



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

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

Year: 2024

**Efficient periodic resolution-of-the-identity Hartree–Fock exchange method with
k-point sampling and Gaussian basis sets**

Bussy, Augustin ; Hutter, Jürg

DOI: <https://doi.org/10.1063/5.0189659>

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

ZORA URL: <https://doi.org/10.5167/uzh-257670>

Journal Article

Published Version



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

Originally published at:

Bussy, Augustin; Hutter, Jürg (2024). Efficient periodic resolution-of-the-identity Hartree–Fock exchange method with k-point sampling and Gaussian basis sets. *Journal of Chemical Physics*, 160(6):064116.

DOI: <https://doi.org/10.1063/5.0189659>

Efficient periodic resolution-of-the-identity Hartree–Fock exchange method with k-point sampling and Gaussian basis sets

Cite as: J. Chem. Phys. 160, 064116 (2024); doi: 10.1063/5.0189659

Submitted: 30 November 2023 • Accepted: 18 January 2024 •

Published Online: 14 February 2024



View Online



Export Citation



CrossMark

Augustin Bussy^{a)}  and Jürg Hutter^{b)} 

AFFILIATIONS

Department of Chemistry, University of Zurich, CH-8057 Zürich, Switzerland

^{a)} Author to whom correspondence should be addressed: augustin.bussy@chem.uzh.ch

^{b)} Electronic mail: hutter@chem.uzh.ch

ABSTRACT

Simulations of condensed matter systems at the hybrid density functional theory level pose significant computational challenges. The elevated costs arise from the non-local nature of the Hartree–Fock exchange (HFX) in conjunction with the necessity to approach the thermodynamic limit. In this work, we address these issues with the development of a new efficient method for the calculation of HFX in periodic systems, employing k-point sampling. We rely on a local atom-specific resolution-of-the-identity scheme, the use of atom-centered Gaussian type orbitals, and the truncation of the Coulomb interaction to limit computational complexity. Our real-space approach exhibits a scaling that is, at worst, linear with the number of k-points. Issues related to basis set diffuseness are effectively addressed through the auxiliary density matrix method. We report the implementation in the CP2K software package, as well as accuracy and performance benchmarks. This method demonstrates excellent agreement with equivalent Γ -point supercell calculations in terms of relative energies and nuclear gradients. Good strong and weak scaling performances, as well as graphics processing unit (GPU) acceleration, make this implementation a promising candidate for high-performance computing.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0189659>

I. INTRODUCTION

Finding their roots in the adiabatic connection theory,^{1–3} hybrid functionals have emerged as an accuracy standard for density functional theory (DFT) calculations of molecules.^{4–8} Beyond the local descriptions provided by functionals in the local density approximation (LDA) and generalized gradient approximation (GGA), hybrid functionals incorporate a fraction of non-local Hartree–Fock exchange (HFX). While such functionals are known to be accurate for the simulation of solids in periodic boundary conditions (PBCs),^{9–15} their computational cost increases dramatically due to their non-local nature and their necessity to approach the thermodynamic limit (TDL). This can be achieved by sampling k-points over the Brillouin zone (BZ) or by considering large supercells at the Γ -point only. Although the former strategy scales better toward the TDL, implementing Γ -point methods in molecular codes is simpler, leading to the availability of both approaches in various

quantum chemistry codes.^{16–23} In addition to an improved description of the electronic structure, hybrid DFT calculations of solids can serve as a foundation for higher level methods, such as GW^{24–26} or RPA.^{27–29} Periodic post-HF methods^{30–33} also benefit from such implementations.

The development of HFX methods for PBC calculations has a rich history and remains an active area of research. In the context of quantum chemistry utilizing local atom-centered basis sets, typically Gaussian type orbitals (GTOs), fundamental work was carried out by Pisani and co-workers in the 1980s.^{34–38} Their approach directly extended molecular HFX to PBCs, involving the calculation of two-electron four-center electron repulsion integrals (ERIs) and the truncation of resulting infinite sums. This latter point is highly non-trivial as the exact exchange energy diverges for systems with finite sampling of the BZ.³⁹ To tackle this, novel approaches were developed, relying on the minimum image convention⁴⁰ (MIC) or the truncated Coulomb (TC) interaction.⁴¹ While these new meth-

ods effectively dealt with the divergence issue, they were initially developed only for Γ -point supercell calculations. More recently, Irmeler and co-workers extended them to \mathbf{k} -point sampling.⁴² Note that both the MIC and TC approaches were adapted from equivalent plane wave schemes.^{43–45} Over the years, various strategies were devised for the efficient evaluation of molecular HFX energy, with some adapted for PBCs. Variations of the resolution-of-the-identity (RI) technique^{46,47} have been employed for calculations with \mathbf{k} -point sampling,^{48–53} while the auxiliary density matrix method (ADMM) was specifically developed for Γ -point calculations.⁵⁴ We note that efficient algorithms have also been developed for plane wave calculations.^{55–59}

