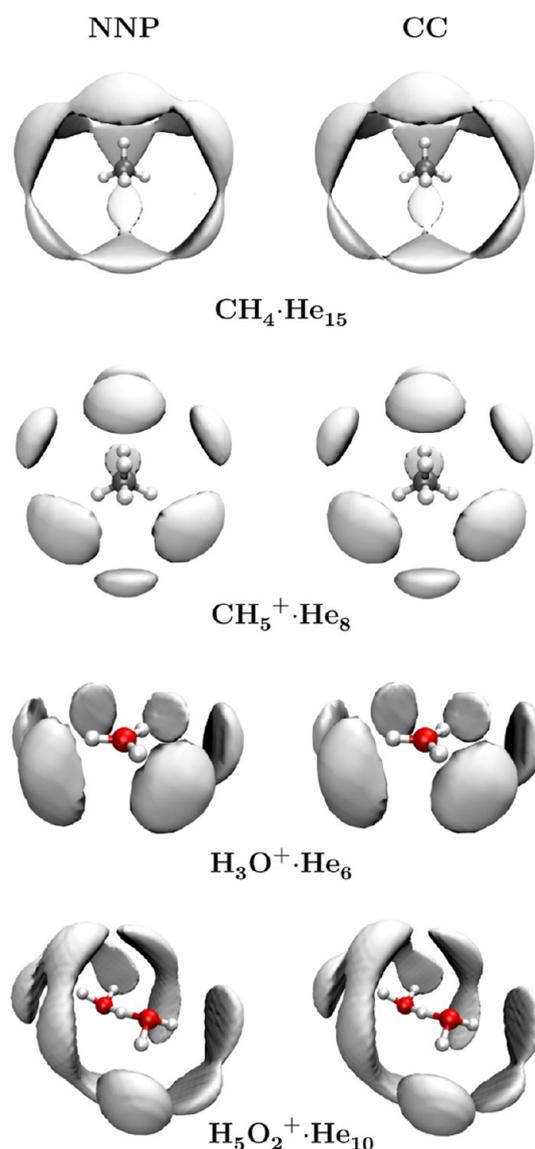

Figure 7. Spatial distributions functions stemming from all N helium atoms around a fixed molecular impurity (i.e., $X \cdot He_N$), as sampled from PIMC simulations (without bosonic exchange) at 1.67 K. The solute$\cdots$helium interaction energies U_{int} used in the PIMC sampling of the helium atoms are interpolated from values that have been precomputed on a grid using either the NNP (left column) or single-point CC calculations (right column) on the identical grid both based on counterpoise-corrected CCSD(T*)-F12a/aug-cc-pVTZ energy differences obtained from Molpro. Reproduced from Figure 7 of ref 386. Copyright 2020, American Institute of Physics.

as explained in section III.C of ref 386 and illustrated schematically in Figure 8, is appropriate. The intramolecular O−H, as well as the intermolecular O−He and H−He distance distribution functions resulting from the input example provided are presented in Figure 9. The corresponding superfluid fraction of the bosonic helium environment, as obtained with the exchange estimator, amounts to $f_s \approx 0.326 \pm 0.012$. This technique is available in CP2K and can be readily used to study the effects of bosonic quantum solvation on fully flexible and reactive molecular solute species embedded in ^{4}He and also in para-H$_2$ environments at temperatures on the order of 1 K,[398] where quantum exchange occurs and thus BE
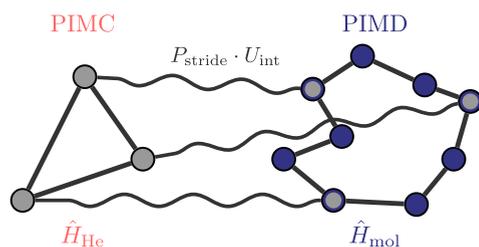
Figure 8. Illustration of PI striding for bosonic HPIMD/MC simulations, as implemented in CP2K: coupling of the PI representation of one helium atom (left) discretized with three Trotter replica (in practice sampled using PIMC based on the pair density matrix representation) to the PI representation of the solute molecule (right) discretized with nine beads (sampled using PIMD based on the one-body density matrix in conjunction with PIQTB thermostatting, as detailed in Section 10.3), thus corresponding to a striding length of three, $P_{\text{stride}} = 3$. Reproduced from Figure 3 of ref 386. Copyright 2020, American Institute of Physics.
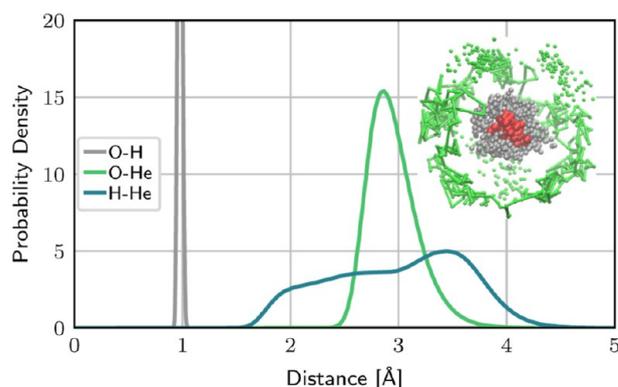
Figure 9. Distance distributions functions of $H_3O^+\cdot{}^4He_8$ at 1 K with BE sampling of the helium microsolvation environment. The intramolecular O−H, as well as the intermolecular O−He and H−He distance distributions are presented. The inset depicts the beads of the hydronium ion in terms of balls (red/gray for O/H), while all ^{4}He atoms involved in a bosonic exchange cycle are visualized by green strings and all others by green balls.

statistics emerge, resulting in strictly nonzero superfluid fractions and densities.[396,403,404,411,413]

10.6. Vibrational Spectroscopy

Vibrational spectroscopy is an important field of chemistry and physics since a very long time. As molecular vibrations (or phonons in solid state materials) are highly sensitive to the chemical environment of a certain molecule, vibrational spectra allow one to study the interactions present in the sample in great detail.

There exist two fundamentally different approaches for predicting vibrational spectra: the so-called "static-harmonic approximation", and the "time-correlation function (TCF)" techniques. Though CP2K directly implements the former approach, it is able to efficiently compute all necessary quantities of the latter method too, so that vibrational spectra can be easily obtained from postprocessing tools such as ASE[416] and TRAVIS.[417,418]

10.6.1. Static-Harmonic Approach. It is possible to predict the vibrational modes of an atomistic system by approximating it as a system of harmonic oscillators. Assuming a reasonably smooth potential energy surface, the potential energy $V(\mathbf{x})$ of a set of N atoms in close proximity to some

point $\mathbf{x}^0$ with respect to the atoms' $3N$ Cartesian coordinates $\mathbf{x} = (x_1, ..., x_{3N})$ can be approximately expressed in terms of a second-order multidimensional Taylor expansion

$$V(\mathbf{x}) \approx V(\mathbf{x}^0) + \sum_{i=1}^{3N}\left(\frac{\partial V}{\partial x_i}\right)_{\mathbf{x}^0}(x_i - x_i^0)$$
$$+ \frac{1}{2}\sum_{i=1}^{3N}\sum_{k=1}^{3N}\left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_{\mathbf{x}^0}(x_i - x_i^0)(x_k - x_k^0)$$

(144)

If the point $\mathbf{x}^0$ is an energy minimum, the gradients $\left(\frac{\partial^2 V}{\partial x_i}\right)_{\mathbf{x}^0}$ vanish, and only the second derivatives remain, i.e.

$$V(\mathbf{x}) \approx V(\mathbf{x}^0) + \frac{1}{2}\sum_{i=1}^{3N}\sum_{k=1}^{3N}\left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_{\mathbf{x}^0}(x_i - x_i^0)(x_k - x_k^0)$$

(145)

Based on this approximation, the force F_i acting on coordinate x_i can be expressed as

$$F_i = -\left(\frac{\partial V}{\partial x_i}\right) = -\frac{1}{2}\sum_{k=1}^{3N}\left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_{\mathbf{x}^0}(x_k - x_k^0)$$

(146)

By inserting eq 146 into Newton's equation

$$F_i = m_i\, a_i = m_i\left(\frac{d^2 x_i}{dt^2}\right), \qquad i = 1, ..., 3N$$

(147)

one obtains a system of equations of motion

$$m_i\left(\frac{d^2 x_i}{dt^2}\right) = -\sum_{k=1}^{3N}\left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_{\mathbf{x}^0}(x_k - x_k^0)$$

(148)

To express these equations more compactly, it is desirable to switch to the set of so-called mass-weighted Cartesian coordinates $\mathbf{q} = (q_1, ..., q_{3N})$ defined as

$$q_i := x_i\sqrt{m_i}$$

(149)

in which eq 148 now reads as

$$\frac{d^2 q_i}{dt^2} = -\sum_{k=1}^{3N} H_{i,k}\, q_k, \qquad i = 1, ..., 3N$$

(150)

with the short-hand notation $H_{i,k}$ for the mass-weighted Hessian matrix

$$H_{i,k} := \frac{1}{\sqrt{m_i m_k}}\left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_{\mathbf{x}^0}$$

(151)

As the above approximation defines a purely (see eq 145), one can assume that the motion of each coordinate q_i can be described by a sinusoidal time evolution

$$q_i(t) = q_i^0 + A\cdot\cos(\omega t)$$

(152)

with some amplitude A and angular frequency ω. Inserting this into eq 150 yields

$$\omega^2 q_i = \sum_{k=1}^{3N} H_{i,k}\, q_k, \qquad i = 1, ..., 3N$$

(153)

which can be rewritten as a matrix eigenvalue problem for the mass-weighted Hessian matrix $\mathbf{H}$ with the eigenvalues ω^2

$$\omega^2 \mathbf{q} = \mathbf{Hq} \qquad (154)$$

In other words, solving this eigenvalue problem directly yields the vibrational frequencies ω of the atomic system as the square roots of the eigenvalues. Furthermore, the eigenvectors $\mathbf{q}$ of the matrix represent the directions in which the atoms are displaced within each normal mode. Therefore, a full set of normal modes, each with a vibrational frequency and an atom displacement vector, can be obtained via this approach. Additional care has to be taken to project out the invariants (i.e., translational and rotational invariances due to the conservation of momentum and angular momentum) of the mass-weighted Hessian matrix, so that typically only $(3N - 6)$ normal modes are obtained.

Such a normal-mode analysis is fully implemented in CP2K. Yet, as explained above, the approach only works for minimum structures, so it should be applied only to geometries that resulted from a converged optimization with very similar computational details. The normal-mode analysis can be requested as follows:

```
&GLOBAL
    RUN_TYPE VIBRATIONAL_ANALYSIS
&END GLOBAL
&VIBRATIONAL_ANALYSIS
    DX 0.01
    FULLY_PERIODIC TRUE
    INTENSITIES TRUE
&END VIBRATIONAL_ANALYSIS
```

Note that CP2K computes the second derivatives of the energy with respect to the displacement (Hessian matrix) numerically via central finite differences, so that for a system of N atoms, $6N$ individual energy and force computations will be required. The DX parameter sets the displacement for the finite differences. Please also note that a very tight SCF threshold (EPS_SCF 1.E-8 in combination with EPS_DEFAULT 1.E-16) is required to reduce numerical noise. The FULLY_PERIODIC keyword inside the VIBRATIONAL_ANALYSIS section should be used for periodic bulk phase systems in order to switch off the pruning of the rotational degrees of freedom from the Hessian. Activating the keyword INTENSITIES facilitates the computation of IR or Raman intensities, but requires the explicit calculation of dipoles or polarizabilities by means of &DFT%PRINT%MOMENTS or &PROPERTIES%LINRES%POLAR, which are described in detail later. Moreover, adding the subsection &MODE_SELECTIVE permits running a so-called mode selective vibrational analysis, which constrains the calculation to either a single FREQUENCY, RANGE of frequencies, or a list of INVOLVED_ATOMS.

The derivation above only shows how to obtain the vibrational frequencies of the normal modes. However, in order to predict real vibrational spectra, the intensities of each mode also need to be computed. This is typically performed by computing derivatives of some more or less complicated electromagnetic properties with respect to the atom displacements $\mathbf{q}$ obtained for each mode. For example, the IR intensity of a certain mode is proportional to the change in the electric dipole moment that occurs when the atoms are displaced along that mode. More details on computing such properties are discussed in the following subsection.

Despite still being the standard approach in the literature, computing spectra via the static-harmonic approach comes with several shortcomings:

- Due to the harmonic approximation of the potential energy surface, all anharmonic effects are neglected.[419] If the system possesses features such as strong hydrogen bonds or hindered rotations, the harmonic approximation of certain modes will be poor, and so will the quality of the predicted spectrum.
- The spectrum can only be computed for one minimum energy structure at a time. If there exist several conformers of the same molecule, they need to be considered separately. If the system can hardly be described by minimum-energy structures, such as bulk phase liquids, it will be hard to obtain reasonable spectra from the outset.
- The method works best for molecules or small clusters in vacuum. Solvent effects on the spectrum can be crudely approximated either via continuum solvation models such as COSMO,[420] PCM,[421] or the SCCS implicit solvation method described in Section 4.1,[135−137] as well as by microsolvation, but the solvent effect cannot be captured comprehensively.
- The approach only yields a discrete line spectrum; no line widths or band shapes can be obtained. Hence, to predict realistic spectra, empirical line broadening needs to be applied.

10.6.2. Time-Correlation Function Approach. Apart from employing the static-harmonic approximation, there exists the possibility to compute vibrational spectra directly from MD simulations. In this approach, the spectra are obtained as the Fourier transform of some TCFs along the simulation trajectory, an idea that is at least 60 years old now,[422−424] and is called the TCF formalism. This approach comes with several advantages over the static-harmonic concept:

- Condensed phase systems can be handled; it is possible to explicitly capture the effects of solvent and entropy on the spectrum.
- Some anharmonic effects, such as line broadening, approximate overtones, and combination bands, are reproduced.
- Realistic band shapes are obtained instead of a discrete line spectrum.
- Intrinsic conformer sampling takes place during the MD simulation.
- No minimum energy structure is required to compute the spectrum.

As CP2K has its main strengths in the description of condensed phase systems, including AIMD, it is particularly well suited to compute vibrational spectra via the TCF formalism. Therefore, the remaining part of this subsection will focus exclusively on this approach. The basic workflow is simple: perform an AIMD simulation, compute some electromagnetic properties along the trajectory, and use some postprocessing software to obtain a spectrum from that time series.

10.6.3. Computing Electromagnetic Moments. Computing vibrational spectra, both via the static-harmonic approach and via the TCF approach, requires knowledge of certain electromagnetic moments of the system, e.g. the electric dipole moment. In the former case, these moments are calculated for excursions of the minimum structure along the normal modes, but they are computed for snapshots along an MD simulation trajectory in the latter case. For nonperiodic

systems, electric moments can be readily derived as expectation values from electron structure calculations by applying the corresponding moment operator to the converged WF.

Under PBCs, however, the standard moment operators are ill-defined, as they often rely on the position operator. The challenge can be easily seen by, e.g. considering the total dipole moment of a simple chain of anions and cations in a periodic cell: the dipole moment depends on the cell origin, but it should not.[425] Resta and King-Smith and Vanderbilt proposed a possible solution, the so-called modern theory of polarization,[426−429] which is based on a Berry phase.[430,431] In short, the ill-defined position operator $\hat{X}$ for the WF Ψ_0 under PBCs is replaced by the following expectation value for the position[429]

$$\langle X \rangle = \frac{L}{2\pi} \text{Im} \ln \left\langle \Psi_0 \left| \exp\left(i\frac{2\pi}{L}\hat{X} \right) \right| \Psi_0 \right\rangle \tag{155}$$

Due to the ambiguity of the complex logarithm function, there exist infinitely many valid results from this equation, all equidistantly separated. This means that the dipole moment is also no longer uniquely defined under PBCs, but only up to a modulus. However, for most applications, this approach works very well.

In CP2K/QUICKSTEP, the total electric dipole moment from the modern theory of polarization can be computed and printed by adding the following section to the input:

```
&FORCE_EVAL
    &DFT
        &PRINT
            &MOMENTS
            &END MOMENTS
        &END PRINT
    &END DFT
&END FORCE_EVAL
```

Note that despite the general name of the `&MOMENTS` subsection and the availability of a `MAX_MOMENT` keyword, it is implemented only for electric dipole moments in the periodic case.

10.6.4. Total vs Molecular Moments. There are several reasons for considering molecular moments instead of system-wide "total" electromagnetic moments. First, it allows for separation of the spectral contributions of the constituents of a mixture, so that, e.g., the solvent spectrum can be suppressed. Second, the sampling of the spectrum is improved and the spectrum contains less noise when the TCF approach is used, as will be explained on the example of the dipole moment in the following. When assuming that the sum of the molecular dipole moments equals the total dipole moment, which is an approximation in the periodic case although the deviation is small,[432] it can be shown that using the total dipole moment is mathematically equivalent to using all autocorrelations of the individual molecules' dipole moments together with all cross-correlations between different molecules. It is often argued that cross-correlations between molecules that are not neighbors in space only introduce noise to the spectrum, as the motions of such molecules are not at all correlated. Therefore, it is often a good idea to consider only molecular autocorrelation functions and to neglect all cross-correlation terms between different molecules.

10.6.5. Orbital Localization. One widely used approach to assign electric dipole moments to individual molecules is the localization of MOs in space.[433,434] For nonperiodic systems,

there exist some well-known and computationally efficient methods such as the Boys,[435] as well as the Pipek−Mezey localization schemes,[436] among others.

In periodic systems, orbital localization is considerably more involved. One commonly used method is the so-called Wannier localization.[437−442] It applies a unitary transformation $\mathbf{U}$ to the set of occupied KS orbitals $|\psi_i\rangle$ so that another set of MOs $|\tilde{\psi}_n\rangle$ is obtained, which are called Wannier orbitals or MLWFs[443]

$$|\tilde{\psi}_n\rangle = \sum_i U_{i,n}|\psi_i\rangle \tag{156}$$

The unitary transformation $\mathbf{U}$ is constructed in a way so that the so-called spread functional

$$\mathbf{\Omega} = \sum_n \sum_I f(|z_{I,n}|^2) \tag{157}$$

$$z_{I,n} = \langle \psi_n | \mathbf{O}^I | \psi_n \rangle \tag{158}$$

is minimized.[443] Here, $\mathbf{O}^I$ is a class of suitable spread operators that are well-defined in periodic space, such as

$$\mathbf{O}^I = \exp(i\mathbf{G}_I \cdot \mathbf{r}) \tag{159}$$

where $\mathbf{G}_I$ are the reciprocal lattice vectors of $I = x, y, z$ and f an appropriate function. Common choices for f are[438,440,443,444]

$$f_1(|z_{I,n}|^2) = \sqrt{|z_{I,n}|^2} = |z_{I,n}| \tag{160}$$

$$f_2(|z_{I,n}|^2) = \log(|z_{I,n}|^2) \tag{161}$$

$$f_3(|z_{I,n}|^2) = |z_{I,n}|^2 \tag{162}$$

Note that both the Boys-Foster localization[435] and the Pipek−Mezey localization[436] for nonperiodic systems can be expressed in terms of the above equations with the choice of $f \equiv f_3$. For Boys-Foster, the operator $\mathbf{O}^I$ is simply defined as $\mathbf{O}^I = \mathbf{r}_I$ with the conventional position operator $\mathbf{r}_I$.