In this work, we present the development of a novel atom-specific RI method tailored for the calculation of periodic HFX with \mathbf{k} -point sampling and GTOs. Particular emphasis is put on efficiency, which is enhanced by the adaptation of the ADMM approximation to \mathbf{k} -point sampling, as well as the use of the truncated Coulomb interaction with a fixed range. We discuss the implementation of the method in the open source CP2K software package²³ and demonstrate its potential for high performance computing applications.

II. THEORY

A. General

For simplicity, the following discussion covers the spin-restricted HF theory. Spin-unrestricted and hybrid DFT equations are readily recoverable. In periodic HF calculations employing \mathbf{k} -point sampling, a Hermitian generalized eigenvalue problem has to be solved for each \mathbf{k} -point,

$$\mathbf{F}^{\mathbf{k}} \mathbf{C}^{\mathbf{k}} = \mathbf{S}^{\mathbf{k}} \mathbf{C}^{\mathbf{k}} \boldsymbol{\epsilon}^{\mathbf{k}}, \quad (1)$$

where $F_{\mu\nu}^{\mathbf{k}}$ denotes the Fock matrix and $S_{\mu\nu}^{\mathbf{k}}$ represents the overlap matrix of Bloch atomic orbitals $\varphi_{\mu}^{\mathbf{k}}(\mathbf{r})$. These orbitals are defined as discrete Fourier transforms of translated real-space atomic orbitals (AOs),

$$\varphi_{\mu}^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \varphi_{\mu}(\mathbf{r} - \mathbf{R}), \quad (2)$$

with \mathbf{R} being a unit cell translation vector. The solutions of the eigenvalue problem, $C_{\mu i}^{\mathbf{k}}$ and $\epsilon_i^{\mathbf{k}}$ correspond to the HF crystalline orbitals (COs) and band energies, respectively.

The exact exchange contribution to the Fock matrix can be expressed as

$$K_{\mu\nu}^{\mathbf{k}} = \sum_{\mathbf{k}'} P_{\sigma\lambda}^{\mathbf{k}'} (\mu^{\mathbf{k}} \sigma^{\mathbf{k}'} | \nu^{\mathbf{k}} \lambda^{\mathbf{k}'}), \quad (3)$$

where $P_{\sigma\lambda}^{\mathbf{k}'}$ is the density matrix for a given \mathbf{k} -point \mathbf{k}' . We use the Mulliken notation for two-electron four-center integrals and the Einstein summation convention for AO indices. Alternatively, the exact exchange matrix can be obtained by a Fourier transform of its real-space counterpart

$$K_{\mu\nu}^{\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} K_{\mu\nu}^{\mathbf{R}}, \quad (4)$$

with

$$K_{\mu\nu}^{\mathbf{b}} = \sum_{\mathbf{a}, \mathbf{c}} P_{\sigma,\lambda}^{\mathbf{c}} (\mu^{\mathbf{a}} \sigma^{\mathbf{a}} | \nu^{\mathbf{b}} \lambda^{\mathbf{a}+\mathbf{c}}). \quad (5)$$

Note that the real-space notation $K_{\mu\nu}^{\mathbf{b}}$ corresponds to a matrix element with AO $\varphi_{\mu}(\mathbf{r})$ in the reference cell and $\varphi_{\nu}(\mathbf{r} - \mathbf{b})$ in a periodic image of the cell translated by \mathbf{b} . For all Hermitian matrices in the \mathbf{k} -space, their real-space equivalents exhibit the $F_{\mu\nu}^{\mathbf{b}} = F_{\nu\mu}^{\mathbf{b}}$ symmetry.