One traditionally applied approach in quantum chemistry for localizing MOs is the method of two-by-two orbital rotations first introduced by Edmiston and Ruedenberg.[445] Unfortunately, the analytical expression for the optimal angle of these rotations can only be derived for the choice of $f \equiv f_3$,[443] i.e. for the Silvestrelli-Marzari-Vanderbilt,[438−440] the Boys-Foster,[435] and the Pipek−Mezey functional.[436] For the choice of $f \equiv f_1$, which is commonly used for performing Wannier localization, one has to resort to iterative numerical methods such as a generalized Jacobi rotation scheme,[446] or CP2K's own so-called "crazy angle" algorithm. Yet, all of these methods require considerable amounts of extra computer time for the localization and are not guaranteed to converge. A discussion of these limitations can be found in the next subsection. A practical alternative to bypass the computational effort to localize the MOs in every AIMD time step is to propagate the matrix $\mathbf{U}$ of eq 156 by means of a second-generation Car−Parrinello-like "Wannier dynamics".[11,447] In CP2K this is enabled by setting `&LOCALIZE%USE_HISTORY` to `TRUE`.[448,449]

The centroids of the Wannier orbitals are called Wannier centers, or maximally localized Wannier centers (MLWCs); they can be seen as the positions of electron pairs in a simple picture. As those are located relatively closely to the atoms, it is well possible to assign Wannier centers to individual molecules.

Based on these Wannier centers, the molecular dipole moment μ^{Mol} can be expressed as

$$\mu^{\mathrm{Mol}} = -2e \sum_{i=1}^{N} \mathbf{r}_i + e \sum_{j=1}^{M} Z_j \mathbf{R}_j \tag{163}$$

where N is the number of Wannier centers in the molecule, $\mathbf{r}_i$ is the position of the i-th Wannier center, M is the number of atoms in the molecule, whereas $\mathbf{r}_j$ and Z_j are the position and the nuclear charge of the j-th atom, respectively, and e is the elementary charge.[432] The sum of all molecular dipole moments computed by this protocol is often a good approximation to the total dipole moment of the system.

In CP2K/QUICKSTEP, the Wannier localization of the MOs can be requested by adding the following section:

```
&FORCE_EVAL
    &DFT
        &LOCALIZE
        &END LOCALIZE
    &END DFT
&END FORCE_EVAL
```

The following keywords inside the &LOCALIZE subsection are most relevant for practical use:

- METHOD: how to minimize the spread functional. Default is JACOBI (slow, but very robust), but should be set to CRAZY (a lot faster, but less robust) in most cases. See next keyword.
- JACOBI_FALLBACK: do a JACOBI minimization ("fallback"), if CRAZY did not converge.
- MAX_ITER: maximum number of iterations for minimizing the spread functional. Default is 10 000, which is almost always enough. Maybe reduce it to not waste time if it is not going to converge anyway.

Several types of results can be obtained from the localization procedure. This can be defined via subsections inside the &LOCALIZE%PRINT section. The following subsections are often useful:

- &WANNIER_CENTERS: outputs the "Wannier centers", i.e. the centroids of the MLWFs to a XYZ file with the specified name. Also works in AIMD runs, where it just writes a consecutive Wannier center trajectory. If the IONS+CENTERS keyword is set, the atom coordinates are added to that file, which can make analysis easier.
- &WANNIER_SPREADS: outputs the remaining spread of the Wannier orbitals. This can be useful to, e.g. approximate polarizabilities,[450−452] as shown below.
- &WANNIER_CUBES: outputs the actual Wannier orbitals on a Cartesian grid as cube files. Useful for visualization.
- &WANNIER_STATES: outputs the Wannier orbitals in the basis of the AOs, i.e. the corresponding coefficient matrix.

10.6.6. Voronoi Integration. The Wannier localization comes with a few well-understood shortcomings:

- the CPU time required for localization can be substantial and scales unfavorably with system size. For a 1000 atom system, the localization typically takes almost as long as the energy and force calculation, i.e. slowing down the simulation by approximately a factor of 2.
- All known methods for computing the Wannier localization are iterative and are not guaranteed to

converge. It happens in practice that for certain AIMD timesteps, neither the crazy angle algorithm, nor the Jacobi fallback converge, so that the electric moments are missing for these timesteps.

- Only molecular electric dipoles can be derived from the Wannier centers; higher-order momenta such as quadrupoles, as for instance required for Raman optical activity (ROA) spectra, are not directly accessible.
- Wannier localization enforces integer molecular charges. Any charge transfer effects between molecules cannot be captured, which leads to artificially increased dipole moments in some cases.
- Systems with delocalized electrons, such as aromatic molecules or metals, may entail convergence issues. Also, for example, the simulated IR spectrum of liquid benzene contains artificial bands that are a consequence from the Wannier localization.[453]

To avoid the computationally demanding localization procedure, one simple idea is to work with the total electron density on a Cartesian grid instead, which is always present in any PW-based electronic structure code including the GPW/GAPW method of CP2K/QUICKSTEP, and to partition it with respect to the atom positions to obtain molecular electromagnetic moments. Even though there exist many such approaches, we will discuss here the so-called Voronoi integration.[453,454]

The Voronoi tessellation is a mathematical tool which partitions an Euclidean space containing some points (Voronoi sites) into nonoverlapping subsets.[455,456] Each Voronoi site corresponds to exactly one such subset, known as a Voronoi cell, which contains all spatial points that are closer to this Voronoi site than to any other Voronoi site. In mathematical form, this is written as

$$C_i := \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{p}_i\| \le \|\mathbf{x} - \mathbf{p}_j\|$$

$$\forall j \in \{1 \ldots k\}, \ j \ne i\}, \quad i \in \{1 \ldots k\} \tag{164}$$

where $\mathbb{R}^n$ stands for any Euclidean space with the norm $\| \ \|$, in which k Voronoi sites, each with position $\mathbf{p}_i \in \mathbb{R}^n$, are given, and the $C_i \subseteq \mathbb{R}^n$ are the resulting Voronoi cells.

By considering atoms in three-dimensional space as Voronoi sites, this concept has widely been applied in different fields of computational chemistry. To name a few advantages of the method, the Voronoi tessellation of a set of atoms is uniquely defined and can be calculated with moderate computational efforts. The Voronoi tessellation can easily be adapted to systems with PBCs, and is therefore well suited for bulk phase systems. Finally, the method does not possess any empirical parameters to tune, and therefore gives a uniquely defined picture.

Voronoi tessellation has already been used since long to partition the total electron density, by placing a simple plane midway between two atoms.[457,458] However, certain limitations do arise from the properties of the standard Voronoi tessellation. As all atoms are treated in the same way, Voronoi polyhedra of light atoms like hydrogen will, on average, have the same size as those around heavier atoms like iodine. From a mathematical point of view, this is not a problem, but from a chemical perspective, this is completely unreasonable. If, for instance, the electron density within the Voronoi cell of a hydrogen atom is integrated, the hydrogen atom would always end up with a heavily negative partial charge, because way too

much electron density would be considered as belonging to this hydrogen atom.

To overcome this problem, radii need to be introduced into the Voronoi tessellation, allowing to treat different atom types differently. Several ways to do so have been proposed. Many such approaches have been proposed in the literature. Here, we employ the so-called "radical Voronoi tessellation", which, in the two-dimensional case, is also known as power diagram.[459] Note that the term "radical" is not related to chemical radicals. In this technique, a radius is assigned to each atom, allowing to model the sizes of the atoms in a chemically reasonable sense. In contrast to other similar approaches,[459,460] the radical Voronoi tessellation does not suffer from the "vertex error", i.e. it does not contain holes. When integrating electron density, this is important to keep the total charge of the system constant. As another advantage, the Voronoi sites, around which the cells are constructed, can be kept on the atoms and do not have to be shifted to obtain a chemically reasonable partitioning.

The definition of the radical Voronoi tessellation as a generalization of the classical tessellation reads as

$$C_i^r := \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{p}_i\|^2 - r_i^2 \leq \|\mathbf{x} - \mathbf{p}_j\|^2 - r_j^2$$

$$\forall j \in \{1 \ldots k\}, j \neq i\}, \quad i \in \{1 \ldots k\} \quad (165)$$

with radius r_i for Voronoi site i. While in the classical case the face between two adjacent Voronoi cells is always placed in the middle between the corresponding Voronoi sites, its position is now determined by the difference of the squared radii. The definition of the radical Voronoi tessellation in eq 165 shows that the tessellation will not change if the set of radii $\{r_i\}$ is transformed to a new set $\{r_i'\}$ by the map

$$r_i' := \sqrt{r_i^2 + C}, \quad i \in \{1 \ldots k\} \quad (166)$$

with some constant $C \in \mathbb{R}$. Due to this relation, the absolute value of the radii does not have a direct meaning.

The crucial parameters in the radical Voronoi tessellation are the radii assigned to the atoms. It was recently shown that vdW radii yield a reasonable separation of molecules in the bulk phase,[453] and that the resulting molecular electromagnetic "Voronoi" moments can readily be used to calculate vibrational spectra of bulk phase systems from AIMD simulations.[461−463]

As soon as the Voronoi tessellation has been computed, the molecular electromagnetic moments can be obtained via simple integration of the total electron density ρ within each molecule's Voronoi cell, e.g. for the electric dipole and quadrupole moment

$$\boldsymbol{\mu}^{\text{Mol}} = \sum_{i=1}^{N_{\text{Mol}}} q_i \mathbf{r}_i - \int_{\text{Mol}} d\mathbf{s}\, \rho(\mathbf{s})\mathbf{s} \quad (167a)$$

$$\mathbf{Q}_{jk}^{\text{Mol}} = \sum_{i=1}^{N_{\text{Mol}}} q_i (3\mathbf{r}_{i,j}\mathbf{r}_{i,k} - \|\mathbf{r}_i\|^2 \delta_{jk})$$

$$- \int_{\text{Mol}} d\mathbf{s}\, \rho(\mathbf{s})(3\mathbf{s}_j \mathbf{s}_k - \|\mathbf{s}\|^2 \delta_{jk}) \quad (167b)$$

where $\mathbf{r}_i$ and q_i are the position and core charge of atom i, respectively.

The Voronoi integration in CP2K is provided by the `libvori` library,[464] which internally relies on `Voro++`.[465] It can be switched on by adding the following section to the input:

```
&DFT
  &PRINT
    &VORONOI
    &END VORONOI
  &END PRINT
&END DFT
```

The following keywords inside of the &VORONOI section are most relevant for practical use:

- VORONOI_RADII: chooses the radii for the radical Voronoi tessellation. Can be set to UNITY (all radii are 1, i.e. classical Voronoi), VdW (vdW radii are used), COVALENT (covalent radii are employed),[466] or USER. Default is VdW.

- USER_RADII: in case of VORONOI_RADII USER, specifies the radii (one number per atom in the system).

- OUTPUT_TEXT: outputs a .voronoi text file with all properties in a human-readable format. The file name can be set via FILENAME. On by default.

- OUTPUT_EMP: outputs a binary .emp file with all electromagnetic properties. Can be read by TRAVIS[418,467] to compute vibrational spectra. Off by default.

- JITTER: randomly displaces all Voronoi sites a tiny bit to avoid problems with highly symmetric structures. The amount can be controlled by the keyword JITTER_-AMPLITUDE. On by default.

The resulting quantities obtained from the Voronoi integration are per atom. For each atom, the Voronoi charge, the center-of-charge vector, the electric dipole vector, and the trace-free electric quadrupole tensor are printed. If molecular properties are required, they can easily be combined as follows by looping over all atoms of the molecule

$$q^{\text{Mol}} = \sum_{i=1}^{N_{\text{Mol}}} q^i \quad (168a)$$

$$\boldsymbol{\mu}^{\text{Mol}} = \sum_{i=1}^{N_{\text{Mol}}} \boldsymbol{\mu}^i + q^i(\mathbf{r}^i - \mathbf{r}^{\text{Ref}}) \quad (168b)$$

$$\mathbf{Q}_{jk}^{\text{Mol}} = \sum_{i=1}^{N_{\text{Mol}}} \mathbf{Q}_{jk}^i + \boldsymbol{\mu}_j^i(\mathbf{r}_k^i - \mathbf{r}_k^{\text{Ref}}) + \boldsymbol{\mu}_k^i(\mathbf{r}_j^i - \mathbf{r}_j^{\text{Ref}})$$

$$+ q^i(\mathbf{r}_j^i - \mathbf{r}_j^{\text{Ref}})(\mathbf{r}_k^i - \mathbf{r}_k^{\text{Ref}}), \quad (168c)$$

where $\mathbf{r}^i$, q^i, $\boldsymbol{\mu}^i$, and $\mathbf{Q}^i$ are the position, Voronoi charge, electric dipole moment, and electric quadrupole tensor of atom i, respectively. The choice of the coordinate origin $\mathbf{r}^{\text{ref}}$ leaves the molecular dipole moment invariant as long as the molecular Voronoi charge is zero, and leaves the molecular quadrupole moment invariant as long as the molecular dipole moment is zero. Otherwise, these quantities depend on the choice of a reference point.

10.6.7. Polarizabilities. After discussing the electric dipole moment, we cover a second type of important electromagnetic property, namely the electric polarizability, or more correctly, the static electric dipole-electric dipole polarizability. In contrast to the dipole moment, it is a response property and therefore can not be directly obtained from the WF of the system.

One straightforward approach to electric polarizabilities is to use finite differences with respect to an external electric

field.[173] In linear approximation, the dipole moment $\boldsymbol{\mu}_{\text{ind}}$ induced by an electric field $\mathbf{E}$ can be expressed as

$$\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\alpha}\mathbf{E} \tag{169}$$

with the second-order electric polarizability tensor $\boldsymbol{\alpha}$. This leads to the central finite differences

$$\alpha_{i,j} = \frac{\mu_i^{j+} - \mu_i^{j-}}{|\mathbf{E}^{j+} - \mathbf{E}^{j-}|}, \quad i, j = x, y, z \tag{170}$$

where $\mathbf{E}^{j+}$ and $\mathbf{E}^{j-}$ are the field vectors of the external electric field applied in the positive and negative j direction, respectively. Moreover, μ_i^{j+} and μ_i^{j-} are the i components of the dipole moment under the influence of these two fields, and $\alpha_{i,j}$ is the (i, j) component of the polarizability tensor. By performing six additional SCF calculations with positive and negative fields in the x, y, and z direction, the full polarizability tensor can thus be obtained. The strength of the electric field $|\mathbf{E}|$ needs to be chosen so that the system is still within the linear regime of polarizability, i.e. eq 169 is still a good approximation. Reasonable choices for organic liquids were found to be $|\mathbf{E}| \approx 5.0 \cdot 10^{-4}$ a.u. $= 2.57 \times 10^{8}$ V m^{-1},[453] or even up to 5.0×10^{-3} a.u. $= 2.57 \times 10^{9}$ V m^{-1}.[461] If one uses molecular electric dipole moments in these calculations, e.g. from Wannier centers or Voronoi integration, the molecular polarizability tensor for each molecule in the system can be obtained.

In CP2K/QUICKSTEP, a periodic electric field, i.e. without any discontinuities and suitable for bulk phase systems, can be applied using the &PERIODIC_EFIELD section. See the following example for a field in the positive x-direction, as controlled by the POLARIZATION keyword:

```
&FORCE_EVAL
  &DFT
    &PERIODIC_EFIELD
      INTENSITY 5.0E-3
      POLARISATION 1.0 0.0 0.0
    &END PERIODIC_EFIELD
  &END DFT
&END FORCE_EVAL
```

However, &PERIODIC_EFIELD requires certain derivatives of the MOs, which are only available via the OT method, so that electric polarizabilities from finite differences can currently not be obtained with diagonalization-based mixing schemes, or with electronic smearing.

When molecular polarizabilities are computed according to eq 170, the changes in the local electric field of a molecule by the polarization of the neighboring molecules are omitted. This effect can be captured by considering the dipole−dipole interaction tensor computed by Ewald summation under PBCs,[468] as explained in refs 469,470, respectively. However, a recent study of water has shown that this has only a minor influence on the resulting spectra.[471]

Another approach to electric polarizabilities is the previously described variational DFPT. Originally proposed and implemented in the CPMD code by Putrino and Sebastiani,[167,173] it is also available in CP2K.[472] It does not suffer from possible nonlinearities, such as in finite difference approaches, and is less prone to numerical noise, but it can be slow and numerically less stable, and it does not allow direct access to molecular polarizability tensors. Such a calculation can be requested by using the following input:

```
&FORCE_EVAL
  &PROPERTIES
    &LINRES
      &POLAR
      &END POLAR
    &END LINRES
  &END PROPERTIES
&END FORCE_EVAL
```

However, there exist further approaches to computing molecular polarizabilities, such as one by Partovi-Azar and Kühne to approximate the polarizability tensor based on the spatial spread of the Wannier centers.[450,451]

10.6.8. Frequency-Dependent Polarizabilities. The electric polarizability is a frequency-dependent property. For many applications, this dependency is completely neglected, and the static (i.e., zero-frequency limit) polarizability is used, such as in predicting Raman spectra. But, such effects need to be explicitly considered when predicting the correct intensities in resonance Raman spectra, for instance. In such cases, one needs to compute the dynamic (i.e., frequency-dependent) polarizability. Doing so turned out to be quite intricate. As the dynamic polarizability involves electronic excitations, it is generally not sufficient to solve the time-independent Schrödinger equation to obtain it. Several approaches to compute this property have been presented in the literature. Many of them are based on the vibronic theory of Albrecht and co-workers,[473−475] or on the time-dependent formalism of Heller and co-workers.[476−478] Another method, based on LR-TDDFT, was published by Jensen and Schatz.[479,480]

Recently, a different approach has appeared in the literature that uses RT-TDDFT[481−483] to obtain the dynamic polarizability of the sample.[484] In contrast to the methods mentioned above, the real-time approach offers the advantage of including all electronic excitations into the calculation, so that the full frequency range is covered and no subset of low-lying excitations needs to be selected. Furthermore, this approach intrinsically includes nonlinear effects that are neglected in perturbative methods such as LR-TDDFT. And last but not least, it allows access to molecular polarizability tensors in bulk phase systems, which is otherwise not possible.

Based on this approach, a protocol for the calculation of molecular dynamic polarizability tensors in condensed phase systems using CP2K has been developed and published,[462] which can be used in the TRAVIS program package.[418,467] The protocol automatically creates all required CP2K input files and works as follows. The initial WF is optimized under the influence of an external periodic electric field, which is switched off in the beginning of the RT-TDDFT run, so that the electron density starts to fluctuate (step response). During the RT-TDDFT run, the temporal development of the total electron density is processed with the Voronoi integration scheme to yield time series of molecular electric dipole vectors $\boldsymbol{\mu}(\tau)$ as a function of RT-TDDFT time τ. The Fourier transform of the three dipole vector components yields three entries of the molecular dynamic polarizability tensor. To obtain the full tensor $\alpha_{ij}(\omega)$ for each molecule (with ω the incident laser frequency), three RT-TDDFT runs are performed from initial WFs optimized under external fields in the x, y and z directions

$$\alpha_{ij}(\omega) = \frac{1}{|\mathbf{E}|} \int_0^T d\tau \, (\mu_i(\tau) - \mu_i^{0,j})$$

$$\times \exp\!\left(\left(-c\frac{\tau}{T}\right)^2\right)\exp(-i\omega\tau) \tag{171}$$

Please note the use of a Gaussian window function with parameter c for the Fourier transform of the RT-TDDFT time series. Furthermore, T is the total RT-TDDFT simulation time, $\boldsymbol{\mu}^{0,j}$ denotes the initial molecular dipole moment after WF optimization under an external electric field in j direction, and $|\mathbf{E}|$ is the absolute value of the external electric field. Please also note that the dynamic polarizability tensor obtained from eq 171 is complex-valued, with dispersion as the real part and absorption as the imaginary part. Speaking of polarizabilities typically refers to the dispersion, i.e. the real part.

10.6.9. Supported Types of Spectra. Computing time series of electromagnetic properties (dipole moment, polarizabilities, etc.) by means of AIMD permits obtaining vibrational spectra by a postprocessing tool. The basic idea is simple: calculate the required cross-correlation functions of the electromagnetic properties along the simulation trajectory, and then Fourier transform these correlation functions to yield the spectrum. In practice, there are several more subtleties and tricks that should be considered to obtain high-quality spectra. Examples are signal processing techniques such as applying window functions and zero padding, but also exploitation of time-reversal symmetry to enhance sampling or finite-difference correction. Using the TRAVIS postprocessing tool,[418,467] IR, Raman, vibrational circular dichroism (VCD), Raman optical activity (ROA) and resonance Raman spectra can be easily computed for periodic condensed phase systems.

10.6.9.1. Storing Compressed Density Cubes. Many of the electromagnetic properties described above can be obtained from the total electron density of a periodic system via Voronoi integration, which can be performed on-the-fly during the CP2K AIMD run. If for some reason the electron density data shall be kept along the trajectory, huge trajectories of Gaussian cube files, which are in the order of many terabytes, would result.[485] For that purpose, CP2K offers to write these volumetric electron density trajectories in bqb file format,[486] which utilizes a lossless compression algorithm that reaches a compression ratio of up to 40:1 by exploiting both the spatial and temporal smoothness of the electron density time series. To switch on this feature, use an input similar to the following example:

```
&DFT
   &PRINT
      &E_DENSITY_BQB
         FILENAME result.bqb
      &END E_DENSITY_BQB
   &END PRINT
&END DFT
```

Apart from trajectories, single cube files can also be compressed by this approach, albeit with a slightly lower compression ratio. The resulting `.bqb` files can be read, processed, and unpacked by the TRAVIS program package.[418,467]

10.6.9.2. Normal Mode Analysis from Ab Initio Molecular Dynamics. When computing vibrational spectra in the static-harmonic approximation via `RUN_TYPE VIBRATIONAL_ANALYSIS` as shown above, each spectral band is obtained

together with a corresponding normal mode so that the spectral features can be easily assigned to specific molecular vibrations. However, when computing spectra from AIMD simulations via the TCF formalism, this cannot so easily be achieved because the spectrum is obtained as a superposition of all vibrational modes in the system without additional information on specific molecular motions that contribute to a certain spectral feature.

To overcome this limitation, several approaches for the extraction of normal modes from MD simulations have been reported in the literature.[487,488] These approaches include the instantaneous normal-mode analysis (INMA), where the Hessian of the system is calculated in certain timesteps along the trajectory.[489−491] Moreover, it includes the principal mode analysis (PMA), where an eigenvalue problem, derived from cross-correlation functions of particle positions and velocities, is solved.[490,492,493] Subsequently, the generalized normal coordinate scheme of Mathias et al. has been developed.[494,495] It is similar to the PMA approach, but does not require the equipartition theorem to be fulfilled. This can be computed using TRAVIS[418,467] and a short summary of the approach will be given below.

The concept of normal modes generally relies on the assumption that the molecule undertakes small oscillations around a fixed reference structure, which is usually a minimum on the potential energy surface. However, in a MD trajectory, there will also be translational and rotational motion, which needs to be removed prior to the normal coordinate analysis. For this purpose, the trajectories are transformed into the Eckart frame of ref 496 by finding a rotation matrix and a translation vector in each time step such that the mass-weighted root-mean-square distance to a reference structure is minimized. The assumption of small oscillations around one single reference structure breaks down if there are conformational changes within the trajectory. In this context, the procedure needs to be extended to several reference structures.[495] Either these can differ only in the ordering of equivalent atoms (e.g., if a methyl group rotates, where the hydrogen atoms are indistinguishable), or they belong to structurally different minima on the potential energy surface (e.g., if a butyl group changes between trans and gauche conformation). In the former case, all conformations can be mapped to a single minimum by considering the permutations of equivalent atoms, whereas in the latter case, the normal coordinate analysis results in independent normal modes for all minima. If there exists more than one reference structure, the probability of a molecule corresponding to a reference structure has to be defined for all snapshots, all molecules, and all reference structures.

For each reference structure, a "matrix" of all atomic velocity cross-correlation functions (i.e., each matrix element is a function) is computed. This matrix shall be diagonalized. As the matrix elements are not numbers, standard diagonalization techniques cannot be applied. Still, it is possible to minimize the integral over the off-diagonal functions by a modified Jacobi algorithm,[446,494] leading to an orthogonal transformation matrix that minimizes the cross-correlation spectra. This transformation matrix consists of the set of new coordinate vectors that characterize the normal modes. Since the trace of a matrix is not changed by an orthogonal transformation, the total power spectrum of the system is not modified by this procedure, but the modes are localized in frequency space. More details can be found in ref 497.

11. TECHNICAL ASPECTS

CP2K is built via CMake and requires a modern Fortran and C compiler; several are known to work, among others, the GNU, Intel, and Cray compilers. A list of supported compiler versions is maintained on the CP2K wiki.[a] Furthermore, many configurations are continuously tested on the CP2K dashboard.[498]

11.1. Installation

11.1.1. Prerequisites and Dependencies. CP2K uses several external libraries, listed in Table 2. The table also

Table 2. List of Internal (Int.), Required (Req.), Optional (Opt.), and Required for Parallel Version (Par.) CP2K Dependencies[a]

library	req ?	GPU	purposes
BLAS/LAPACK	req		general
DBCSR[90]	req	C,H,O	sparse matrix
DBM	int.	C, H	sparse matrix
grid	int.	C, H	integration
MPI	par.		general
ScaLAPACK	par.		general
FFTW3[499]	opt.		FFT
FPGA[500−503]	opt.	O	PW-FFT
COSMA[103]	opt.	C, H	matrix multipl.
SPLA[91]	opt.	C, H	matrix multipl.
LIBXSMM[504]	opt.		DBCSR, DBM
LIBINT[105]	opt.		HF exchange
LIBXC[94]	opt.		XC functionals
ELPA[505]	opt.	C	diagonalization
cuSOLVERMp[506]	opt.	C	diagonalization
DLA-Future[507]	opt.	C, H	diagonalization
SIRIUS[508]	opt.	C, H	separate PW code
libvori[464]	opt.		Voronoi integration
DFT-D4[509]	opt.		dispersion corr.
tblite[510]	opt.		GFN2-xTB
PW	int.	C, H	solvation models
libgrpp[511]	int.		ECPs
PLUMED[512]	opt.		sampling methods
spglib[513]	opt.		symmetry detection
LibTorch	opt.		ML library
DeePMD-kit[319]	opt.		ML potentials
ACE[318]	opt.		ML potentials
NequIP[320]	opt.		ML potentials
Allegro[321]	opt.		ML potentials
Smeagol[514]	opt.		NEGF transport
TREXIO[515]	opt.		IO formats
GreenX[516]	opt.		Green's functions

[a]In order to use GPUs, either CUDA (C), HIP (H), or OpenCL (O) can be employed.

indicates which libraries can make use of GPU acceleration via CUDA, HIP, or OpenCL. In this case, the corresponding GPU computing libraries have to be present, and the provided MPI distribution needs to support them. Table 2 also lists the purposes of external libraries. Most of the libraries are optional and are only required for performance improvements, or to enable specific functionalities. The second column in Table 2 categorizes the dependencies as internal, required, and optional. For MPI parallelization, an MPI implementation and ScaLAPACK are additionally required.

11.1.2. Building CP2K. There are four suggested ways for the installation of CP2K: (1) via distribution packages, (2) via containers, (3) from the source with CMake, and (4) with the Spack package manager.

The quickest installation methods are via a Linux or macOS distribution,[b] or through ready-made containers[c] run with Apptainer, Podman, or Docker. These methods also include the ability to run CP2K with MPI parallelization and GPU acceleration. Both installation methods require no build steps or tuning, and containers also offer optimized versions for essential platforms.[d,e] Generally, a distribution package or container, if available, is the easiest way to install CP2K. However, beware that many packages include only a small subset of the optional dependencies.

Installation via CMake is recommended for users with a strong technical background.[f] Using CMake, it is possible to fully customize the CP2K installation, enable any of the optional dependencies, and fine-tune the various compilation options.

The Spack package manager[517] is recommended for advanced users, or HPC clusters to obtain a feature-complete installation of CP2K.[g] Spack is especially suited for the installation and management of multiple versions of CP2K together with different versions of its dependencies. One can also use Spack to only install CP2K's dependencies and then build CP2K itself manually via CMake.

Many HPC centers also provide CP2K as a software module. Such preinstalled binaries are generally preferable because they typically have been carefully tuned to ensure good performance on the specific hardware.

11.1.3. Sanity Checks after Installation. For CMake and Spack installations, as well as MPI-parallel and GPU-enabled containers, it is highly recommended to perform some sanity checks before starting extensive calculations with CP2K.

It is recommended to check that (1) CP2K was installed correctly and produces correct results, (2) that MPI and OpenMP parallelization work, and (3) for GPU installations, that the GPU is actually used. The Python script do_regtest.py provided as part of the CP2K source code can be used to test all three aspects. The script can be invoked as follows: ./do_regtest.py /cp2k/path/bin/ psmp

In order to run the tests, it may be required to define the data directory (located in the CP2K directory) via export CP2K_DATA_DIR = /Path/to/data. For GPU builds, its usage should be checked with the help of the nvidia-smi (or similar) command-line tool during the execution of the tests.

The CP2K test suite consists of more than 4000 regtests. However, although these are all complete and syntactically correct CP2K input files, they should not be misunderstood as a source for meaningful settings, as the parameters are often chosen to minimize execution time of a specific functionality instead of reliability of the resulting data. For that purpose, the examples of the present work at https://github.com/cp2k/cp2k-examples are recommended instead. A detailed description of CP2K's regression testing is available at https://www.cp2k.org/dev:regtesting. Depending on the build options, the relevant tests will be executed. For each test, an input file is run, and then parts of the output are compared against reference values. When a test crashes midway with a runtime failure, this clearly indicates a serious problem with the binary. It is more ambiguous when a test is complete, but the result

does not agree with the reference value within the given threshold. In such cases, additional investigation is required to discern whether the threshold was simply a bit too tight, the test did not converge due to numerical instabilities, or there is an actual problem with the binary.

Persistent errors can be reported to CP2K developers using the GitHub issue tracker,[h] the dedicated Google group,[i] or GitHub discussions.[j] These resources also contain other known problems and possible solutions.

Beside directly contributing your code, as the constraint DFT implementation of Holmberg and Laasonen for instance,[518,519] or interfacing to CP2K (e.g., PLUMED,[520] phonopy,[521] Libra,[522] PyRETIS,[523−525] Wannier90,[526] Green-X,[516] Newton-X,[527] DeepMD-kit,[528] MiMiC[529]), it is possible to connect to CP2K via community interfaces (e.g., ASE,[530] i-Pi,[531,532] or Spicy[533]) or domain-specific file formats (e.g., MD-TRACKS,[534] ZEOBUILDER,[535] TAMkin,[536] IOData,[537] openPMD,[538,539] TREXIO,[515] Multiwfn[540]), as well as compiling CP2K as a library (e.g., ISA,[541,542] PIMD,[543,544] phonopy,[545] OpenMolcas,[546] or Qiskit Nature,[547] MixPI[548]).

11.2. Performance Aspects

In addition to checking for functionality, the tests can also be used to get a first idea of performance. At the end of each CP2K run, the essential timings will be listed. For the provided test cases, they can be compared with the CP2K dashboard.[498] Additionally, a folder with benchmarks is available in the distribution in the folder `cp2k/benchmarks`. The timings for different systems are listed at,[549] which is periodically updated for new relevant hardware platforms. Both lists should be consulted to get realistic reference values for the expected execution times on the given hardware. A large discrepancy from the reference values suggests suboptimal builds, or a failure to actually use GPUs even though they are present.

A `TIMING` section like the following can be found at the end of each CP2K output file:

```
 -------------------------------------------------------------------------------
 -                                                                             -
 -                                T I M I N G                                  -
 -                                                                             -
 -------------------------------------------------------------------------------
 SUBROUTINE                       CALLS  ASD         SELF TIME        TOTAL TIME
                                MAXIMUM       AVERAGE MAXIMUM AVERAGE MAXIMUM
 CP2K                                 1  1.0   0.029   0.040 342.578 342.580
 qs_mol_dyn_low                       1  2.0   0.017   0.019 342.056 342.057
 qs_forces                           11  3.9   0.010   0.011 341.993 342.006
 qs_energies                         11  4.9   0.003   0.004 316.732 316.771
 scf_env_do_scf                      11  5.9   0.001   0.002 291.649 291.652
 [...]
 grid_collocate_task_list           107  9.7  80.861 126.411  80.861 126.411
 grid_integrate_task_list           107 12.3  81.736 125.015  81.736 125.015
 density_rs2pw                       107  9.7   0.015   0.018  53.039 117.863
 transfer_rs2pw                     439 10.6   0.015   0.018  49.033 115.612
 mp_waitany                        4720 13.7  47.543 113.939  47.543 113.939
 [...]
```

It lists different subroutines and in the second column the number of calls is given. The column `ASD` indicates the nesting level of the code. The timings are split into two groups: `SELF TIME` is the time the routine needs without nested subroutine calls inside, whereas `TOTAL TIME` includes them. For each group there are two columns: `MAXIMUM` gives the time of the MPI rank with the longest execution time, the `AVERAGE` value is the average over all ranks. A large difference between these two indicates load balancing issues. Routines and library calls that are computationally expensive can be identified from the `SELF TIME` column. In the example a significant amount is spent in the integration of the DFT grid (i.e., grid_* routines). The first part of the routine name relates the routine to different parts of the code. For example, "mp_" indicates that it is part of the MPI part, related to communication.

Depending on the type of computation, additional performance statistics appear earlier in the output file. In most calculations, the DBCSR library will be used, which reports its own statistics like:

```
 -------------------------------------------------------------------------------
 -                                                                             -
 -                             DBCSR STATISTICS                                -
 -                                                                             -
 -------------------------------------------------------------------------------
 COUNTER                               TOTAL       BLAS     SMM      ACC
 flops     9 x    32 x    32       207618048       0.0%   100.0%    0.0%
 [...]
 flops inhomo. stacks           159221022720     100.0%     0.0%    0.0%
 flops total                      1.710628E+12      9.3%    90.7%    0.0%
 flops max/rank                 114.624955E+09     12.4%    87.6%    0.0%
 matmuls inhomo. stacks                54240      100.0%     0.0%    0.0%
 matmuls total                      82804340        0.1%    99.9%    0.0%
 number of processed stacks           894728        6.1%    93.9%    0.0%
 average stack size                        1.0              98.5     0.0
 marketing flops                   1.904784E+12
 -------------------------------------------------------------------------------
```

Besides the number of flops (floating point operations per second) listed in the column `TOTAL` the percentage of operations executed with different modules is listed. DBCSR can use different libraries for basic linear algebra operations, which is indicated in the last three columns. This can be different `BLAS` implementations like cuBLAS, the LIBXSMM[504]/libsmm libraries (`SMM`) or execution on GPUs (`ACC`).

The general recommendation for faster execution is to switch to GPUs when available and whenever most of the runtime is reported for libraries that have a GPU option. If there are considerable runtime shares that cannot be replaced by GPUs, then using them will still entail a gain in performance, yet in a limited and inefficient way.

If this is not advised, then the only general alternative is to scale up the number of MPI ranks for multiple compute nodes and OpenMP threads for CPU cores inside of the compute nodes. Often it is beneficial to balance this toward more MPI ranks and fewer OpenMP threads (for the same product of MPI ranks × OpenMP threads). There is no general optimum, and performance experiments of your own for the type of computations performed may be required to reach the sweet spot.

For a series of many independent calculations like in parameter studies or high-throughput computation projects,[55] there is a general recommendation, however: reduce the parallelism of each individual run and execute more separate cases concurrently (trivial parallelization). Make sure that each run has its exclusive computing resources. This will bring the best parallel efficiency, at least up to the point where memory usage or other constraints will prevent it. In most cases, a sensible compromise between the extremes will be best for practical purposes.

■ ASSOCIATED CONTENT

Data Availability Statement

The data underlying this study are openly available at https://github.com/cp2k/cp2k-examples

■ AUTHOR INFORMATION

Corresponding Author

Thomas D. Kühne − *Center for Advanced Systems Understanding (CASUS), Helmholtz Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany; Institute of Artificial Intelligence, Technische Universität Dresden, D-01187 Dresden, Germany; Cluster of Excellence "Physics of Life", Technische Universität Dresden, D-01307 Dresden,*

Germany; ⓘ orcid.org/0000-0001-5471-2407;
Email: tkuehne@cp2k.org

Authors

Marcella Iannuzzi − *Department of Chemistry, University of Zurich, CH-8057 Zürich, Switzerland;* ⓘ orcid.org/0000-0001-9717-2527

Jan Wilhelm − *Regensburg Center for Ultrafast Nanoscopy (RUN) and Institute of Theoretical Physics, University of Regensburg, D-93053 Regensburg, Germany;* ⓘ orcid.org/0000-0001-8678-8246

Frederick Stein − *Center for Advanced Systems Understanding (CASUS), Helmholtz Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany;* ⓘ orcid.org/0000-0003-3689-4926

Augustin Bussy − *Swiss National Supercomputing Centre (CSCS), ETH Zurich, CH-6900 Lugano, Switzerland*

Hossam Elgabarty − *Department of Chemistry, Paderborn University, D-33098 Paderborn, Germany*

Dorothea Golze − *Department of Chemistry and Food Chemistry, Technische Universität Dresden, D-01069 Dresden, Germany;* ⓘ orcid.org/0000-0002-2196-9350

Anna-Sophia Hehn − *Department of Chemistry, Christian-Albrechts-University Kiel, D-24118 Kiel, Germany;* ⓘ orcid.org/0000-0002-0498-740X

Maximilian Graml − *Regensburg Center for Ultrafast Nanoscopy (RUN) and Institute of Theoretical Physics, University of Regensburg, D-93053 Regensburg, Germany;* ⓘ orcid.org/0000-0002-4279-8511

Stepan Marek − *Regensburg Center for Ultrafast Nanoscopy (RUN) and Institute of Theoretical Physics, University of Regensburg, D-93053 Regensburg, Germany;* ⓘ orcid.org/0000-0002-8806-1670

Beliz Sertcan Gökmen − *Department of Chemistry, University of Zurich, CH-8057 Zürich, Switzerland*

Christoph Schran − *Cavendish Laboratory, Department of Physics, University of Cambridge, CB3 0HE Cambridge, U.K.;* ⓘ orcid.org/0000-0003-4595-5073

Harald Forbert − *Center for Solvation Science ZEMOS, Ruhr-Universität Bochum, D-44801 Bochum, Germany*

Rustam Z. Khaliullin − *Department of Chemistry, McGill University, Montreal QC H3A 0B8, Canada;* ⓘ orcid.org/0000-0002-9073-6753

Anton Kozhevnikov − *Swiss National Supercomputing Centre (CSCS), ETH Zurich, CH-6900 Lugano, Switzerland*

Mathieu Taillefumier − *Swiss National Supercomputing Centre (CSCS), ETH Zurich, CH-6900 Lugano, Switzerland*

Rocco Meli − *Swiss National Supercomputing Centre (CSCS), ETH Zurich, CH-6900 Lugano, Switzerland;* ⓘ orcid.org/0000-0002-2845-3410

Vladimir V. Rybkin − *HQS Quantum Simulations GmbH, D-76131 Karlsruhe, Germany;* ⓘ orcid.org/0000-0001-5136-6035

Martin Brehm − *Department of Chemistry, Paderborn University, D-33098 Paderborn, Germany*

Robert Schade − *Paderborn Center for Parallel Computing (PC2), Paderborn University, D-33098 Paderborn, Germany*

Ole Schütt − *CP2K Foundation, CH-8057 Zürich, Switzerland;* ⓘ orcid.org/0009-0005-8661-3418

Johann V. Pototschnig − *Center for Advanced Systems Understanding (CASUS), Helmholtz Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany;* ⓘ orcid.org/0000-0002-9982-0556

Hossein Mirhosseini − *Center for Advanced Systems Understanding (CASUS), Helmholtz Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany*

Andreas Knüpfer − *Center for Advanced Systems Understanding (CASUS), Helmholtz Zentrum Dresden-Rossendorf, D-02826 Görlitz, Germany*

Dominik Marx − *Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany;* ⓘ orcid.org/0009-0005-0865-1419

Matthias Krack − *PSI Center for Scientific Computing, Theory and Data, Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland;* ⓘ orcid.org/0000-0002-2082-7027

Jürg Hutter − *Department of Chemistry, University of Zurich, CH-8057 Zürich, Switzerland*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.5c05851

Notes

The authors declare no competing financial interest.