Both approaches to building the exact exchange matrix are found in the literature, each with its variations, advantages, and drawbacks. The direct approach outlined in Eq. (3) confines the Bloch AO indices to the unit cell, facilitating the efficient handling of diffuse basis functions. This approach is favored in most plane-wave codes as well as in the occ-RI-K method of Lee and co-workers.^{52,53} Its primary limitation lies in the quadratic scaling cost with respect to the number of \mathbf{k} -points. On the other hand, the cost of building the exact exchange matrix in real-space, as depicted in Eq. (5), is, in principle, independent of the number of \mathbf{k} -points. It entirely depends on the extent of the cell translation vectors \mathbf{a} , \mathbf{b} , and \mathbf{c} , determined by the basis set diffuseness and the range of the exchange potential. Variations of this latter approach are followed by the pioneering work of Pisani and co-workers,³⁴ as well as in more recent publications.^{42,48,51} Hybrid methods combining aspects of both techniques have also emerged.⁶⁰

Considering that most periodic hybrid DFT calculations are performed with small unit cells and dense \mathbf{k} -point meshes, we believe that the real-space approach is more suitable. Indeed, the overall cost scales, at worst, linearly with the number of \mathbf{k} -points, provided that diagonalization dominates. In this work, we undertake several measures to reduce the cost of building the real-space exact exchange matrices and implement them in the CP2K software package. To achieve this, we reduce the extent of \mathbf{a} , \mathbf{b} , and \mathbf{c} cell indices by using the truncated Coulomb operator. Concurrently, we improve the overall computational scaling and efficiency with a novel atom-specific resolution-of-the-identity (RI) scheme. Furthermore, we address the challenge of basis set diffuseness by employing the auxiliary density matrix method (ADMM).⁵⁴ Each of these steps is detailed in the following sections: Sec. II B for the truncated Coulomb operator, Sec. II C for the RI approximation, and Sec. II D for ADMM.

Note that real-space approaches are also well suited for band structure calculations. Given a converged SCF calculation with a dense enough \mathbf{k} -point mesh, the resulting collection of real-space Fock matrices $\mathbf{F}^{\mathbf{b}}$ can be back Fourier transformed to any \mathbf{k} -point path. After diagonalization, eigenvalues are used for the band structure along the given path.

B. Truncated Coulomb operator

In periodic HF calculations, the exact exchange energy is calculated from the trace of Eq. (3) with the density matrix

$$E_x^{\text{HF}} = -\frac{1}{2} \sum_{\mathbf{k}, \mathbf{k}'} P_{\mu\nu}^{\mathbf{k}} P_{\sigma\lambda}^{\mathbf{k}'} (\mu^{\mathbf{k}} \sigma^{\mathbf{k}'} | \nu^{\mathbf{k}} \lambda^{\mathbf{k}'}), \quad (6)$$

which suffers from an integrable singularity at $\mathbf{k} = \mathbf{k}'$ when the standard $1/r_{12}$ Coulomb potential is used. This singularity vanishes in the limit of an infinite number of \mathbf{k} -points. This issue also exists

in periodic Γ -point only calculations and requires infinitely large supercells. This is evidently impractical for any kind of realistic calculations, and over the years, numerous schemes have been developed to tackle the problem. Of particular relevance to this work are real-space approaches applicable to Eq. (5). These include the minimal image convention (MIC) method by Tymczak and co-workers,⁴⁰ the truncated Coulomb (TC) interaction by Guidon and co-workers,⁴¹ and the screened Coulomb interaction by Heyd and co-workers.⁶¹

In this work, our emphasis lies on the truncated Coulomb (TC) technique, which has served as the cornerstone of CP2K periodic exact exchange calculations for over a decade. The machinery for calculating electron repulsion integrals (ERIs) based on the TC operator is readily available and the Γ -point TC implementation is widely established.^{41,62–68} The truncated Coulomb potential is defined as

$$g_{\text{TC}}(r_{12}) = \begin{cases} \frac{1}{r_{12}} & \text{if } r_{12} \leq R_C \\ 0 & \text{if } r_{12} > R_C, \end{cases} \quad (7)$$

where R_C is the truncation radius. In the original TC paper by Spencer and Alavi,⁴⁴ the truncation radius increases with the number of \mathbf{k} -points such that R_C is the radius of the largest sphere fitting in the corresponding Born–von Karman⁶⁹ (BvK) supercell. In CP2K, the cutoff radius is typically fixed to a value yielding converged exact exchange contributions. The Γ -point approach allows for HFX calculations that are linear-scaling with the system size. When applied to a real-space \mathbf{k} -point method, this approach leads to constant costs for building exact exchange matrices as a function of the number of \mathbf{k} -points. The ideal fixed value for the truncation radius comes from a trade-off between performance and accuracy and is system dependent. It is discussed in Sec. III B of this paper. The usual constraint that the truncation radius is, at most, the radius of the largest sphere fitting in the BvK supercell remains unchanged. While we focus on the TC operator in this work, the short range screened Coulomb interaction was also implemented. Extension to any other limited range potential is straightforward.