■ ADDITIONAL NOTES

[a]https://www.cp2k.org/dev:compiler_support.
[b]https://manual.cp2k.org/trunk/getting-started/distributions.html.
[c]https://github.com/cp2k/cp2k-containers.
[d]Docker hub https://hub.docker.com/r/cp2k/cp2k.
[e]NVIDIA NGC catalog https://catalog.ngc.nvidia.com/orgs/hpc/containers/cp2k.
[f]https://manual.cp2k.org/trunk/getting-started/build-from-source.html.
[g]https://manual.cp2k.org/trunk/getting-started/build-with-spack.html.

[h]CP2K issue tracker: https://github.com/cp2k/cp2k/issues.
[i]CP2K Google group: https://groups.google.com/g/cp2k.
[j]CP2K GitHub discussions: https://github.com/cp2k/cp2k/discussions.

■ REFERENCES

(1) Dirac, P. A. M. Quantum mechanics of many-electron systems. *Math. Proc. Cambridge Philos. Soc.* **1929**, *123*, 714−733.

(2) Billeter, S. R.; Curioni, A.; Andreoni, W. Efficient linear scaling geometry optimization and transition-state search for direct wave-function optimization schemes in density functional theory using a plane-wave basis. *Comput. Mater. Sci.* **2003**, *27*, 437−445.

(3) Schlegel, H. B. Geometry Optimization. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 790−809.

(4) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(5) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(6) Rahman, A. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.* **1964**, *136*, A405−A411.

(7) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press, 2010.

(8) Frenkel, D.; Smit, B. *Understanding Molecular Simulation From Algorithms to Applications*; Academic Press, 2023.

(9) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press, 2009.

(10) Hutter, J. Car-Parrinello molecular dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 604−612.

(11) Kühne, T. D.; Krack, M.; Mohamed, F. R.; Parrinello, M. Efficient and accurate Car-Parrinello-like approach to Born-Oppenheimer molecular dynamics. *Phys. Rev. Lett.* **2007**, *98*, 066401.

(12) Kühne, T. D.; Prodan, E. Disordered crystals from first principles I: Quantifying the configuration space. *Ann. Phys.* **2018**, *391*, 120−149.

(13) Kühne, T. D. Second generation Car-Parrinello molecular dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 391−406.

(14) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An electronic structure and molecular dynamics software package-Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(15) Hutter, J.; Iannuzzi, M.; Kühne, T. D.. In *Comprehensive Computational Chemistry*; Yáñez, M., Boyed, R. J., Eds.; Elsevier, 2024; pp 493−517.

(16) Kohn, W. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **1999**, *71*, 1253−1266.

(17) Pople, J. A. Nobel Lecture: Quantum chemical models. *Rev. Mod. Phys.* **1999**, *71*, 1267−1274.

(18) Ihm, J.; Zunger, A.; Cohen, M. L. Momentum-space formalism for the total energy of solids. *J. Phys. C: Solid State Phys.* **1979**, *12*, 4409−4422.

(19) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953−17979.

(20) Andersen, O. K. Linear methods in band theory. *Phys. Rev. B* **1975**, *12*, 3060−3083.

(21) Aryasetiawan, F.; Gunnarsson, O. The GW method. *Rep. Prog. Phys.* **1998**, *61*, 237.

(22) Krack, M.; Parrinello, M. *High Performance Computing in Chemistry*; NIC Series; NIC-Directors, 2004; Vol. *25*, pp 29−51.

(23) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

(24) Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477−488.

(25) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **1999**, *103*, 124−140.

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

(27) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618.

(28) Bohm, D.; Pines, D. A. collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas. *Phys. Rev.* **1953**, *92*, 609.

(29) Gell-Mann, M.; Brueckner, K. A. Correlation energy of an electron gas at high density. *Phys. Rev.* **1957**, *106*, 364.

(30) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

(31) Goedecker, S.; Teter, M.; Hutter, J. Separable dual−space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703−1710.

(32) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641−3662.

(33) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145−152.

(34) Reiher, M. Relativistic douglas−kroll−hess theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 139−149.

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

(36) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

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

(38) Lu, J.-B.; Cantu, D. C.; Nguyen, M.-T.; Li, J.; Glezakou, V.-A.; Rousseau, R. Norm-conserving pseudopotentials and basis sets to explore lanthanide chemistry in complex environments. *J. Chem. Theory Comput.* **2019**, *15*, 5987−5997.

(39) Lu, J.-B.; Cantu, D. C.; Xu, C.-Q.; Nguyen, M.-T.; Hu, H.-S.; Glezakou, V.-A.; Rousseau, R.; Li, J. Norm-conserving pseudopotentials and basis sets to explore actinide chemistry in complex environments. *J. Chem. Theory Comput.* **2021**, *17*, 3360−3371.

(40) Lu, J.-B.; Jiang, X.-L.; Hu, H.-S.; Li, J. Norm-conserving 4f-in-core pseudopotentials and basis sets optimized for trivalent lanthanides (Ln= Ce−Lu). *J. Chem. Theory Comput.* **2023**, *19*, 82−96.

(41) Lu, J.-B.; Zhang, Y.-Y.; Liu, J.-B.; Li, J. Norm-Conserving 5f-in-Core Pseudopotentials and Gaussian Basis Sets Optimized for Tri-and Tetra-Valent Actinides (An= Pa−Lr). *J. Chem. Theory Comput.* **2025**, *21*, 170−182.

(42) Willand, A.; Kvashnin, Y. O.; Genovese, L.; Vázquez-Mayagoitia, Á.; Deb, A. K.; Sadeghi, A.; Deutsch, T.; Goedecker, S. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *J. Chem. Phys.* **2013**, *138*, 104109.

(43) Oleynichenko, A. V.; Zaitsevskii, A.; Mosyagin, N. S.; Petrov, A. N.; Eliav, E.; Titov, A. V. LIBGRPP: A library for the evaluation of molecular integrals of the generalized relativistic pseudopotential operator over Gaussian functions. *Symmetry* **2023**, *15*, 197.

(44) 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.* **2019**, *59*, 4814−4820.

(45) Zijlstra, E. S.; Huntemann, N.; Kalitsov, A.; Garcia, M. E.; Von Barth, U. Optimized Gaussian basis sets for Goedecker−Teter−Hutter pseudopotentials. *Model. Simul. Mater. Sci. Eng.* **2009**, *17*, 015009.

(46) Peintinger, M. F.; Oliveira, D. V.; Bredow, T. Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *J. Comput. Chem.* **2013**, *34*, 451−459.

(47) Vilela Oliveira, D.; Laun, J.; Peintinger, M. F.; Bredow, T. BSSE-correction scheme for consistent gaussian basis sets of double- and triple-zeta valence with polarization quality for solid-state calculations. *J. Comput. Chem.* **2019**, *40*, 2364−2376.

(48) Laun, J.; Bredow, T. BSSE-corrected consistent Gaussian basis sets of triple-zeta valence with polarization quality of the sixth period for solid-state calculations. *J. Comput. Chem.* **2021**, *42*, 1064−1072.

(49) Laun, J.; Bredow, T. BSSE-corrected consistent Gaussian basis sets of triple-zeta valence with polarization quality of the fifth period for solid-state calculations. *J. Comput. Chem.* **2022**, *43*, 839−846.

(50) Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2348−2364.

(51) Hedin, L. New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Phys. Rev.* **1965**, *139*, A796−A823.

(52) Aryasetiawan, F.; Gunnarsson, O. The GW method. *Rep. Prog. Phys.* **1998**, *61*, 237−312.

(53) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496−4501.

(54) Dunlap, B. I.; Connolly, J.; Sabin, J. On some approximations in applications of $X\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396−3402.

(55) Bosoni, E.; Beal, L.; Bercx, M.; Blaha, P.; Blügel, S.; Bröder, J.; Callsen, M.; Cottenier, S.; Degomme, A.; Dikan, V.; et al. How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. *Nat. Rev. Phys.* **2024**, *6*, 45−58.

(56) Krack, M.; Parrinello, M. All-electron ab-initio molecular dynamics. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2105−2112.

(57) Dorado, B.; Amadon, B.; Freyss, M.; Bertolus, M. DFT + U calculations of the ground state and metastable states of uranium dioxide. *Phys. Rev. B* **2009**, *79*, 235125.

(58) Krack, M. On the ground state electronic structure of uranium dioxide. *Phys. Scr.* **2015**, *90*, 094014.

(59) Krack, M.; Gambirasio, A.; Parrinello, M. *Ab initio* X-ray scattering of liquid water. *J. Chem. Phys.* **2002**, *117*, 9409−9412.

(60) Müller, P.; Karhan, K.; Krack, M.; Gerstmann, U.; Schmidt, W. G.; Bauer, M.; Kühne, T. D. Impact of finite-temperature and condensed-phase effects on theoretical X-ray absorption spectra of transition metal complexes. *J. Comput. Chem.* **2019**, *40*, 712−716.

(61) Krack, M.; Köster, A. M. An adaptive numerical integrator for molecular integrals. *J. Chem. Phys.* **1998**, *108*, 3226−3234.

(62) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365−4369.

(63) Kerker, G. Efficient iteration scheme for self-consistent pseudopotential calculations. *Phys. Rev. B* **1981**, *23*, 3082.

(64) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

(65) Marks, L. D.; Luke, D. Robust mixing for ab initio quantum mechanical calculations. *Phys. Rev. B* **2008**, *78*, 075114.

(66) Broyden, C. G. A Class of Methods for Solving Nonlinear Simultaneous Equations. *Math. Comput.* **1965**, *19*, 577.

(67) Mermin, N. D. Thermal Properties of the Inhomogeneous Electron Gas. *Phys. Rev.* **1965**, *137*, A1441−A1443.

(68) Hutter, J.; Parrinello, M.; Vogel, S. Exponential transformation of molecular orbitals. *J. Chem. Phys.* **1994**, *101*, 3862−3865.

(69) Gan, C.; Haynes, P.; Payne, M. Preconditioned conjugate gradient method for the sparse generalized eigenvalue problem in electronic structure calculations. *Comput. Phys. Commun.* **2001**, *134*, 33−40.

(70) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin−zone integrations. *Phys. Rev. B* **1976**, *13*, 5188−5192.

(71) Guidon, M.; Hutter, J.; VandeVondele, J. Robust Periodic Hartree-Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *J. Chem. Theory Comput.* **2009**, *5*, 3010−3021.

(72) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(73) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(74) Merlot, P.; Izsák, R.; Borgoo, A.; Kjærgaard, T.; Helgaker, T.; Reine, S. Charge-constrained auxiliary-density-matrix methods for the Hartree−Fock exchange contribution. *J. Chem. Phys.* **2014**, *141*, 094104.

(75) Jensen, F. Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074−1085.

(76) Kumar, C.; Fliegl, H.; Jensen, F.; Teale, A. M.; Reine, S.; Kjærgaard, T. Accelerating Kohn-Sham response theory using density fitting and the auxiliary-density-matrix method. *Int. J. Quantum Chem.* **2018**, *118*, No. e25639.

(77) Gavini, V.; Baroni, S.; Blum, V.; Bowler, D. R.; Buccheri, A.; Chelikowsky, J. R.; Das, S.; Dawson, W.; Delugas, P.; Dogan, M.; et al. Roadmap on electronic structure codes in the exascale era. *Model. Simul. Mater. Sci. Eng.* **2023**, *31*, 063301.

(78) Bussy, A.; Schütt, O.; Hutter, J. Sparse tensor based nuclear gradients for periodic Hartree−Fock and low-scaling correlated wave function methods in the CP2K software package: A massively parallel and GPU accelerated implementation. *J. Chem. Phys.* **2023**, *158*, 164109.

(79) Vahtras, O.; Almlöf, J.; Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.

(80) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285−4291.

(81) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic generation of auxiliary basis sets. *J. Chem. Theory Comput.* **2017**, *13*, 554−562.

(82) Bussy, A.; Hutter, J. Efficient periodic resolution-of-the-identity Hartree−Fock exchange method with k-point sampling and Gaussian basis sets. *J. Chem. Phys.* **2024**, *160*, 064116.

(83) Del Ben, M.; Hutter, J.; VandeVondele, J. Second-Order Møller−Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach. *J. Chem. Theory Comput.* **2012**, *8*, 4177−4188.

(84) Del Ben, M.; Hutter, J.; VandeVondele, J. Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme. *J. Chem. Theory Comput.* **2013**, *9*, 2654−2671.

(85) Weigend, F.; Häser, M. RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **1997**, *97*, 331−340.

(86) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143−152.

(87) Del Ben, M.; Hutter, J.; VandeVondele, J. Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach. *J. Chem. Phys.* **2015**, *143*, 102803.

(88) Rybkin, V. V.; VandeVondele, J. Spin-Unrestricted Second-Order Møller−Plesset (MP2) Forces for the Condensed Phase: From Molecular Radicals to F-Centers in Solids. *J. Chem. Theory Comput.* **2016**, *12*, 2214−2223.

(89) Schwartz, C. Importance of Angular Correlations between Atomic Electrons. *Phys. Rev.* **1962**, *126*, 1015−1019.

(90) Borštnik, U.; VandeVondele, J.; Weber, V.; Hutter, J. Sparse matrix multiplication: The distributed block-compressed sparse row library. *Parallel Comput.* **2014**, *40*, 47−58.

(91) SPLA: Specialized Parallel Linear Algebra. https://github.com/eth-cscs/spla, (accessed, 2025 05 29).

(92) Stein, F.; Hutter, J. Double-hybrid density functionals for the condensed phase: Gradients, stress tensor, and auxiliary-density-matrix method acceleration. *J. Chem. Phys.* **2022**, *156*, 074107.

(93) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.

(94) (a) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in libxc — A comprehensive library of

functionals for density functional theory. *SoftwareX* **2018**, *7*, 1−5. See. https://gitlab.com/libxc/libxc.

(95) Stein, F.; Hutter, J.; Rybkin, V. V. Double-hybrid DFT functionals for the condensed phase: Gaussian and plane waves implementation and evaluation. *Molecules* **2020**, *25*, 5174.

(96) Furche, F. Developing the random phase approximation into a practical post-Kohn−Sham correlation model. *J. Chem. Phys.* **2008**, *129*, 114105.

(97) Boyd, J. P. Exponentially convergent Fourier-Chebshev quadrature schemes on bounded and infinite intervals. *J. Sci. Comput.* **1987**, *2*, 99−109.

(98) Braess, D.; Hackbusch, W. Approximation of 1/x by exponential sums in [1, ∞). *IMA J. Numer. Anal.* **2005**, *25*, 685−697.

(99) Ren, X.; Rinke, P.; Scuseria, G. E.; Scheffler, M. Renormalized second-order perturbation theory for the electron correlation energy: Concept, implementation, and benchmarks. *Phys. Rev. B* **2013**, *88*, 035120.

(100) Bates, J. E.; Furche, F. Communication: Random phase approximation renormalized many-body perturbation theory. *J. Chem. Phys.* **2013**, *139*, 171103.

(101) Freeman, D. L. Coupled-cluster expansion applied to the electron gas: Inclusion of ring and exchange effects. *Phys. Rev. B* **1977**, *15*, 5512−5521.

(102) Stein, F.; Hutter, J. Massively parallel implementation of gradients within the random phase approximation: Application to the polymorphs of benzene. *J. Chem. Phys.* **2024**, *160*, 024120.

(103) Kwasniewski, G.; Kabić, M.; Besta, M.; VandeVondele, J.; Solcà, R.; Hoefler, T. Red-blue pebbling revisited: Near optimal parallel matrix-matrix multiplication. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; Association for Computing Machinery, 2019; pp 1−22.

(104) Wilhelm, J.; Seewald, P.; Del Ben, M.; Hutter, J. Large-scale cubic-scaling random phase approximation correlation energy calculations using a Gaussian basis. *J. Chem. Theory Comput.* **2016**, *12*, 5851−5859.

(105) Valeev, E. F. Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions. http://libint.valeyev.net/, 2024; (accessed, 2025 05 29).

(106) Anisimov, V. I.; Zaanen, J.; Andersen, O. K. Band theory and Mott insulators: Hubbard U instead of Stoner I. *Phys. Rev. B* **1991**, *44*, 943−954.

(107) Dudarev, S. L.; Manh, D. N.; Sutton, A. P. Effect of Mott-Hubbard correlations on the electronic structure and structural stability of uranium dioxide. *Philos. Mag. B* **1997**, *75*, 613−628.

(108) Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B* **1998**, *57*, 1505−1509.

(109) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(110) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365−375.

(111) Blaha, P.; Schwarz, K.; Tran, F.; Laskowski, R.; Madsen, G. K. H.; Marks, L. D. WIEN2k: An APW+lo program for calculating the properties of solids. *J. Chem. Phys.* **2020**, *152*, 074101.

(112) The FLEUR project. https://www.flapw.de/, (accessed: 2025 10 17).

(113) The Elk Code. https://elk.sourceforge.io/, (accessed, 2025 10 17).

(114) Gulans, A.; Kontur, S.; Meisenbichler, C.; Nabok, D.; Pavone, P.; Rigamonti, S.; Sagmeister, S.; Werner, U.; Draxl, C. exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory. *J. Phys.: Condens. Matter* **2014**, *26*, 363202.

(115) Hamann, D. R.; Schlüter, M.; Chiang, C. Norm-Conserving Pseudopotentials. *Phys. Rev. Lett.* **1979**, *43*, 1494−1497.

(116) Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **1990**, *41*, 7892−7895.

(117) Slater, J. C. Wave functions in a periodic potential. *Phys. Rev.* **1937**, *51*, 846.

(118) Ceperley, D. M.; Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **1980**, *45*, 566.

(119) Golze, D.; Dvorak, M.; Rinke, P. The GW compendium: A practical guide to theoretical photoemission spectroscopy. *Front. Chem.* **2019**, *7*, 377.

(120) Capelle, K. A. bird's-eye view of density-functional theory. *Braz. J. Phys.* **2006**, *36*, 1318−1343.

(121) Caravati, S.; Bernasconi, M.; Kühne, T.; Krack, M.; Parrinello, M. First-principles study of crystalline and amorphousGe2Sb2Te5 and the effects of stoichiometric defects. *J. Phys.: Condens. Matter* **2009**, *21*, 255501.

(122) Caravati, S.; Colleoni, D.; Mazzarello, R.; Kühne, T.; Krack, M.; Bernasconi, M.; Parrinello, M. First-principles study of nitrogen doping in cubic and amorphousGe2Sb2Te5. *J. Phys.: Condens. Matter* **2011**, *23*, 265801.

(123) Los, J. H.; Kühne, T. D.; Gabardi, S.; Bernasconi, M. First-principles study of the amorphous In 3 SbTe 2 phase change compound. *Phys. Rev. B* **2013**, *88*, 174203.

(124) Gabardi, S.; Caravati, S.; Los, J. H.; Kühne, T. D.; Bernasconi, M. Influence of the exchange and correlation functional on the structure of amorphous InSb and In3SbTe2 compounds. *J. Chem. Phys.* **2016**, *144*, 204508.

(125) Schambeck, M.; Golze, D.; Wilhelm, J. Solving multipole challenges in the *GW*100 benchmark enables precise low-scaling *GW* calculations. *Phys. Rev. B* **2024**, *110*, 125146.

(126) Pasquier, R.; Camarasa-Gómez, M.; Hehn, A.-S.; Hernangómez-Pérez, D.; Wilhelm, J. Efficient *GW* band structure calculations using Gaussian basis functions and application to atomically thin transition-metal dichalcogenides. *Phys. Rev. B* **2025**, *112*, 205103.

(127) Graml, M.; Zollner, K.; Hernangómez-Pérez, D.; Faria Junior, P. E.; Wilhelm, J. Low-Scaling GW Algorithm Applied to Twisted Transition-Metal Dichalcogenide Heterobilayers. *J. Chem. Theory Comput.* **2024**, *20*, 2202−2208.

(128) van Setten, M. J.; Caruso, F.; Sharifzadeh, S.; Ren, X.; Scheffler, M.; Liu, F.; Lischner, J.; Lin, L.; Deslippe, J. R.; Louie, S. G.; et al. GW100: Benchmarking G_0W_0 for Molecular Systems. *J. Chem. Theory Comput.* **2015**, *11*, 5665−5687.

(129) Setyawan, W.; Curtarolo, S. High-throughput electronic band structure calculations: Challenges and tools. *Comput. Mater. Sci.* **2010**, *49*, 299−312.

(130) Schusteritsch, G.; Kühne, T. D.; Guo, Z. X.; Kaxiras, E. The effect of Ag, Pb and Bi impurities on grain boundary sliding and intergranular decohesion in Copper. *Philos. Mag.* **2016**, *96*, 2868−2886.

(131) Praprotnik, M.; Site, L. D.; Kremer, K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545−571.

(132) Cortes-Huerto, R.; Praprotnik, M.; Kremer, K.; Delle Site, L. From adaptive resolution to molecular dynamics of open systems. *Eur. Phys. J. B* **2021**, *94*, 189.

(133) Panahian Jand, S.; Kühne, T. D.; Delle Site, L. On the physical consistency of an open quantum region with a classical reservoir in molecular simulation. *Adv. Theory Simul.* **2024**, *7*, 2400833.

(134) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999−3094.

(135) Fattebert, J.-L.; Gygi, F. Density functional theory for efficient ab initio molecular dynamics simulations in solution. *J. Comput. Chem.* **2002**, *23*, 662−666.

(136) Fattebert, J.-L.; Gygi, F. First-principles molecular dynamics simulations in a continuum solvent. *Int. J. Quantum Chem.* **2003**, *93*, 139−147.

(137) Andreussi, O.; Dabo, I.; Marzari, N. Revised self-consistent continuum solvation in electronic-structure calculations. *J. Chem. Phys.* **2012**, *136*, 064102.

(138) Yin, W.-J.; Krack, M.; Li, X.; Chen, L.-Z.; Liu, L.-M. Periodic continuum solvation model integrated with first-principles calculations for solid surfaces. *Prog. Nat. Sci.* **2017**, *27*, 283−288.

(139) Yin, W.-J.; Krack, M.; Wen, B.; Ma, S.-Y.; Liu, L.-M. CO2 Capture and Conversion on Rutile TiO2(110) in the Water Environment: Insight by First-Principles Calculations. *J. Phys. Chem. Lett.* **2015**, *6*, 2538−2545.

(140) Dupont, C.; Andreussi, O.; Marzari, N. Self-consistent continuum solvation (SCCS): The case of charged systems. *J. Chem. Phys.* **2013**, *139*, 214110.

(141) Martyna, G. J.; Tuckerman, M. E. A reciprocal space based method for treating long range interactions inab initioand force-field-based calculations in clusters. *J. Chem. Phys.* **1999**, *110*, 2810−2821.

(142) Cococcioni, M.; Mauri, F.; Ceder, G.; Marzari, N. Electronic-Enthalpy Functional for Finite Systems Under Pressure. *Phys. Rev. Lett.* **2005**, *94*, 145501.

(143) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(144) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.

(145) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(146) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1− 86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.

(147) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super-computers. *SoftwareX* **2015**, *1−2*, 19−25.

(148) QM/MM Documentation - GROMACS 2022-beta1. https://manual.gromacs.org/documentation/2022-beta1/reference-manual/special/qmmm.html, 2022; (accessed, 2024 10 29).

(149) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(150) Zhang, Y.; Lee, T.-S.; Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46−54.

(151) Ihrig, A. C.; Schiffmann, C.; Sebastiani, D. Specific quantum mechanical/molecular mechanical capping-potentials for biomolecular functional groups. *J. Chem. Phys.* **2011**, *135*, 214107.

(152) Schiffmann, C.; Sebastiani, D. Artificial bee colony optimization of capping potentials for hybrid quantum mechanical/ molecular mechanical calculations. *J. Chem. Theory Comput.* **2011**, *7*, 1307−1315.

(153) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient real space multigrid QM/MM electrostatic coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176−1184.

(154) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Linear-Scaling Electrostatic Coupling for treating periodic boundary conditions in QM/MM Simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370−1378.

(155) Golze, D.; Iannuzzi, M.; Nguyen, M.-T.; Passerone, D.; Hutter, J. Simulation of Adsorption Processes at Metallic Interfaces: An Image Charge Augmented QM/MM Approach. *J. Chem. Theory Comput.* **2013**, *9*, 5086−5097.

(156) Devynck, F.; Iannuzzi, M.; Krack, M. Frenkel pair recombinations in UO2: Importance of explicit description of polarizability in core-shell molecular dynamics simulations. *Phys. Rev. B* **2012**, *85*, 184103.

(157) Blöchl, P. E. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *J. Chem. Phys.* **1995**, *103*, 7422−7428.

(158) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370−1378.

(159) Image Charges. https://manual.cp2k.org/trunk/methods/qm_mm/image_charges.html, 2015; (accessed, 2025 02 05).

(160) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(161) Golze, D.; Hutter, J.; Iannuzzi, M. Wetting of water on hexagonal boron nitrideRh(111): a QM/MM model based on atomic charges derived for nano-structured substrates. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14307−14316.

(162) https://www.cp2k.org/howto:resp, 2015; (accessed, 2025 06 13).

(163) Campañá, C.; Mussard, B.; Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **2009**, *5*, 2866−2878.

(164) Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.

(165) Wu, Q.; Yang, W. A direct optimization method for calculating density functionals and exchange-correlation potentials from electron densities. *J. Chem. Phys.* **2003**, *118*, 2498−2509.

(166) Van Leeuwen, R.; Baerends, E. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A* **1994**, *49*, 2421.

(167) Putrino, A.; Sebastiani, D.; Parrinello, M. Generalized variational density functional perturbation theory. *J. Chem. Phys.* **2000**, *113*, 7102−7109.

(168) Sebastiani, D.; Parrinello, M. A New ab-Initio Approach for NMR Chemical Shifts in Periodic Systems. *J. Phys. Chem. A* **2001**, *105*, 1951−1958.

(169) Sebastiani, D. Ab-initio calculations of NMR parameters in condensed phases. *Mod. Phys. Lett. B* **2003**, *17*, 1301−1319.

(170) Gonze, X. Adiabatic density-functional perturbation theory. *Phys. Rev. A* **1995**, *52*, 1096−1114.

(171) Sebastiani, D. Current Densities and Nucleus-Independent Chemical Shift Maps from Reciprocal-Space Density Functional Perturbation Theory Calculations. *ChemPhysChem* **2006**, *7*, 164−175.

(172) Weber, V.; Iannuzzi, M.; Giani, S.; Hutter, J.; Declerck, R.; Waroquier, M. Magnetic linear response properties calculations with the Gaussian and augmented-plane-wave method. *J. Chem. Phys.* **2009**, *131*, 014106.

(173) Putrino, A.; Parrinello, M. Anharmonic Raman spectra in high-pressure ice from ab initio simulations. *Phys. Rev. Lett.* **2002**, *88*, 176401.

(174) Scherrer, A.; Vuilleumier, R.; Sebastiani, D. Vibrational circular dichroism from ab initio molecular dynamics and nuclear velocity perturbation theory in the liquid phase. *J. Chem. Phys.* **2016**, *145*, 084101.

(175) Resta, R. Why are insulators insulating and metals conducting? *J. Phys.: Condens. Matter* **2002**, *14*, R625−R656.

(176) Mauri, F.; Louie, S. Magnetic Susceptibility of Insulators from First Principles. *Phys. Rev. Lett.* **1996**, *76*, 4246−4249.

(177) Keith, T. A.; Bader, R. F. W. Calculation of magnetic response properties using atoms in molecules. *Chem. Phys. Lett.* **1992**, *194*, 1− 8.

(178) Keith, T. A.; Bader, R. F. W. Calculation of magnetic response properties using a continuous set of gauge transformations. *Chem. Phys. Lett.* **1993**, *210*, 223−231.

(179) van Wüllen, C. *Calculation of NMR and EPR Parameters: Theory and Applications*; Wiley, 2004; pp 83−100.

(180) Cowan, B. P. *Nuclear Magnetic Resonance and Relaxation*, online-ausg ed.; Cambridge University Press: Cambridge, 1997.

(181) Mondal, A.; Gaultois, M. W.; Pell, A. J.; Iannuzzi, M.; Grey, C. P.; Hutter, J.; Kaupp, M. Large-scale computation of nuclear magnetic resonance shifts for paramagnetic solids using CP2K. *J. Chem. Theory Comput.* **2018**, *14*, 377−394.

(182) Schreckenbach, G.; Ziegler, T. Calculation of NMR shielding tensors based on density functional theory and a scalar relativistic Pauli-type Hamiltonian. The application to transition metal complexes. *Int. J. Quantum Chem.* **1997**, *61*, 899−918.

(183) Pickard, C.; Mauri, F. First-Principles Theory of the EPR G Tensor in Solids: Defects in Quartz. *Phys. Rev. Lett.* **2002**, *88*, 086403.

(184) Van Yperen-De Deyne, A.; Pauwels, E.; Van Speybroeck, V.; Waroquier, M. Accurate Spin−orbit and Spin−other-orbit Contributions to the G-tensor for Transition Metal Containing Systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10690−10704.

(185) Neese, F. Efficient and Accurate Approximations to the Molecular Spin-orbit Coupling Operator and Their Use in Molecular G-tensor Calculations. *J. Chem. Phys.* **2005**, *122*, 034107.

(186) Malkina, O. L.; Vaara, J.; Schimmelpfennig, B.; Munzarova, M.; Malkin, V. G.; Kaupp, M. Density Functional Calculations of Electronic G-tensors Using Spin-orbit Pseudopotentials and Mean-field All-electron Spin-orbit Operators. *J. Am. Chem. Soc.* **2000**, *122*, 9206−9218.

(187) Pigliapochi, R.; Pell, A. J.; Seymour, I. D.; Grey, C. P.; Ceresoli, D.; Kaupp, M. DFT investigation of the effect of spin-orbit coupling on the NMR shifts in paramagnetic solids. *Phys. Rev. B* **2017**, *95*, 054412.

(188) Declerck, R.; Pauwels, E.; Van Speybroeck, V.; Waroquier, M. First-principles Calculations of Hyperfine Parameters with the Gaussian and Augmented-plane-wave Method: Application to Radicals Embedded in a Crystalline Environment. *Phys. Rev. B* **2006**, *74*, 245103.

(189) Blase, X.; Duchemin, I.; Jacquemin, D. The Bethe−Salpeter equation in chemistry: relations with TD-DFT, applications and challenges. *Chem. Soc. Rev.* **2018**, *47*, 1022−1043.

(190) Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm-Dancoff approximation. *Chem. Phys. Lett.* **1999**, *314*, 291.

(191) Hutter, J. Excited state nuclear forces from the Tamm−Dancoff approximation to time−dependent density functional theory within the plane wave basis set framework. *J. Chem. Phys.* **2003**, *118*, 3928−3934.

(192) Grimme, S. A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules. *J. Chem. Phys.* **2013**, *138*, 244104.

(193) Hehn, A.-S.; Sertcan, B.; Belleflamme, F.; Chulkov, S. K.; Watkins, M. B.; Hutter, J. Excited-State Properties for Extended Systems: Efficient Hybrid Density Functional Methods. *J. Chem. Theory Comput.* **2022**, *18*, 4186−4202.

(194) Sertcan Gokmen, B.; Hutter, J.; Hehn, A.-S. Excited-State Forces with the Gaussian and Augmented Plane Wave Method for the Tamm−Dancoff Approximation of Time-Dependent Density Functional Theory. *J. Chem. Theory Comput.* **2024**, *20*, 8494−8504.

(195) Vogt, J.-R.; Wilhelm, J.; Hehn, A.-S. Perturbative spin−orbit couplings for the simulation of extended framework materials. *J. Chem. Phys.* **2025**, *162*, 082502.

(196) van Wüllen, C. Molecular density functional calculations in the regular relativistic approximation: Method, application to coinage metal diatomics, hydrides, fluorides and chlorides, and comparison with first-order relativistic calculations. *J. Chem. Phys.* **1998**, *109*, 392−399.

(197) Bussy, A.; Hutter, J. Efficient and low-scaling linear-response time-dependent density functional theory implementation for core-level spectroscopy of large and periodic systems. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4736−4746.

(198) Ljungberg, M. P.; Koval, P.; Ferrari, F.; Foerster, D.; Sánchez-Portal, D. Cubic-Scaling Iterative Solution of the Bethe-Salpeter Equation for Finite Systems. *Phys. Rev. B* **2015**, *92*, 075422.

(199) Jacquemin, D.; Duchemin, I.; Blondel, A.; Blase, X. Assessment of the Accuracy of the Bethe−Salpeter (BSE/*GW*) Oscillator Strengths. *J. Chem. Theory Comput.* **2016**, *12*, 3969−3981.

(200) Bruneval, F.; Hamed, S. M.; Neaton, J. B. A systematic benchmark of the ab initio Bethe-Salpeter equation approach for low-lying optical excitations of small organic molecules. *J. Chem. Phys.* **2015**, *142*, 244101.

(201) Sander, T.; Maggio, E.; Kresse, G. Beyond the Tamm-Dancoff approximation for extended systems using exact diagonalization. *Phys. Rev. B* **2015**, *92*, 045209.

(202) Liu, C.; Kloppenburg, J.; Yao, Y.; Ren, X.; Appel, H.; Kanai, Y.; Blum, V. All-electron ab initio Bethe-Salpeter equation approach to neutral excitations in molecules with numeric atom-centered orbitals. *J. Chem. Phys.* **2020**, *152*, 044105.

(203) Blase, X.; Duchemin, I.; Jacquemin, D.; Loos, P.-F. The Bethe−Salpeter Equation Formalism: From Physics to Chemistry. *J. Phys. Chem. Lett.* **2020**, *11*, 7371−7382.

(204) Bechstedt, F. *Many-Body Approach to Electronic Excitations: Concepts and Applications*; *Springer Series in Solid-State Sciences*; Springer-Verlag: Berlin, Heidelberg, 2015; Vol. 181.

(205) Mewes, S. A.; Plasser, F.; Krylov, A.; Dreuw, A. Benchmarking Excited-State Calculations Using Exciton Properties. *J. Chem. Theory Comput.* **2018**, *14*, 710−725.

(206) Ullrich, C. A. *Time-Dependent Density-Functional Theory: Concepts and Applications*; Oxford University Press: New York, 2012.

(207) Bruneval, F.; Rangel, T.; Hamed, S. M.; Shao, M.; Yang, C.; Neaton, J. B. Molgw 1: Many-body Perturbation Theory Software for Atoms, Molecules, and Clusters. *Comput. Phys. Commun.* **2016**, *208*, 149−161.

(208) Knysh, I.; Lipparini, F.; Blondel, A.; Duchemin, I.; Blase, X.; Loos, P.-F.; Jacquemin, D. Reference CC3 Excitation Energies for Organic Chromophores: Benchmarking TD-DFT, BSE/GW, and Wave Function Methods. *J. Chem. Theory Comput.* **2024**, *20*, 8152−8174.

(209) Kendall, R. A.; Dunning, 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.

(210) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.

(211) Kunert, T.; Schmidt, R. Non-adiabatic quantum molecular dynamics: General formalism and case study H2+ in strong laser fields. *Eur. Phys. J. D* **2003**, *25*, 15−24.

(212) Castro, A.; Marques, M. A. L.; Rubio, A. Propagators for the time-dependent Kohn-Sham equations. *J. Chem. Phys.* **2004**, *121*, 3425−3433.

(213) Moler, C.; Loan, C. V. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Rev.* **2003**, *45*, 3−49.

(214) Andermatt, S.; Bani-Hashemian, M. H.; Brück, S.; VandeVondele, J.; Luisier, M. Transport simulations with density-matrix-based real-time time-dependant density functional theory. In *2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*; Institute of Electrical and Electronics Engineers, 2017, pp 177−180.

(215) Bertsch, G. F.; Iwata, J.-I.; Rubio, A.; Yabana, K. Real-space, real-time method for the dielectric function. *Phys. Rev. B* **2000**, *62*, 7998−8002.

(216) Yabana, K.; Sugiyama, T.; Shinohara, Y.; Otobe, T.; Bertsch, G. F. Time-dependent density functional theory for strong electro-magnetic fields in crystalline solids. *Phys. Rev. B* **2012**, *85*, 045134.

(217) Pemmaraju, C.; Vila, F.; Kas, J.; Sato, S.; Rehr, J.; Yabana, K.; Prendergast, D. Velocity-gauge real-time TDDFT within a numerical atomic orbital basis set. *Comput. Phys. Commun.* **2018**, *226*, 30−38.

(218) Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601−659.

(219) Marek, S.; Wilhelm, J. Linear and Nonlinear Optical Properties of Molecules from Real-Time Propagation Based on the Bethe-Salpeter Equation. *J. Chem. Theory Comput.* **2025**, *21*, 9814−9822.

(220) Attaccalite, C.; Grüning, M.; Marini, A. Real-time approach to the optical properties of solids and nanostructures: Time-dependent Bethe-Salpeter equation. *Phys. Rev. B* **2011**, *84*, 245110.

(221) Casida, M. E.; Chong, D. P. Physical interpretation and assessment of the Coulomb-hole and screened-exchange approximation for molecules. *Phys. Rev. A* **1989**, *40*, 4837−4848.

(222) Farid, B.; Daling, R.; Lenstra, D.; van Haeringen, W. GW approach to the calculation of electron self-energies in semiconductors. *Phys. Rev. B* **1988**, *38*, 7530−7534.

(223) Betzinger, M.; Friedrich, C.; Görling, A.; Blügel, S. Precise all-electron dynamical response functions: Application to COHSEX and the RPA correlation energy. *Phys. Rev. B* **2015**, *92*, 245101.

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

(225) Mattiat, J.; Luber, S. Comparison of Length, Velocity, and Symmetric Gauges for the Calculation of Absorption and Electric Circular Dichroism Spectra with Real-Time Time-Dependent Density Functional Theory. *J. Chem. Theory Comput.* **2022**, *18*, 5513−5526.

(226) Müller, C.; Sharma, M.; Sierka, M. Real-time time-dependent density functional theory using density fitting and the continuous fast multipole method. *J. Comput. Chem.* **2020**, *41*, 2573−2582.

(227) 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* **2006**, *243*, 1121−1138.

(228) Li, X.; Govind, N.; Isborn, C.; DePrince, A. E., III; Lopata, K. Real-time time-dependent electronic structure theory. *Chem. Rev.* **2020**, *120*, 9951−9993.

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

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

(231) Leucke, M.; Panadés-Barrueta, R. L.; Bas, E. E.; Golze, D. Analytic continuation component of the GreenX library: robust Padé approximants with symmetry constraints. *J. Open Source Softw.* **2025**, *10*, 7859.