C. Atom-specific resolution-of-the-identity scheme

From the real-space exact exchange matrix outlined in Eq. (5), it becomes apparent that calculating two-electron four-center electron repulsion integrals (ERIs) comes with an unfavorable scaling with respect to the system and basis set sizes. In Γ -point calculations, ERIs stemming from different periodic images can be summed in advance before being contracted with the density matrix. This typically allows for the computation and storage of all ERIs at the first SCF step, while they can be efficiently retrieved from the memory for all subsequent iterations. However, in four-center \mathbf{k} -point HFX calculations, the ERIs can only be partially pre-summed and storing them in the memory would rapidly become intractable. This would require a complete recalculation of the ERIs at each SCF step, leading to substantial additional costs. On the other hand, RI schemes only require two- and three-center ERIs, making integral storage more manageable. Moreover, previous research⁷⁰ has demonstrated that RI-HFX is particularly efficient for small- to medium-sized dense

systems, which is the typical application case for \mathbf{k} -point calculations. Finally, adopting an RI-based method shifts the computational bottleneck from four-center ERI evaluations to tensor contractions, a task better suited for graphics processing unit (GPU) acceleration. For all these reasons, we developed an RI-based method for HFX periodic \mathbf{k} -point calculations.

In a global resolution-of-the-identity scheme, RI based elements would need to span all possible image cells, leading to the following formula:

$$K_{\mu\nu}^{\mathbf{b}} = \sum_{\mathbf{a}, \mathbf{c}, \mathbf{d}, \mathbf{e}} P_{\sigma, \lambda}^{\mathbf{c}} (\mu^{\mathbf{0}} \sigma^{\mathbf{a}} | P^{\mathbf{e}}) [(P|Q)^{-1}]_{\mathbf{e}, \mathbf{f}} (Q^{\mathbf{f}} | \nu^{\mathbf{b}} \lambda^{\mathbf{a}+\mathbf{c}}), \quad (8)$$

where P and Q represent atom-centered GTOs. $[(P|Q)^{-1}]_{\mathbf{e}, \mathbf{f}}$ corresponds to the \mathbf{e} and \mathbf{f} blocks of an inverse matrix involving RI based elements from all possible periodic images. Such an approach introduces an additional double loop over periodic cell indices \mathbf{e} and \mathbf{f} , which is detrimental for computational complexity. Moreover, storing three-center ERIs with two such periodic cell indices would be challenging. Finally, inverting the two-center RI potential matrix $(P|Q)$ could be prohibitively expensive due to the cubic scaling of such operations. To address these issues, we introduce a local atom-specific RI scheme, in which each atom has a different RI basis defined in its vicinity. With the addition of an RI metric denoted with operator $|\cdot|$, we get

$$K_{\mu_i, \nu_j}^{\mathbf{b}} = \sum_{\mathbf{a}, \mathbf{c}} P_{\sigma, \lambda}^{\mathbf{c}} (\mu_i^{\mathbf{0}} \sigma^{\mathbf{a}} | P_i^{\mathbf{0}}) (P_i^{\mathbf{0}} | R_i^{\mathbf{0}})^{-1} (R_i^{\mathbf{0}} | S_j^{\mathbf{b}}) \times (S_j^{\mathbf{b}} | Q_j^{\mathbf{b}})^{-1} (Q_j^{\mathbf{b}} | \nu_j^{\mathbf{b}} \lambda^{\mathbf{a}+\mathbf{c}}), \quad (9)$$

where indices i and j refer to atoms. The RI metric operator $|\cdot|$ is typically chosen to have a smaller range than the HFX potential for increased sparsity. The RI-SVS scheme⁴⁷ corresponds to the most extreme case, where the RI metric is taken to be the overlap. Note that the choice of an atom-specific RI scheme is driven by the need for translational invariance, which would not be enforced if the local RI basis were defined in terms of the unit cell. We refer to this method as RI-HFXk.