(232) Barbatti, M.; Bondanza, M.; Crespo-Otero, R.; Demoulin, B.; Dral, P. O.; Granucci, G.; Kossoski, F.; Lischka, H.; Mennucci, B.; Mukherjee, S.; et al. Newton-X Platform: New Software Developments for Surface Hopping and Nuclear Ensembles. *J. Chem. Theory Comput.* **2022**, *18*, 6851−6865.

(233) Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H. Newton-X: a surface-hopping program for nonadiabatic molecular dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 26−33.

(234) Barbatti, M.; Granucci, G.; Persico, M.; Ruckenbauer, M.; Vazdar, M.; Eckert-Maksić, M.; Lischka, H. The on-the-fly surface-hopping program system Newton-X: Application to ab initio simulation of the nonadiabatic photodynamics of benchmark systems. *J. Photochem. Photobiol., A* **2007**, *190*, 228−240.

(235) Cederbaum, L. S.; Domcke, W.; Schirmer, J. Many-body theory of core holes. *Phys. Rev. A* **1980**, *22*, 206.

(236) DeBeer George, S.; Petrenko, T.; Neese, F. Time-dependent density functional calculations of ligand K-edge X-ray absorption spectra. *Inorg. Chim. Acta* **2008**, *361*, 965−972.

(237) Müller, P.; Karhan, K.; Krack, M.; Gerstmann, U.; Schmidt, W. G.; Bauer, M.; Kühne, T. D. Impact of finite-temperature and condensed-phase effects on theoretical X-ray absorption spectra of transition metal complexes. *J. Comput. Chem.* **2019**, *40*, 712−716.

(238) Franco de Carvalho, F.; Curchod, B. F.; Penfold, T. J.; Tavernelli, I. Derivation of spin-orbit couplings in collinear linear-response TDDFT: A rigorous formulation. *J. Chem. Phys.* **2014**, *140*, 144103.

(239) Bussy, A.; Hutter, J. Data for: Efficient and low-scaling linear-response time-dependent density functional theory implementation for core-level spectroscopy of large and periodic systems, 2021. accessed: 2025−08−23..

(240) Lopata, K.; Van Kuiken, B.; Khalil, M.; Govind, N. Linear-response and real-time time-dependent density functional theory

studies of core-level near-edge X-ray absorption. *J. Chem. Theory Comput.* **2012**, *8*, 3284−3292.

(241) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753−8765.

(242) Khaliullin, R. Z.; Kühne, T. D. Microscopic properties of liquid water from combined ab initio molecular dynamics and energy decomposition studies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15746−15766.

(243) Mo, Y.; Gao, J.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530−5538.

(244) Azar, R. J.; Horn, P. R.; Sundstrom, E. J.; Head-Gordon, M. Useful lower limits to polarization contributions to intermolecular interactions using a minimal basis of localized orthogonal orbitals: Theory and analysis of the water dimer. *J. Chem. Phys.* **2013**, *138*, 084102.

(245) Stoll, H.; Wagenblast, G.; Preuß, H. On the use of local basis sets for localized molecular orbitals. *Theor. Chim. Acta* **1980**, *57*, 169−178.

(246) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.

(247) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. *J. Chem. Phys.* **2008**, *128*, 184112.

(248) Khaliullin, R. Z.; VandeVondele, J.; Hutter, J. Efficient Linear-Scaling Density Functional Theory for Molecular Systems. *J. Chem. Theory Comput.* **2013**, *9*, 4421−4427.

(249) Staub, R.; Iannuzzi, M.; Khaliullin, R. Z.; Steinmann, S. N. Energy Decomposition Analysis for Metal Surface‚ÄìAdsorbate Interactions by Block Localized Wave Functions. *J. Chem. Theory Comput.* **2019**, *15*, 265−275.

(250) Shi, Y.; Scheiber, H.; Khaliullin, R. Z. Contribution of the Covalent Component of the Hydrogen-Bond Network to the Properties of Liquid Water. *J. Phys. Chem. A* **2018**, *122*, 7482−7490.

(251) Scheiber, H.; Shi, Y.; Khaliullin, R. Z. Communication: Compact orbitals enable low-cost linear-scaling ab initio molecular dynamics for weakly-interacting systems. *J. Chem. Phys.* **2018**, *148*, 231103.

(252) Luo, Z.; Friscic, T.; Khaliullin, R. Z. Why pregnenolone and progesterone, two structurally similar steroids, exhibit remarkably different cocrystallization with aromatic molecules. *Phys. Chem. Chem. Phys.* **2018**, *20*, 898−904.

(253) Ojha, D.; Kaliannan, N. K.; Kühne, T. D. Time-dependent vibrational sum-frequency generation spectroscopy of the air-water interface. *Commun. Chem.* **2019**, *2*, 116.

(254) Ojha, D.; Kühne, T. D. Hydrogen bond dynamics of interfacial water molecules revealed from two-dimensional vibrational sum-frequency generation spectroscopy. *Sci. Rep.* **2021**, *11*, 2456.

(255) Kühne, T. D.; Khaliullin, R. Z. Electronic signature of the instantaneous asymmetry in the first coordination shell of liquid water. *Nat. Commun.* **2013**, *4*, 1450.

(256) Elgabarty, H.; Khaliullin, R. Z.; Kühne, T. D. Covalency of hydrogen bonds in liquid water can be probed by proton nuclear magnetic resonance experiments. *Nat. Commun.* **2015**, *6*, 8318.

(257) Elgabarty, H.; Kühne, T. D. Tumbling with a limp: local asymmetry in water's hydrogen bond network and its consequences. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10397−10411.

(258) Kühne, T. D.; Khaliullin, R. Z. Nature of the Asymmetry in the Hydrogen-Bond Networks of Hexagonal Ice and Liquid Water. *J. Am. Chem. Soc.* **2014**, *136*, 3395−3399.

(259) Yun, Y.; Khaliullin, R. Z.; Jung, Y. Low-Dimensional Confined Ice Has the Electronic Signature of Liquid Water. *J. Phys. Chem. Lett.* **2019**, *10*, 2008−2016.

(260) Salem, M. A.; Kühne, T. D. Insight from energy decomposition analysis on a hydrogen-bond-mediated mechanism for on-water catalysis. *Mol. Phys.* **2020**, *118*, No. e1797920.

(261) Heske, J.; Kühne, T. D.; Antonietti, M. Water in PHI nanopores: modeling adsorption, solvent structure, and thermodynamics. *ACS Omega* **2023**, *8*, 26526−26532.

(262) Ramos-Cordoba, E.; Lambrecht, D. S.; Head-Gordon, M. Charge-transfer and the hydrogen bond: Spectroscopic and structural implications from electronic structure calculations. *Faraday Discuss.* **2011**, *150*, 345−362.

(263) Fransson, T.; Harada, Y.; Kosugi, N.; Besley, N. A.; Winter, B.; Rehr, J. J.; Pettersson, L. G. M.; Nilsson, A. X-ray and Electron Spectroscopy of Water. *Chem. Rev.* **2016**, *116*, 7551−7569.

(264) Lenz, A.; Ojamäe, L. Theoretical IR Spectra for Water Clusters (H2O)n(n= 6−22, 28, 30) and Identification of Spectral Contributions from Different H-Bond Conformations in Gaseous and Liquid Water. *J. Phys. Chem. A* **2006**, *110*, 13388−13393.

(265) Ojha, D.; Karhan, K.; Kühne, T. D. On the hydrogen bond strength and vibrational spectroscopy of liquid water. *Sci. Rep.* **2018**, *8*, 16888.

(266) Elgabarty, H.; Kaliannan, N. K.; Kühne, T. D. Enhancement of the local asymmetry in the hydrogen bond network of liquid water by an ultrafast electric field pulse. *Sci. Rep.* **2019**, *9*, 10002.

(267) Yun, Y.; Khaliullin, R. Z.; Jung, Y. Correlated Local Fluctuations in the Hydrogen Bond Network of Liquid Water. *J. Am. Chem. Soc.* **2022**, *144*, 13127−13136.

(268) Balos, V.; Kaliannan, N. K.; Elgabarty, H.; Wolf, M.; Kühne, T. D.; Sajadi, M. Time-resolved terahertz−Raman spectroscopy reveals that cations and anions distinctly modify intermolecular interactions of water. *Nat. Chem.* **2022**, *14*, 1031−1037.

(269) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Sprik, M.; Hutter, J.; Mohamed, F.; Krack, M.; Parrinello, M. Toward a Monte Carlo program for simulating vapor−liquid phase equilibria from first principles. *Comput. Phys. Commun.* **2005**, *169*, 289−294.

(270) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. Isobaric−isothermal Monte Carlo simulations from first principles: application to liquid water at ambient conditions. *ChemPhysChem* **2005**, *6*, 1894−1901.

(271) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. Simulating fluid-phase equilibria of water from first principles. *J. Phys. Chem. A* **2006**, *110*, 640−646.

(272) Kuo, I. W.; Mundy, C. J.; Eggimann, B. L.; McGrath, M. J.; Siepmann, J. I.; Chen, B.; Vieceli, J.; Tobias, D. J. Structure and Dynamics of the Aqueous Liquid- Vapor Interface: A Comprehensive Particle-Based Simulation Study. *J. Phys. Chem. B* **2006**, *110*, 3738−3746.

(273) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car−Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003**, *90*, 238302.

(274) Schade, R.; Kenter, T.; Elgabarty, H.; Lass, M.; Schütt, O.; Lazzaro, A.; Pabst, H.; Mohr, S.; Hutter, J.; Kühne, T. D.; et al. Towards electronic structure-based ab-initio molecular dynamics simulations with hundreds of millions of atoms. *Parallel Comput.* **2022**, *111*, 102920.

(275) Schade, R.; Kenter, T.; Elgabarty, H.; Lass, M.; Kühne, T. D.; Plessl, C. Breaking the exascale barrier for the electronic structure problem in ab-initio molecular dynamics. *Int. J. High Perform. Comput. Appl.* **2023**, *37*, 530−538.

(276) Arias, T. A.; Payne, M. C.; Joannopoulos, J. D. Ab initio molecular dynamics: Analytically continued energy functionals and insights into iterative solutions. *Phys. Rev. Lett.* **1992**, *69*, 1077−1080.

(277) Arias, T. A.; Payne, M. C.; Joannopoulos, J. D. Ab initio molecular-dynamics techniques extended to large-length-scale systems. *Phys. Rev. B* **1992**, *45*, 1538−1549.

(278) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558.

(279) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal−amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251.

(280) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.

(281) Caravati, S.; Bernasconi, M.; Kühne, T.; Krack, M.; Parrinello, M. Coexistence of tetrahedral-and octahedral-like sites in amorphous phase change materials. *Appl. Phys. Lett.* **2007**, *91*, 171906.

(282) Kühne, T. D.; Krack, M.; Parrinello, M. Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car-Parrinello-like Approach. *J. Chem. Theory Comput.* **2009**, *5*, 235−241.

(283) Cucinotta, C. S.; Miceli, G.; Raiteri, P.; Krack, M.; Kühne, T. D.; Bernasconi, M.; Parrinello, M. Superionic Conduction in Substoichiometric LiAl Alloy: An Ab Initio Study. *Phys. Rev. Lett.* **2009**, *103*, 125901.

(284) Caravati, S.; Bernasconi, M.; Kühne, T. D.; Krack, M.; Parrinello, M. Unravelling the Mechanism of Pressure Induced Amorphization of Phase Change Materials. *Phys. Rev. Lett.* **2009**, *102*, 205502.

(285) Luduena, G. A.; Kühne, T. D.; Sebastiani, D. Mixed Grotthuss and Vehicle Transport Mechanism in Proton Conducting Polymers from Ab initio Molecular Dynamics Simulations. *Chem. Mater.* **2011**, *23*, 1424−1429.

(286) Kühne, T. D.; Pascal, T. A.; Kaxiras, E.; Jung, Y. New Insights into the Structure of the Vapor/Water Interface from Large-Scale First-Principles Simulations. *J. Phys. Chem. Lett.* **2011**, *2*, 105−113.

(287) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

(288) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.

(289) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, No. e1701816.

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

(291) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31*, 1902765.

(292) Kang, P. L.; Shang, C.; Liu, Z. P. Large-Scale Atomic Simulation via Machine Learning Potentials Constructed by Global Potential Energy Surface Exploration. *Acc. Chem. Res.* **2020**, *53*, 2119−2129.

(293) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037−10072.

(294) Fiedler, L.; Shah, K.; Bussmann, M.; Cangi, A. Deep dive into machine learning density functional theory for materials science and chemistry. *Phys. Rev. Mater.* **2022**, *6*, 040301.

(295) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(296) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(297) Natarajan, S. K.; Morawietz, T.; Behler, J. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8356−8371.

(298) Schran, C.; Behler, J.; Marx, D. Automated Fitting of Neural Network Potentials at Coupled Cluster Accuracy: Protonated Water Clusters as Testing Ground. *J. Chem. Theory Comput.* **2020**, *16*, 88−99.

(299) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 8368−8373.

(300) Daru, J.; Forbert, H.; Behler, J.; Marx, D. Coupled Cluster Molecular Dynamics of Condensed Phase Systems Enabled by Machine Learning Potentials: Liquid Water Benchmark. *Phys. Rev. Lett.* **2022**, *129*, 226001.

(301) Quaranta, V.; Hellström, M.; Behler, J. Proton-Transfer Mechanisms at the Water-ZnO Interface: The Role of Presolvation. *J. Phys. Chem. Lett.* **2017**, *8*, 1476−1483.

(302) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.* **2008**, *100*, 185501.

(303) Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. Ab initio quality neural-network potential for sodium. *Phys. Rev. B* **2010**, *81*, 184107.

(304) Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. Microscopic origins of the anomalous melting behavior of sodium under high pressure. *Phys. Rev. Lett.* **2012**, *108*, 115701.

(305) Eckhoff, M.; Behler, J. From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5. *J. Chem. Theory Comput.* **2019**, *15*, 3793−3809.

(306) Eckhoff, M.; Behler, J. Insights into lithium manganese oxide-water interfaces using machine learning potentials. *J. Chem. Phys.* **2021**, *155*, 244703.

(307) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Graphite-diamond phase coexistence study employing a neural-network mapping of the ab initio potential energy surface. *Phys. Rev. B* **2010**, *81*, 100103.

(308) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.* **2011**, *10*, 693−697.

(309) Omranpour, A.; Montero De Hijes, P.; Behler, J.; Dellago, C. Perspective: Atomistic simulations of water and aqueous systems with machine learning potentials. *J. Chem. Phys.* **2024**, *160*, 170901.

(310) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 153101.

(311) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.

(312) Khajehpasha, E. R.; Finkler, J. A.; Kühne, T. D.; Ghasemi, S. A. CENT2: Improved charge equilibration via neural network technique. *Phys. Rev. B* **2022**, *105*, 144106.

(313) Thiemann, F. L.; O'Neill, N.; Kapil, V.; Michaelides, A.; Schran, C. Introduction to machine learning potentials for atomistic simulations. *J. Phys.: Condens. Matter* **2025**, *37*, 073002.

(314) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930−17955.

(315) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032−1050.

(316) Brieuc, F.; Schran, C.; Forbert, H.; Marx, D. *RubNNet4MD: Ruhr-Universitat Bochum Neural Networks for Molecular Dynamics Simulations.* Software Package 2020−2025..

(317) Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **2019**, *99*, 014104.

(318) (a) Bochkarev, A.; Lysogorskiy, Y.; Drautz, R. Graph Atomic Cluster Expansion for Semilocal Interactions beyond Equivariant Message Passing. *Phys. Rev. X* **2024**, *14*, 021036. https://github.com/ICAMS/lammps-user-pace.

(319) Zeng, J.; Zhang, D.; Lu, D.; Mo, P.; Li, Z.; Chen, Y.; Rynik, M.; Huang, L.; Li, Z.; Shi, S.; et al. DeePMD-kit v2: A software package for deep potential models. *J. Chem. Phys.* **2023**, *159*, 054801.

(320) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.

(321) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579.

(322) Schran, C.; Brezina, K.; Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *J. Chem. Phys.* **2020**, *153*, 104105.

(323) Imbalzano, G.; Zhuang, Y.; Kapil, V.; Rossi, K.; Engel, E. A.; Grasselli, F.; Ceriotti, M. Uncertainty estimation for molecular dynamics and sampling. *J. Chem. Phys.* **2021**, *154*, 074102.

(324) Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.

(325) Podryabinkin, E. V.; Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **2017**, *140*, 171−180.

(326) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.

(327) Schran, C.; Thiemann, F. L.; Rowe, P.; Müller, E. A.; Marsalek, O.; Michaelides, A. Machine learning potentials for complex aqueous systems made simple. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2110077118.

(328) Stolte, N.; Daru, J.; Forbert, H.; Marx, D.; Behler, J. Random sampling versus active learning algorithms for machine learning potentials of quantum liquid water. *J. Chem. Theory Comput.* **2025**, *21*, 886−899.

(329) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **1999**, *397*, 601−604.

(330) Marx, D.. In *Computer Simulations in Condensed Matter: From Materials to Chemical Biology (Lecture Notes in Physics 704)*; Ferrario, M., Binder, K., Ciccotti, G., Eds.; Springer, 2006; *Chapter 2*, pp 507−539.

(331) Marx, D. Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations. *ChemPhysChem* **2006**, *7*, 1848−1870.

(332) Li, X.-Z.; Walker, B.; Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6369−6373.

(333) Singh, R.; Azadi, S.; Kühne, T. D. Anharmonicity and finite-temperature effects on the structure, stability, and vibrational spectrum of phase III of solid molecular hydrogen. *Phys. Rev. B* **2014**, *90*, 014110.

(334) Spura, T.; Elgabarty, H.; Kühne, T. D. "On-the-fly" coupled cluster path-integral molecular dynamics: impact of nuclear quantum effects on the protonated water dimer. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14355−14359.

(335) Kessler, J.; Elgabarty, H.; Spura, T.; Karhan, K.; Partovi-Azar, P.; Hassanali, A. A.; Kühne, T. D. Structure and dynamics of the instantaneous water/vapor interface revisited by path-integral and ab initio molecular dynamics simulations. *J. Phys. Chem. B* **2015**, *119*, 10079−10086.

(336) Marsalek, O.; Markland, T. E. Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects. *J. Phys. Chem. Lett.* **2017**, *8*, 1545−1551.

(337) Azadi, S.; Singh, R.; Kühne, T. D. Nuclear quantum effects induce metallization of dense solid molecular hydrogen. *J. Comput. Chem.* **2018**, *39*, 262−268.

(338) Ojha, D.; Henao, A.; Kühne, T. D. Nuclear quantum effects on the vibrational dynamics of liquid water. *J. Chem. Phys.* **2018**, *148*, 102328.

(339) Clark, T.; Heske, J.; Kühne, T. D. Opposing electronic and nuclear quantum effects on hydrogen bonds in H2O and D2O. *ChemPhysChem* **2019**, *20*, 2461−2465.

(340) Schran, C.; Marx, D. Quantum nature of the hydrogen bond from ambient conditions down to ultra-low temperatures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24967−24975.

(341) Ohto, T.; Dodia, M.; Xu, J.; Imoto, S.; Tang, F.; Zysk, F.; Kühne, T. D.; Shigeta, Y.; Bonn, M.; Wu, X.; et al. Accessing the accuracy of density functional theory through structure and dynamics of the water−air interface. *J. Phys. Chem. Lett.* **2019**, *10*, 4914−4919.

(342) Kaliannan, N. K.; Henao Aristizabal, A.; Wiebeler, H.; Zysk, F.; Ohto, T.; Nagata, Y.; Kühne, T. D. Impact of intermolecular

vibrational coupling effects on the sum-frequency generation spectra of the water/air interface. *Mol. Phys.* **2020**, *118*, 1620358.

(343) Ojha, D.; Henao, A.; Zysk, F.; Kühne, T. D. Nuclear quantum effects on the vibrational dynamics of the water−air interface. *J. Chem. Phys.* **2024**, *160*, 204114.