The atom-specific RI basis for atom i in the reference cell, $P_i^{\mathbf{0}}$, should span the same space as all possible $\mu_i^{\mathbf{0}} \sigma^{\mathbf{a}}$ products, where $\varphi_{\mu_i}(\mathbf{r})$ is an AO centered on atom i and $\varphi_{\sigma}(\mathbf{r} - \mathbf{a})$ is a general AO in cell \mathbf{a} . For these products to be non-zero, the AOs have to overlap, which only happens within a region around atom i , where $\varphi_{\mu_i}(\mathbf{r}) > 0$. In practical terms, we define the range of a primitive Gaussian function as R_{\max} , such that $\varphi_{\mu_i}(R_{\max}) = \varepsilon_{\text{PGF}}$. Two functions overlap if $\varphi_{\mu_i}(\mathbf{r}) \varphi_{\sigma}(\mathbf{r} - \mathbf{a}) > \varepsilon_{\text{PGF}}^2$. The extent of the atom-specific RI basis is defined by a sphere centered on atom i , with a radius equal to the range of the most diffuse AO in the system. All atoms within that sphere contribute to their RI basis functions. To avoid a discontinuous change of the basis upon atomic displacements, atoms on the edge of the RI region have lesser weight, enforced with a bump function, as introduced by Sodt and co-workers.⁷¹ Figure 1 illustrates how the atom-specific RI basis is constructed. This choice of a local RI basis is based on the properties of the system and does not require user input, beside the general ε_{PGF} accuracy parameter. Any standard optimized or automatically generated RI basis can be used. Note that the extent of the RI basis is independent of the system size. Further details about the bump function used in this work can be

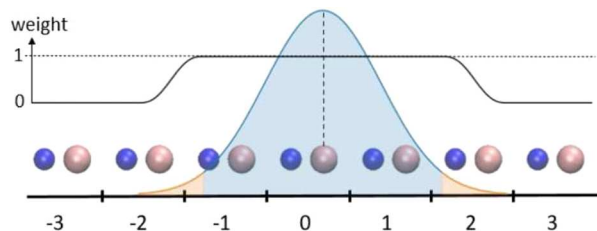


FIG. 1. Illustration of the atom-specific RI basis extent for a fictional one-dimensional two-atom system. Cell indices are on the x axis, and the atom of interest is on the RHS of the reference cell. The Gaussian basis set represents the extent of the most diffuse AO, within which range all atoms contribute to RI basis functions. Atoms in the blue zone contribute fully, while atoms on the edges (orange zone) have a decaying weight. Note that the RI region may extend beyond the reference cell. In CP2K, the default value of ϵ_{PGF} is 10^{-5} . In this implementation, the weight of a distant atom starts decaying at 85% of the RI range, by default.

found in the supplementary material. Other local RI schemes, where each $\mu_i^0 \sigma_j^a$ product is projected on GTOs centered on atoms i in the reference cell and j in cell \mathbf{a} , can be found in the literature.^{48,51} The RI method used in this work is not robust in the definition given by Dunlap.⁷² A strictly robust density fitting^{51,73} scheme would, however, lead to a much greater number of tensor contractions and steep computational costs. Furthermore, the accuracy of RI-HFXk is not very sensitive to the value of ϵ_{PGF} (see the supplementary material).

D. Auxiliary density matrix method

In the auxiliary density matrix method, the HFX energy is approximated using an auxiliary basis set, typically chosen to be smaller and/or less diffuse than the main orbital basis,

$$E_x^{\text{HF}} = E_x^{\text{HF}}[\hat{\mathbf{P}}] + (E_x^{\text{HF}}[\mathbf{P}] - E_x^{\text{HF}}[\hat{\mathbf{P}}]) \approx E_x^{\text{HF}}[\hat{\mathbf{P}}] + (E_x^{\text{GGA}}[\mathbf{P}] - E_x^{\text{GGA}}[\hat{\mathbf{P}}]), \quad (10)$$

where \mathbf{P} and $\hat{\mathbf{P}}$ are the electronic density in the primary and auxiliary basis, respectively. The energy difference between both the basis sets is approximated with a local GGA level functional, which is evaluated much more efficiently than in HF. The ADMM approximation is assumed to be accurate, provided that $\mathbf{P} \approx \hat{\mathbf{P}}$. In its simplest formulation, ADMM2, the density in the auxiliary basis set is obtained by basis projection,

$$\hat{\mathbf{P}}^{\mathbf{k}} = \hat{\mathbf{C}}^{\mathbf{k}} \hat{\mathbf{C}}^{\mathbf{k}\text{T}} = \hat{\mathbf{S}}^{\mathbf{k}-1} \mathbf{Q}^{\mathbf{k}} \mathbf{P}^{\mathbf{k}} \mathbf{Q}^{\mathbf{k}\text{T}} \hat{\mathbf{S}}^{\mathbf{k}-1}, \quad (11)$$

where $\hat{\mathbf{S}}^{\mathbf{k}}$ is the auxiliary–auxiliary basis overlap matrix and $\mathbf{Q}^{\mathbf{k}}$ is the auxiliary–primary basis overlap matrix. Note that with \mathbf{k} -point sampling, this basis projection is only well defined in \mathbf{k} -space. Since we have chosen to work with real-space matrices, the ADMM method involves a Fourier transform of the primary real-space density matrices to \mathbf{k} -space. Subsequently, a basis projection is performed at each \mathbf{k} -point, followed by back Fourier transforms of the auxiliary density matrices from \mathbf{k} -space to real space. Note that the alternative flavors of the ADMM developed by Merlot and co-workers⁷⁴ (ADMMp,

ADMMq, and ADMMs) are also straightforwardly extended to \mathbf{k} -points. Purified wave function fitting, also referred to as ADMM1, is much more cumbersome and was not implemented.

The use of the ADMM with the RI-HFXk method described in this work offers several advantages. Calculations are accelerated by selecting less diffuse auxiliary basis sets, thereby reducing the extent of cell indices \mathbf{a} , \mathbf{b} , and \mathbf{c} in Eq. (9). Additionally, the use of smaller auxiliary basis sets contributes to faster matrix–matrix multiplications and tensor contractions. Finally, the approach enables calculations with more diffuse primary basis sets that would typically be prohibitively expensive.

E. Implementation

Equation (9) for the exact exchange contribution to the Fock matrix is straightforward, but its efficient implementation requires effort. There are indeed numerous aspects that hinder performance. The triple loop over periodic cells is problematic, given that the number of images to consider typically reaches the hundreds. Coupled with the additional double loop over atom indices, these result in a very large number of small tensor contractions, not ideal for distributed computing. Additionally, a trade-off has to be found between what can be stored in memory and what needs recomputing at each SCF step. The most important aspects of our implementation are discussed below.

As previously discussed, the efficient storage of two- and three-center ERI tensors is limited to cases where they depend on a single periodic image index. Therefore, any tensor contractions resulting in a two-index quantity cannot be pre-computed, even if the operation needs to be repeated at each SCF step. Only the following two- and three-center tensors can be pre-contracted and stored in memory without extra cost,

$$I_{\mu_i^0 \sigma^a R_i^0} = (\mu_i^0 \sigma^a | P_i^0) (P_i^0 | R_i^0)^{-1}. \quad (12)$$

The computational complexity associated with the triple loop on periodic image indices can be significantly simplified by selecting a judicious order of operations. By initially looping over images \mathbf{a} and \mathbf{c} and contracting the density matrix with the three-center integrals, we build a single-index three-center tensor with $\mathbf{d} = \mathbf{a} + \mathbf{c}$,

$$T_{\mu_i^0 \lambda^d R_i^0} = \sum_{\mathbf{a}, \mathbf{c}} P_{\sigma, \lambda}^{\mathbf{c}} I_{\mu_i^0 \sigma^a R_i^0}, \quad (13)$$

where different \mathbf{a} and \mathbf{c} pairs can sum up to the same \mathbf{d} . Finally, a loop over cell vector \mathbf{b} and atom pairs i, j is executed,

$$K_{\mu_i, \nu_j}^{\mathbf{b}} = \sum_{\mathbf{d}} T_{\mu_i^0 \lambda^d R_i^0} (R_i^0 | S_j^{\mathbf{b}}) I_{\nu_j^{\mathbf{b}} \lambda^d S_j^{\mathbf{b}}}, \quad (14)$$

where atomic blocks are computed one at a time. This leads to a computational complexity of $\mathcal{O}(N_{\text{img}}^2)$ because of the two subsequent double loops on periodic cells, where N_{img} is the number of periodic images \mathbf{b} , for which $(R_i^0 | S_j^{\mathbf{b}})$ is non-zero. With the loop on atom pairs i and j and the fact that only σ increases with system size in tensor $(\mu_i^0 \sigma^a | P_i^0)$, the overall complexity is $\mathcal{O}(N_{\text{atom}}^3 N_{\text{img}}^2)$. Note that as the simulation cell gets larger, the number of periodic images to consider decreases, leading to a non-trivial scaling of the method with the system size.

Equation (13) can be efficiently handled by batched contractions. Large matrices/tensors can be built by stacking $P_{\sigma,\lambda}^c$ and $I_{\mu_i^0 \sigma^a R_i^0}$ in a proper order such that various \mathbf{a} and \mathbf{c} pairs sum up to the same \mathbf{d} . While stacking along \mathbf{d} can also take place in Eq. (14), the tensors remain much smaller, leading to a hard limit on strong scaling. The triple loop over \mathbf{b} , i , and j becomes, in most cases, the bottleneck of the calculation. To efficiently deal with it, we distribute the \mathbf{b} , i , and j triplets over multiple MPI subgroups working independently. Because these subgroups each have fewer resources, they perform small matrix/tensor contractions at higher a parallel efficiency. This comes with some communication and memory overhead, as the three-center tensors $I_{\nu_j \lambda^d S_j^b}$ and $T_{\mu_i^0 \lambda^d R_i^0}$ must be replicated on each subgroup. The former can be replicated once and for all at the first SCF step, while the latter has to be broadcast at each step. Note that i, j atomic pairs are calculated on different subgroups according to index i such that the $T_{\mu_i^0 \lambda^d R_i^0}$ tensor is only partially copied.