(344) Kac, M. On Distributions of Certain Wiener Functionals. *Trans. Am. Math. Soc.* **1949**, *65*, 1−13.

(345) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, 1965.

(346) Feynman, R. P. *Statistical Mechanics*; Addison-Wesley: Redwood City, 1972.

(347) Barker, J. A. A quantum−statistical Monte Carlo method; path integrals with boundary conditions. *J. Chem. Phys.* **1979**, *70*, 2914−2918.

(348) Chandler, D.; Wolynes, P. G. Exploiting the isomorphism between quantum-theory and classical statistical-mechanics of polyatomic fluids. *J. Chem. Phys.* **1981**, *74*, 4078−4095.

(349) Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **1995**, *67*, 279−355.

(350) Markland, T. E.; Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2018**, *2*, 0109.

(351) Callaway, D. J. E.; Rahman, A. Micro-canonical ensemble formulation of lattice gauge-theory. *Phys. Rev. Lett.* **1982**, *49*, 613−616.

(352) Parrinello, M.; Rahman, A. Study of an F−center in molten KCl. *J. Chem. Phys.* **1984**, *80*, 860−867.

(353) Coalson, R. D. On the connection between Fourier coefficient and discretized cartesian path integration. *J. Chem. Phys.* **1986**, *85*, 926−936.

(354) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117−1157.

(355) Ceriotti, M.; Parrinello, M.; Markland, T. E.; Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *J. Chem. Phys.* **2010**, *133*, 124104.

(356) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains - The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.

(357) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. Efficient and general algorithms for path integral Car−Parrinello molecular dynamics. *J. Chem. Phys.* **1996**, *104*, 5579−5588.

(358) Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J. Chem. Phys.* **2005**, *123*, 144104.

(359) Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932−942.

(360) Schran, C.; Brieuc, F.; Marx, D. Converged Colored Noise Path Integral Molecular Dynamics Study of the Zundel Cation Down to Ultralow Temperatures at Coupled Cluster Accuracy. *J. Chem. Theory Comput.* **2018**, *14*, 5068−5078.

(361) John, C.; Spura, T.; Habershon, S.; Kühne, T. D. Quantum ring-polymer contraction method: Including nuclear quantum effects at no additional computational cost in comparison to ab initio molecular dynamics. *Phys. Rev. E* **2016**, *93*, 043305.

(362) Ceriotti, M.; Manolopoulos, D. E.; Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *J. Chem. Phys.* **2011**, *134*, 084104.

(363) Ceriotti, M.; Manolopoulos, D. E. Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei. *Phys. Rev. Lett.* **2012**, *109*, 100604.

(364) Brieuc, F.; Dammak, H.; Hayoun, M. Quantum Thermal Bath for Path Integral Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2016**, *12*, 1351−1359.

(365) Uhl, F.; Marx, D.; Ceriotti, M. Accelerated path integral methods for atomistic simulations at ultra-low temperatures. *J. Chem. Phys.* **2016**, *145*, 054101.

(366) Cao, J. S.; Voth, G. A. A new perspective on quantum time-correlation functions. *J. Chem. Phys.* **1993**, *99*, 10070−10073.

(367) Craig, I. R.; Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* **2004**, *121*, 3368−3373.

(368) Hele, T. J.; Willatt, M. J.; Muolo, A.; Althorpe, S. C. Communication: Relation of centroid molecular dynamics and ring-polymer molecular dynamics to exact quantum dynamics. *J. Chem. Phys.* **2015**, *142*, 191101.

(369) Voth, G. A. Path−integral centroid methods in quantum statistical mechanics and dynamics. *Adv. Chem. Phys.* **1996**, *93*, 135−218.

(370) Althorpe, S. C. Path-integral approximations to quantum dynamics. *Eur. Phys. J. B* **2021**, *94*, 155.

(371) Althorpe, S. C. Path Integral Simulations of Condensed-Phase Vibrational Spectroscopy. *Annu. Rev. Phys. Chem.* **2024**, *75*, 397−420.

(372) Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *J. Chem. Phys.* **2009**, *130*, 194510.

(373) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* **2014**, *140*, 234116.

(374) Shiga, M. Path integral Brownian chain molecular dynamics: A simple approximation of quantum vibrational dynamics. *J. Comput. Chem.* **2022**, *43*, 1864−1879.

(375) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J.; Klein, M. L. Efficient molecular-dynamics and hybrid Monte-Carlo algorithms for path-integrals. *J. Chem. Phys.* **1993**, *99*, 2796−2808.

(376) Spura, T.; John, C.; Habershon, S.; Kühne, T. D. Nuclear quantum effects in liquid water from path-integral simulations using an ab initio force-matching approach. *Mol. Phys.* **2015**, *113*, 808−822.

(377) Stolte, N.; Daru, J.; Forbert, H.; Behler, J.; Marx, D. Nuclear Quantum Effects in Liquid Water Are Marginal for Its Average Structure but Significant for Dynamics. *J. Phys. Chem. Lett.* **2024**, *15*, 12144−12150.

(378) Malik, R.; Stolte, N.; Forbert, H.; Chandra, A.; Marx, D. Accurate Determination of Isotope Effects on the Dynamics of H-Bond Breaking and Making in Liquid Water. *J. Phys. Chem. Lett.* **2025**, *16*, 3727−3733.

(379) Feynman, R. P. Atomic Theory of the λ Transition in Helium. *Phys. Rev.* **1953**, *91*, 1291−1301.

(380) Troyer, M.; Wiese, U.-J. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Phys. Rev. Lett.* **2005**, *94*, 170201.

(381) Calcavecchia, F.; Pederiva, F.; Kalos, M. H.; Kühne, T. D. Sign problem of the fermionic shadow wave function. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2014**, *90*, 053304.

(382) Ceperley, D. M.; Pollock, E. L. Path-Integral Computation of the Low-Temperature Properties of Liquid ^{4}He. *Phys. Rev. Lett.* **1986**, *56*, 351−354.

(383) Pollock, E. L.; Ceperley, D. M. Path-integral computation of superfluid densities. *Phys. Rev. B* **1987**, *36*, 8343−8352.

(384) Boninsegni, M.; Prokof'ev, N.; Svistunov, B. Worm Algorithm for Continuous-Space Path Integral Monte Carlo Simulations. *Phys. Rev. Lett.* **2006**, *96*, 070601.

(385) Boninsegni, M.; Prokof'ev, N. V.; Svistunov, B. V. Worm algorithm and diagrammatic Monte Carlo: A new approach to continuous-space path integral Monte Carlo simulations. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2006**, *74*, 036701.

(386) Brieuc, F.; Schran, C.; Uhl, F.; Forbert, H.; Marx, D. Converged quantum simulations of reactive solutes in superfluid helium: The Bochum perspective. *J. Chem. Phys.* **2020**, *152*, 210901.

(387) Walewski, L.; Forbert, H.; Marx, D. Reactive path integral quantum simulations of molecules solvated in superfluid helium. *Comput. Phys. Commun.* **2014**, *185*, 884−899.

(388) Pollock, E. L.; Ceperley, D. M. Simulation of quantum many-body systems by path-integral methods. *Phys. Rev. B* **1984**, *30*, 2555−2568.

(389) Aziz, R. A.; Janzen, A. R.; Moldover, M. R. Ab Initio Calculations for Helium: A Standard for Transport Property Measurements. *Phys. Rev. Lett.* **1995**, *74*, 1586−1589.

(390) Pollock, E. L.; Runge, K. J. Finite-size-scaling analysis of a simulation of the 4He superfluid transition. *Phys. Rev. B* **1992**, *46*, 3535−3539.

(391) Keesom, W. H.; Keesom, A. P. New measurements on the specific heat of liquid helium. *Physica* **1935**, *2*, 557−572.

(392) Klein, M. L.; Venables, J. A. *Rare Gas Solids Vol. II*; Academic Press: London, New York, San Francisco, 1977.

(393) Sindzingre, P.; Klein, M. L.; Ceperley, D. M. Path-integral Monte Carlo study of low-temperature 4He clusters. *Phys. Rev. Lett.* **1989**, *63*, 1601−1604.

(394) Zeng, T.; Guillon, G.; Cantin, J. T.; Roy, P.-N. Probing the Superfluid Response of para-Hydrogen with a Sulfur Dioxide Dopant. *J. Phys. Chem. Lett.* **2013**, *4*, 2391−2396.

(395) Zeng, T.; Roy, P. N. Microscopic molecular superfluid response: Theory and simulations. *Rep. Prog. Phys.* **2014**, *77*, 046601.

(396) Brieuc, F.; Schran, C.; Marx, D. Manifestations of local supersolidity of 4He around a charged molecular impurity. *Phys. Rev. Res.* **2023**, *5*, 043083.

(397) Walewski, Ł.; Forbert, H.; Marx, D. Solvation of molecules in superfluid helium enhances the "interaction induced localization" effect. *J. Chem. Phys.* **2014**, *140*, 144305.

(398) Duran Caballero, L.; Schran, C.; Brieuc, F.; Marx, D. Neural network interaction potentials for para-hydrogen with flexible molecules. *J. Chem. Phys.* **2022**, *157*, 074302.

(399) Marx, D.; Parrinello, M. Ab initio path−integral molecular dynamics. *Z. Phys. B* **1994**, *95*, 143−144.

(400) Marx, D.; Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. *J. Chem. Phys.* **1996**, *104*, 4077−4082.

(401) Boese, A. D.; Forbert, H.; Masia, M.; Tekin, A.; Marx, D.; Jansen, G. Constructing simple yet accurate potentials for describing the solvation of HCl/water clusters in bulk helium and nanodroplets. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14550−14564.

(402) Kuchenbecker, D.; Uhl, F.; Forbert, H.; Jansen, G.; Marx, D. Constructing accurate interaction potentials to describe the micro-solvation of protonated methane by helium atoms. *Phys. Chem. Chem. Phys.* **2017**, *19*, 8307−8321.

(403) Uhl, F.; Marx, D. Helium Tagging of Protonated Methane in Messenger Spectroscopy: Does It Interfere with the Fluxionality of CH$_5$$^+$? *Angew. Chem., Int. Ed.* **2018**, *57*, 14792−14795.

(404) Uhl, F.; Marx, D. Quantum Microsolvation of Protonated Methane with 4He: Large-Amplitude Motion Heavily Influences Bosonic Exchange. *Phys. Rev. Lett.* **2019**, *123*, 123002.

(405) Schran, C.; Uhl, F.; Behler, J.; Marx, D. High-dimensional neural network potentials for solvation: The case of protonated water clusters in helium. *J. Chem. Phys.* **2018**, *148*, 102310.

(406) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Black, J. A.; Doll, K.; Heßelmann, A.; Kats, D.; Köhn, A.; Korona, T.; Kreplin, D. A.; et al. The Molpro quantum chemistry package. *J. Chem. Phys.* **2020**, *152*, 144107.

(407) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.

(408) Topolnicki, R.; Brieuc, F.; Schran, C.; Marx, D. Deciphering High-Order Structural Correlations within Fluxional Molecules from Classical and Quantum Configurational Entropy. *J. Chem. Theory Comput.* **2020**, *16*, 6785−6794.

(409) Schran, C.; Brieuc, F.; Marx, D. Transferability of machine learning potentials: Protonated water neural network potential applied to the protonated water hexamer. *J. Chem. Phys.* **2021**, *154*, 051101.

(410) Larsson, H. R.; Schröder, M.; Beckmann, R.; Brieuc, F.; Schran, C.; Marx, D.; Vendrell, O. State-resolved infrared spectrum of the protonated water dimer: revisiting the characteristic proton transfer doublet peak. *Chem. Sci.* **2022**, *13*, 11119−11125.

(411) Beckmann, R.; Topolnicki, R.; Marx, D. Deciphering the Impact of Helium Tagging on Flexible Molecules: Probing Micro-

solvation Effects of Protonated Acetylene by Quantum Configurational Entrop. *J. Phys. Chem. A* **2023**, *127*, 2460−2471.

(412) Simko, I.; Schran, C.; Brieuc, F.; Fabri, C.; Asvany, O.; Schlemmer, S.; Marx, D.; Császár, A. G. Quantum Nuclear Delocalization and its Rovibrational Fingerprints. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202306744.

(413) Davies, J. A.; Schran, C.; Brieuc, F.; Marx, D.; Ellis, A. M. Onset of Rotational Decoupling for a Molecular Ion Solvated in Helium: From Tags to Rings and Shells. *Phys. Rev. Lett.* **2023**, *130*, 083001.

(414) Beckmann, R.; Schran, C.; Brieuc, F.; Marx, D. Theoretical infrared spectroscopy of protonated methane isotopologues. *Phys. Chem. Chem. Phys.* **2024**, *26*, 22846−22852.

(415) Beckmann, R.; Brieuc, F.; Schran, C.; Marx, D. Infrared Spectra at Coupled Cluster Accuracy from Neural Network Representations. *J. Chem. Theory Comput.* **2022**, *18*, 5492−5501.

(416) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The atomic simulation environment - a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

(417) Brehm, M.; Kirchner, B. TRAVIS − A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **2011**, *51* (8), 2007−2023.

(418) Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS − A Free Analyzer for Trajectories from Molecular Simulation. *J. Chem. Phys.* **2020**, *152* (16), 164105.

(419) Lindner, J.; Cringus, D.; Pshenichnikov, M. S.; Vöhringer, P. Anharmonic Bend−Stretch Coupling in Neat Liquid Water. *Chem. Phys.* **2007**, *341*, 326−335.

(420) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(421) Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications. *J. Phys. Chem. B* **1997**, *101*, 10506−10517.

(422) Gordon, R. G. Molecular motion in infrared and raman spectra. *J. Chem. Phys.* **1965**, *43*, 1307−1312.

(423) Shimizu, H. Time-Correlation Function of Molecular Random Motion and Shape of Spectral Bands. *J. Chem. Phys.* **1965**, *43*, 2453−2465.

(424) Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics II*; Springer: Berlin, 1991.

(425) Spaldin, N. A. A beginner's guide to the modern theory of polarization. *J. Solid State Chem.* **2012**, *195*, 2−10.

(426) King-Smith, R. D.; Vanderbilt, D. Theory of Polarization of Crystalline Solids. *Phys. Rev. B* **1993**, *47*, 1651−1654.

(427) Resta, R. Macroscopic Electric Polarization As a Geometric Quantum Phase. *Europhys. Lett.* **1993**, *22*, 133−138.

(428) Resta, R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Rev. Mod. Phys.* **1994**, *66*, 899−915.

(429) Resta, R. Quantum-Mechanical Position Operator in Extended Systems. *Phys. Rev. Lett.* **1998**, *80*, 1800−1803.

(430) Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. London, Ser. A* **1984**, *392*, 45−57.

(431) Mead, C. A. The geometric phase in molecular systems. *Rev. Mod. Phys.* **1992**, *64*, 51.

(432) Kirchner, B.; Hutter, J. Solvent effects on electronic properties from Wannier functions in a dimethyl sulfoxide/water mixture. *J. Chem. Phys.* **2004**, *121*, 5133−5142.

(433) Silvestrelli, P. L.; Parrinello, M. Water Molecule Dipole in the Gas and in the Liquid Phase — Erratum. *Phys. Rev. Lett.* **1999**, *82*, 5415.

(434) Silvestrelli, P. L.; Parrinello, M. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **1999**, *111*, 3572−3580.

(435) Boys, S. F. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296−299.

(436) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for *ab initio* and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916−4926.

(437) Wannier, G. H. The Structure of Electronic Excitation Levels in Insulating Crystals. *Phys. Rev.* **1937**, *52*, 191−197.

(438) Marzari, N.; Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **1997**, *56*, 12847−12865.

(439) Silvestrelli, P. L.; Marzari, N.; Vanderbilt, D.; Parrinello, M. Maximally-localized Wannier functions for disordered systems: Application to amorphous silicon. *Solid State Commun.* **1998**, *107*, 7−11.

(440) Silvestrelli, P. L. Maximally localized Wannier functions for simulations with supercells of general symmetry. *Phys. Rev. B* **1999**, *59*, 9703−9706.

(441) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally localized Wannier functions: Theory and applications. *Rev. Mod. Phys.* **2012**, *84*, 1419−1475.

(442) Ambrosetti, A.; Silvestrelli, P. L. Introduction to maximally localized Wannier functions. *Rev. Comput. Chem.* **2016**, *29*, 327−368.

(443) Berghold, G.; Mundy, C. J.; Romero, A. H.; Hutter, J.; Parrinello, M. General and efficient algorithms for obtaining maximally localized Wannier functions. *Phys. Rev. B* **2000**, *61*, 10040−10048.

(444) Resta, R.; Sorella, S. Electron Localization in the Insulating State. *Phys. Rev. Lett.* **1999**, *82*, 370−373.

(445) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457−464.

(446) Jacobi, C. Über ein Leichtes Verfahren die in der Theorie der Säcularstörungen Vorkommenden Gleichungen Numerisch Aufzulösen. *J. für die Reine Angewandte Math.* **1846**, *30*, 51−94.

(447) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6654−6659.

(448) Ojha, D.; Kühne, T. D. "On-the-fly" calculation of the Vibrational Sum-frequency Generation Spectrum at the Air-water Interface. *Molecules* **2020**, *25*, 3939.

(449) Partovi-Azar, P.; Kühne, T. D. Full Assignment of ab-initio Raman Spectra at Finite Temperatures Using Wannier Polarizabilities: Application to Cyclohexane Molecule in Gas Phase. *Micromachines* **2021**, *12*, 1212.

(450) Partovi-Azar, P.; Kühne, T. D. Efficient "On-the-Fly" calculation of Raman Spectra from Ab-Initio molecular dynamics: Application to hydrophobic/hydrophilic solutes in bulk water. *J. Comput. Chem.* **2015**, *36*, 2188−2192.

(451) Partovi-Azar, P.; Kühne, T. D.; Kaghazchi, P. Evidence for the existence of Li 2 S 2 clusters in lithium−sulfur batteries: ab initio Raman spectroscopy simulation. *Phys. Chem. Chem. Phys.* **2015**, *17*, 22009−22014.

(452) Partovi-Azar, P. Efficient method for estimating the dynamics of the full polarizability tensor during ab initio molecular dynamics simulations. *Phys. Rev. B* **2023**, *108*, 235157.

(453) Thomas, M.; Brehm, M.; Kirchner, B. Voronoi Dipole Moments for the Simulation of Bulk Phase Vibrational Spectra. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3207−3213.

(454) Brehm, M.; Thomas, M. Optimized Atomic Partial Charges and Radii Defined by Radical Voronoi Tessellation of Bulk Phase Simulations. *Molecules* **2021**, *26* (7), 1875.

(455) Voronoi, G. Nouvelles Applications Des Paramètres Continus à La Théorie Des Formes Quadratiques. Deuxième Mémoire. Recherches Sur Les Parallélloèdres Primitifs. *J. für die Reine Angewandte Math.* **1908**, *1908*, 198−287.

(456) Medvedev, N. N. The Algorithm for Three-Dimensional Voronoi Polyhedra. *J. Comput. Phys.* **1986**, *67*, 223−229.

(457) Pollak, M.; Rein, R. Molecular-Orbital Studies of Inter-molecular Interaction Energies. II. Approximations Concerned with

Coulomb Interactions and Comparison of the Two London Schemes. *J. Chem. Phys.* **1967**, *47*, 2045−2052.

(458) Becke, A. D. A Multicenter Numerical Integration Scheme for Polyatomic Molecules. *J. Chem. Phys.* **1988**, *88*, 2547−2553.

(459) Gellatly, B. J.; Finney, J. L. Calculation of Protein Volumes: An Alternative to the Voronoi Procedure. *J. Mol. Biol.* **1982**, *161*, 305−322.