Finally, we exploit the linearity of the exact exchange matrix with respect to the density matrix. At the n th SCF step, we calculate $K_{\mu_i, \nu_j}^b[\mathbf{P}^n] = K_{\mu_i, \nu_j}^b[\mathbf{P}^{(n-1)}] + K_{\mu_i, \nu_j}^b[\Delta\mathbf{P}]$. As the SCF converges, $\Delta\mathbf{P}$ becomes smaller, leading to sparser tensors contractions and faster SCF iterations. To maintain a low overall number of SCF steps, we implemented a \mathbf{k} -point-specific DIIS solver.⁷⁵ We also leverage the $K_{\mu_i, \nu_j}^b = K_{\nu_j, \mu_i}^{-b}$ symmetry to reduce the computational effort.

Depending on the system and the method used to generate \mathbf{k} -point meshes, symmetries can be exploited to reduce the effective number of \mathbf{k} -points. This can be useful to lower the time spent in diagonalization [Eq. (1)]. Because these operations are usually not a bottleneck in the real-space approach developed in this work, only the straightforward time-reversal symmetry ($\mathbf{k} = -\mathbf{k}$) is exploited.

In our implementation, we use the Libint library⁷⁶ to evaluate all two- and three-center ERIs. Matrix and tensor contractions

are efficiently managed by the GPU accelerated block-sparse DBM and DBT libraries.⁷⁰ The default filtering threshold for sparsity is set to 10^{-9} . By default, the RI basis is generated on the fly according to the scheme of Stoychev and co-workers,⁷⁷ but user provided RI basis sets are also possible. Nuclear gradients and stress tensors were also implemented, included within the ADMM approximation for the ADMM2, ADMMp, ADMMq, and ADMMS flavors.

III. RESULTS

Many calculations and benchmarks were performed with the aim of showcasing the method's accuracy and computational performance. First, a discussion concerning the choice of the RI metric is conducted. A study on the TC truncation radius to be used in RI-HFXk calculations and on the application of the ADMM approximation follows. Finally, various scaling tests are reported. Unless specified otherwise, we systematically use Monkhorst-Pack⁷⁸ \mathbf{k} -point meshes in this work.

We measure accuracy in terms of relative energy errors, i.e., the error made on the energy difference between two systems of the same composition, but with different atomic structures. For a given accuracy test, we subtract the energy of the pristine crystal structure from that of a system with a broken symmetry, which is achieved by randomly displacing atoms by a maximum of ± 0.1 Å. The same structures are used for all data points of the test. While absolute energies may vary, the focus on relative energies ensures that consistent comparisons between different calculations at the same level of theory can be done. We systematically report relative energy errors in terms of $\mu\text{Ha}/\text{atom}$, referring to the number of atoms in the calculation unit cell (independent of the \mathbf{k} -point mesh).

A. Choice of the RI metric

The selection of the RI metric, denoted by the operator \lfloor in Eq. (9), is an input parameter to any RI-HFX and low-scaling