(460) Richards, F. M. The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density. *J. Mol. Biol.* **1974**, *82*, 1−14.

(461) Brehm, M.; Thomas, M. Computing Bulk Phase Raman Optical Activity Spectra from ab initio Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2017**, *8* (14), 3409−3414.

(462) Brehm, M.; Thomas, M. Computing Bulk Phase Resonance Raman Spectra from ab initio Molecular Dynamics and Real-Time TDDFT. *J. Chem. Theory Comput.* **2019**, *15* (7), 3901−3905.

(463) Yang, Y.; Cheramy, J.; Brehm, M.; Xu, Y. Raman Optical Activity of *N*-Acetyl-l-Cysteine in Water and in Methanol: The "Clusters-in-a-Liquid" Model and ab initio Molecular Dynamics Simulations. *ChemPhysChem* **2022**, *23*, No. e202200161.

(464) Brehm, M. libvori: Voronoi Integration for Electrostatic Properties from Electron Density, 2025. https://brehm-research.de/libvori.php (accessed 13 09.

(465) Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos* **2009**, *19*, 041111.

(466) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. Covalent radii revisited. *Dalton Trans.* **2008**, *2832*, 2832.

(467) Brehm, M.; Kirchner, B. TRAVIS - A free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **2011**, *51*, 2007−2023.

(468) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253−287.

(469) Heaton, R. J.; Madden, P. A.; Clark, S. J.; Jahn, S. Condensed Phase Ionic Polarizabilities from Plane Wave Density Functional Theory Calculations. *J. Chem. Phys.* **2006**, *125*, 144104.

(470) Salanne, M.; Vuilleumier, R.; Madden, P. A.; Simon, C.; Turq, P.; Guillot, B. Polarizabilities of Individual Molecules and Ions in Liquids from First Principles. *J. Phys.: Condens. Matter* **2008**, *20*, 494207.

(471) Wan, Q.; Spanu, L.; Galli, G. A.; Gygi, F. Raman Spectra of Liquid Water from ab initio Molecular Dynamics: Vibrational Signatures of Charge Fluctuations in the Hydrogen Bond Network. *J. Chem. Theory Comput.* **2013**, *9*, 4124−4130.

(472) Luber, S.; Iannuzzi, M.; Hutter, J. Raman spectra from ab initio molecular dynamics and its application to liquid S-methyloxirane. *J. Chem. Phys.* **2014**, *141*, 094503.

(473) Albrecht, A. C. On the Theory of Raman Intensities. *J. Chem. Phys.* **1961**, *34*, 1476−1484.

(474) Tang, J.; Albrecht, A. C. Studies in Raman Intensity Theory. *J. Chem. Phys.* **1968**, *49*, 1144−1154.

(475) Albrecht, A. C.; Hutley, M. C. On the Dependence of Vibrational Raman Intensity on the Wavelength of Incident Light. *J. Chem. Phys.* **1971**, *55*, 4438−4443.

(476) Lee, S.-Y.; Heller, E. J. Time-Dependent Theory of Raman Scattering. *J. Chem. Phys.* **1979**, *71*, 4777−4788.

(477) Heller, E. J. The Semiclassical Way to Molecular Spectros-copy. *Acc. Chem. Res.* **1981**, *14*, 368−375.

(478) Heller, E. J.; Sundberg, R.; Tannor, D. J. Simple Aspects of Raman Scattering. *J. Phys. Chem.* **1982**, *86*, 1822−1833.

(479) Jensen, L.; Zhao, L. L.; Autschbach, J.; Schatz, G. C. Theory and Method for Calculating Resonance Raman Scattering from Resonance Polarizability Derivatives. *J. Chem. Phys.* **2005**, *123*, 174110.

(480) Jensen, L.; Autschbach, J.; Schatz, G. C. Finite Lifetime Effects on the Polarizability within Time-Dependent Density-Functional Theory. *J. Chem. Phys.* **2005**, *122*, 224115.

(481) Andermatt, S.; Cha, J.; Schiffmann, F.; VandeVondele, J. Combining Linear-Scaling DFT with Subsystem DFT in Born−

Oppenheimer and Ehrenfest Molecular Dynamics Simulations: From Molecules to a Virus in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 3214−3227.

(482) Provorse, M. R.; Isborn, C. M. Electron Dynamics with Real-Time Time-Dependent Density Functional Theory. *Int. J. Quantum Chem.* **2016**, *116*, 739−749.

(483) Goings, J. J.; Lestrange, P. J.; Li, X. Real-Time Time-Dependent Electronic Structure Theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1341.

(484) Chen, H.; Mcmahon, J. M.; Ratner, M. A.; Schatz, G. C. Classical Electrodynamics Coupled to Quantum Mechanics for Calculation of Molecular Optical Properties: A RT-TDDFT/FDTD Approach. *J. Phys. Chem. C* **2010**, *114*, 14384−14392.

(485) Ibaceta-Jaña, J.; Chugh, M.; Novikov, A. S.; Mirhosseini, H.; Kühne, T. D.; Szyszka, B.; Wagner, M. R.; Muydinov, R. Do lead halide hybrid perovskites have hydrogen bonds? *J. Phys. Chem. C* **2022**, *126*, 16215−16226.

(486) Brehm, M.; Thomas, M. An Efficient Lossless Compression Algorithm for Trajectories of Atom Positions and Volumetric Data. *J. Chem. Inf. Model.* **2018**, *58* (10), 2092−2107.

(487) Zhang, C.; Khaliullin, R. Z.; Bovi, D.; Guidoni, L.; Kühne, T. D. Vibrational Signature of Water Molecules in Asymmetric Hydrogen Bonding Environments. *J. Phys. Chem. Lett.* **2013**, *4*, 3245−3250.

(488) Zhang, C.; Guidoni, L.; Kühne, T. D. Competing factors on the frequency separation between the OH stretching modes in water. *J. Mol. Liq.* **2015**, *205*, 42−45.

(489) Nonella, M.; Mathias, G.; Tavan, P. Infrared spectrum of p−benzoquinone in water obtained from a QM/MM hybrid molecular dynamics simulation. *J. Phys. Chem. A* **2003**, *107*, 8638−8647.

(490) Schmitz, M.; Tavan, P. Vibrational spectra from atomic fluctuations in dynamics simulations I. Theory, limitations and a sample application. *J. Chem. Phys.* **2004**, *121*, 12233−12246.

(491) Pejov, L.; Spångberg, D.; Hermansson, K. Using MD Snapshots in ab initio and DFT Calculations: OH Vibrations in the First Hydration Shell around Li^+(aq). *J. Phys. Chem. A* **2005**, *109*, 5144−5152.

(492) Wheeler, R. A.; Dong, H.; Boesch, S. E. Quasiharmonic Vibrations of Water, Water Dimer, and Liquid Water from Principal Component Analysis of Quantum or QM/MM Trajectories. *ChemPhysChem* **2003**, *4*, 382−384.

(493) Gaigeot, M.-P.; Sprik, M. Ab initio molecular dynamics computation of the infrared spectrum of aqueous uracil. *J. Phys. Chem. B* **2003**, *107*, 10344−10358.

(494) Mathias, G.; Baer, M. D. Generalized Normal Coordinates for the Vibrational Analysis of Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2028−2039.

(495) Mathias, G.; Ivanov, S. D.; Witt, A.; Baer, M. D.; Marx, D. Infrared Spectroscopy of Fluxional Molecules from (ab initio) Molecular Dynamics: Resolving Large-Amplitude Motion, Multiple Conformations, and Permutational Symmetries. *J. Chem. Theory Comput.* **2012**, *8*, 224−234.

(496) Eckart, C. Some Studies Concerning Rotating Axes and Polyatomic Molecules. *Phys. Rev.* **1935**, *47*, 552−558.

(497) Thomas, M.; Brehm, M.; Hollóczki, O.; Kelemen, Z.; Nyulászi, L.; Pasinszki, T.; Kirchner, B. Simulating the Vibrational Spectra of Ionic Liquid Systems: 1-Ethyl-3-Methylimidazolium Acetate and its Mixtures. *J. Chem. Phys.* **2014**, *141*, 024510.

(498) CP2K Dashboard. https://dashboard.cp2k.org/index.html, 2024; (accessed, 2025 10 18).

(499) (a) Frigo, M.; Johnson, S. G. The design and implementation of FFTW3. *Proc. IEEE* **2005**, *93*, 216−231. See. http://www.fftw.org.

(500) Ramaswami, A.; Kenter, T.; Kühne, T. D.; Plessl, C. Efficient ab-initio molecular dynamic simulations by offloading fast Fourier transformations to FPGAs. In *2020 30th International Conference on Field-Programmable Logic and Applications (FPL)*; Institute of Electrical and Electronics Engineers, 2020, pp 353−354.

(501) Ramaswami, A.; Kenter, T.; Kühne, T. D.; Plessl, C. Evaluating the design space for offloading 3D FFT calculations to an FPGA for high-performance computing. In *International Symposium on Applied Reconfigurable Computing*; Springer Nature, 2021, pp 285−294.

(502) Wu, X.; Kenter, T.; Schade, R.; Kühne, T. D.; Plessl, C. Computing and compressing electron repulsion integrals on FPGAs. In *2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)*; Institute of Electrical and Electronics Engineers, 2023, pp 162−173.

(503) *FFTFPGA: An OpenCL based library for Fast Fourier Transformations for FPGAs.* https://github.com/pc2/fft3d-fpga, 2024; (accessed, 2025 10 01).

(504) Heinecke, A.; Henry, G.; Hutchinson, M.; Pabst, H. LIBXSMM: Accelerating Small Matrix Multiplications by Runtime Code Generation. In *SC16: International Conference for High Performance Computing, Networking, Storage and Analysis*; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2016.

(505) Marek, A.; Blum, V.; Johanni, R.; Havu, V.; Lang, B.; Auckenthaler, T.; Heinecke, A.; Bungartz, H.-J.; Lederer, H. The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *J. Phys.: Condens. Matter* **2014**, *26*, 213201.

(506) *cuSOLVERMp: A High-Performance CUDA Library for Distributed Dense Linear Algebra.* https://docs.nvidia.com/cuda/cusolvermp, 2024; (accessed, 2025 10 19).

(507) (a) Solcà, R.; Simberg, M.; Meli, R.; Invernizzi, A.; Reverdell, A.; Biddiscombe, J. DLA-Future: A Task-Based Linear Algebra Library Which Provides a GPU-Enabled Distributed Eigensolver. In *Asynchronous Many-Task Systems and Applications*; Springer Nature: Cham, 2024; pp 135−141. See https://github.com/eth-cscs/DLA-Future.

(508) Kozhevnikov, A.; Taillefumier, M.; Frasch, S.; Pintarelli, V. J., Simon; Schulthess, T. Domain specific library for electronic structure calculations. 2019; https://github.com/electronic-structure/SIRIUS, (accessed: 2025 10 02).

(509) (a) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8499−8512. See. https://github.com/dftd4/dftd4.

(510) (a) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671. See. https://github.com/tblite/tblite.

(511) (a) Oleynichenko, A. V.; Zaitsevskii, A.; Mosyagin, N. S.; Petrov, A. N.; Eliav, E.; Titov, A. V. LIBGRPP: A library for the evaluation of molecular integrals of the generalized relativistic pseudopotential operator over Gaussian functions. *Symmetry* **2023**, *15*, 197. See. https://github.com/aoleynichenko/libgrpp.

(512) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.

(513) Togo, A.; Tanaka, I. Spglib: a software library for crystal symmetry search, 2018. https://github.com/spglib/spglib (accessed 18 10 2025).

(514) (a) Rocha, A. R.; Garcia-Suarez, V. M.; Bailey, S. W.; Lambert, C. J.; Ferrer, J.; Sanvito, S. Towards molecular spintronics. *Nat. Mater.* **2005**, *4*, 335−339. See. https://github.com/StefanoSanvitoGroup/libsmeagol.

(515) (a) Posenitskiy, E.; Chilkuri, V. G.; Ammar, A.; Hapka, M.; Pernal, K.; Shinde, R.; Landinez Borda, E. J.; Filippi, C.; Nakano, K.; Kohulák, O.; et al. TREXIO: A file format and library for quantum chemistry. *J. Chem. Phys.* **2023**, *158*, 174801. See. https://github.com/trex-coe/trexio.

(516) Azizi, M.; Wilhelm, J.; Golze, D.; Delesma, F. A.; Panadés-Barrueta, R. L.; Rinke, P.; Giantomassi, M.; Gonze, X. Validation of the GreenX library time-frequency component for efficient GW and RPA calculations. *Phys. Rev. B* **2024**, *109*, 245101.

(517) Gamblin, T.; LeGendre, M.; Collette, M. R.; Lee, G. L.; Moody, A.; de Supinski, B. R.; Futral, S. The Spack package manager: bringing order to HPC software chaos. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; Association for Computing Machinery, 2015, pp 1−12.

(518) Holmberg, N.; Laasonen, K. Efficient constrained density functional theory implementation for simulation of condensed phase electron transfer reactions. *J. Chem. Theory Comput.* **2017**, *13*, 587−601.

(519) Holmberg, N.; Laasonen, K. Diabatic model for electrochemical hydrogen evolution based on constrained DFT configuration interaction. *J. Chem. Phys.* **2018**, *149*, 104702.

(520) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.

(521) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1−5.

(522) Akimov, A. V. Libra: an open-source "methodology discovery" library for quantum and classical dynamics simulations. *J. Comput. Chem.* **2016**, *37*, 1626−1649.

(523) Lervik, A.; Riccardi, E.; van Erp, T. S. PyRETIS: A well-done, medium-sized python library for rare events. *J. Comput. Chem.* **2017**, *38*, 2439−2451.

(524) Riccardi, E.; Lervik, A.; Roet, S.; Aarøen, O.; van Erp, T. S. PyRETIS 2: an improbability drive for rare events. *J. Comput. Chem.* **2020**, *41*, 370−377.

(525) Vervust, W.; Zhang, D. T.; Ghysels, A.; Roet, S.; van Erp, T. S.; Riccardi, E. PyRETIS 3: Conquering rare and slow events without boundaries. *J. Comput. Chem.* **2024**, *45*, 1224−1234.

(526) Pizzi, G.; Vitale, V.; Arita, R.; Bluegel, S.; Freimuth, F.; Géranton, G.; Gibertini, M.; Gresch, D.; Johnson, C.; Koretsune, T.; et al. Wannier90 as a Community Code: New Features and Applications. *J. Phys.: Condens. Matter* **2020**, *32*, 165902.

(527) Vogt, J.-R.; Schulz, M.; Mattos, R. S.; Barbatti, M.; Persico, M.; Granucci, G.; Hutter, J.; Hehn, A.-S. A mixed density functional theory and semi-empirical framework for trajectory surface hopping on extended systems. *ChemRxiv* **2025**, 2025-ddxzg.

(528) Zeng, J.; Zhang, D.; Peng, A.; Zhang, X.; He, S.; Wang, Y.; Liu, X.; Bi, H.; Li, Y.; Cai, C.; et al. DeePMD-kit v3: a multiple-backend framework for machine learning potentials. *J. Chem. Theory Comput.* **2025**, *21*, 4375−4385.

(529) Antalík, A.; Levy, A.; Johnson, S. K.; Olsen, J. M. H.; Rothlisberger, U. Making Puzzle Pieces Fit or Reshaping MiMiC for Multiscale Simulations with CP2K and More. *J. Chem. Inf. Model.* **2025**, *65*, 4994−5005.

(530) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

(531) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Comput. Phys. Commun.* **2014**, *185*, 1019−1026.

(532) Kapil, V.; Rossi, M.; Marsalek, O.; Petraglia, R.; Litman, Y.; Spura, T.; Cheng, B.; Cuzzocrea, A.; Meißner, R. H.; Wilkins, D. M.; et al. i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **2019**, *236*, 214−223.

(533) Seeber, P.; Seidenath, S.; Steinmetzer, J.; Gräfe, S. Growing Spicy ONIOMs: Extending and generalizing concepts of ONIOM and many body expansions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2023**, *13*, No. e1644.

(534) Verstraelen, T.; Van Houteghem, M.; Van Speybroeck, V.; Waroquier, M. MD-TRACKS: A Productive Solution for the Advanced Analysis of Molecular Dynamics and Monte Carlo simulations. *J. Chem. Inf. Model.* **2008**, *48*, 2414−2424.

(535) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. ZEOBUILDER: A GUI Toolkit for the Construction of Complex Molecular Structures on the Nanoscale with Building Blocks. *J. Chem. Inf. Model.* **2008**, *48*, 1530−1541.

(536) Ghysels, A.; Verstraelen, T.; Hemelsoet, K.; Waroquier, M.; Van Speybroeck, V. TAMkin: a versatile package for vibrational analysis and chemical kinetics. *J. Chem. Inf. Model.* **2010**, *50*, 1736−1750.

(537) Verstraelen, T.; Adams, W.; Pujal, L.; Tehrani, A.; Kelly, B. D.; Macaya, L.; Meng, F.; Richer, M.; Hernández-Esparza, R.; Yang, X. D.; et al. IOData: A python library for reading, writing, and converting computational chemistry file formats and generating input files. *J. Comput. Chem.* **2021**, *42*, 458−464.

(538) Poeschel, F.; Juncheng, E.; Godoy, W. F.; Podhorszki, N.; Klasky, S.; Eisenhauer, G.; Davis, P. E.; Wan, L.; Gainaru, A.; Gu, J.; et al. Transitioning from file-based HPC workflows to streaming data pipelines with openPMD and ADIOS2. In *Smoky Mountains Computational Sciences and Engineering Conference*; Springer Nature, 2021, pp 99−118.

(539) Gu, J.; Davis, P.; Eisenhauer, G.; Godoy, W.; Huebl, A.; Klasky, S.; Parashar, M.; Podhorszki, N.; Poeschel, F.; Vay, J.; et al. Organizing large data sets for efficient analyses on HPC systems. *J. Phys.: Conf. Ser.* **2022**, *2224*, 012042.

(540) Lu, T. A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn. *J. Chem. Phys.* **2024**, *161*, 082503.

(541) Los, J. H.; Kühne, T. D. Inverse simulated annealing for the determination of amorphous structures. *Phys. Rev. B* **2013**, *87*, 214202.

(542) Los, J. H.; Gabardi, S.; Bernasconi, M.; Kühne, T. D. Inverse simulated annealing: Improvements and application to amorphous InSb. *Comput. Mater. Sci.* **2016**, *117*, 7−14.

(543) Thomsen, B.; Shiga, M. Ab initio study of nuclear quantum effects on sub-and supercritical water. *J. Chem. Phys.* **2021**, *155*, 194107.

(544) Thomsen, B.; Shiga, M. Structures of liquid and aqueous water isotopologues at ambient temperature from ab initio path integral simulations. *Phys. Chem. Chem. Phys.* **2022**, *24*, 10851−10859.

(545) Togo, A.; Chaput, L.; Tadano, T.; Tanaka, I. Implementation strategies in phonopy and phono3py. *J. Phys.: Condens. Matter* **2023**, *35*, 353001.

(546) Schreder, L.; Luber, S. Implementation of frozen density embedding in CP2K and OpenMolcas: CASSCF wavefunctions embedded in a Gaussian and plane wave DFT environment. *J. Chem. Phys.* **2024**, *161*, 144110.

(547) Battaglia, S.; Rossmannek, M.; Rybkin, V. V.; Tavernelli, I.; Hutter, J. A general framework for active space embedding methods with applications in quantum computing. *npj Comput. Mater.* **2024**, *10*, 297.

(548) Johnson, B. A.; Bu, S.; Mundy, C. J.; Ananth, N. MixPI: Mixed-Time Slicing Path Integral Software for Quantized Molecular Dynamics Simulations. *arXiv* **2024**, 11988.

(549) CP2K Benchmark Suite, 2024. https://www.cp2k.org/performance (accessed, 2025 10 19).