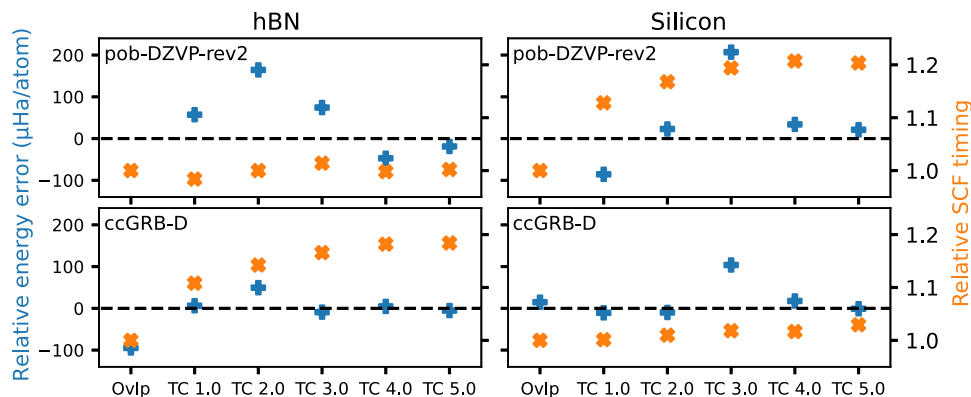


FIG. 2. Relative HF energy errors and SCF timings measured as a function of the RI metric, for 2D hBN and bulk silicon, with two atoms per unit cell each. Ovlp stands for the overlap metric, while TC R_C stands for the truncated Coulomb metric with a cutoff radius of R_C Å. The HFX potential used for all calculations is the truncated Coulomb with a cutoff radius of 5 Å. The reference relative energy is calculated at the Γ -point with a supercell of $7 \times 7 \times 1$ and $5 \times 5 \times 5$ for hBN and silicon, respectively. This corresponds exactly to the sampling from the \mathbf{k} -point meshes used for RI-HFXk. For each system and basis set combination, the reference timing of 1.0 was set for the calculation with the overlap metric. Note that the errors for the overlap metric when using the pob-DZVP-rev2 basis set are off the charts at 1090 and 708 $\mu\text{Ha}/\text{atom}$ for hBN and silicon, respectively.

post-HF calculation within CP2K. We are interested in making the optimal choice, with the assumption that a short range RI metric is more efficient due to sparsity, while longer ranged metrics improve accuracy. Figure 2 illustrates the relative errors and SCF timings observed when running RI-HFXk calculations on two-dimensional hBN and bulk silicon, with RI metrics of increasing range. Two basis sets are used: the solid-state specific all-electron pob-DZVP-rev2⁷⁹ and the more diffuse and general purpose ccGRB-D,²³ with GTH pseudopotentials.⁸⁰ The results reveal a general trend of error reduction with an increasing RI metric range, albeit with non-monotonic convergence and occasional significant jumps, as exemplified by TC 3.0 for silicon. Additionally, the SCF timings vary little, with a maximum increase of 20%. This is due to the almost constant levels of sparsity because of the local nature of the RI basis. This is in stark contrast with global RI schemes, where such long range RI metric choices lead to a significant increase in computational costs.⁷⁰ In conclusion, for optimal accuracy with a maximum error of 21 $\mu\text{Ha}/\text{atom}$ (for the specified setup of bulk silicon with pob-DZVP-rev2), we recommend employing an RI metric identical to the HFX potential. Furthermore, we note that the ccGRB-D basis set consistently outperforms pob-DZVP-rev2. A similar investigation on nuclear gradients corroborates these findings, and additional details can be found in the supplementary material. Unless explicitly stated, all calculations in this work use the same operator for the RI metric and the HFX potential.

B. Ideal TC truncation radius

All RI-HFXk calculations performed in this work use the truncated Coulomb operator as the HFX potential. As discussed in Sec. II B, this choice allows for a finite number of periodic images to be considered, while ensuring convergence of the exact exchange energy term. In CP2K, unlike the original work of Spencer and Alavi⁴⁴ and the GTO-based implementation of Irmeler and co-workers,⁴² we do not systematically increase the truncation radius of

TABLE I. Number of periodic images considered as a function of the HFX potential TC cutoff radius (\AA) for different systems, simulated with the pob-TZVP-rev2 basis set.

R_C	hBN	LiH	Silicon	Diamond
4.0	116	229	270	713
5.0	136	274	321	916
6.0	155	313	407	1117
7.0	165	413	530	1371
8.0	192	428	599	1650
9.0	213	540	703	2005
10.0	256	634	798	2352

the TC operator with the extent of the BvK supercell. Instead, a fixed value is chosen as an input parameter, large enough such that relative energies are converged. The general requirement that R_C must be smaller than the radius of the largest sphere fitting in the BvK supercell still holds. Figure 3 shows the relative energy errors and timings measured on 2D hBN, bulk LiH, bulk silicon, and diamond, as a function of the HFX potential truncation radius. The errors exhibit a clear monotonic decrease, with $R_C = 6 \text{ \AA}$ being converged enough for all systems considered. The maximum relative error for this cutoff is 43 $\mu\text{Ha}/\text{atom}$ for silicon using the pob-TZVP-rev2 basis set. Note that increasing the truncation radius leads to significant additional costs for dense solids because of the larger number of periodic images involved (Table I). In the case of diamond, switching from 6 to 10 \AA induces a 2.5 \times increase in cost. Section III B summarizes the number of periodic images considered as a function of the HFX potential truncation radius for the considered systems. We stress that R_C is a parameter of the HFXk method, and its value should be properly tested for convergence. In particular, small gaps and metallic systems might require larger truncation radii.

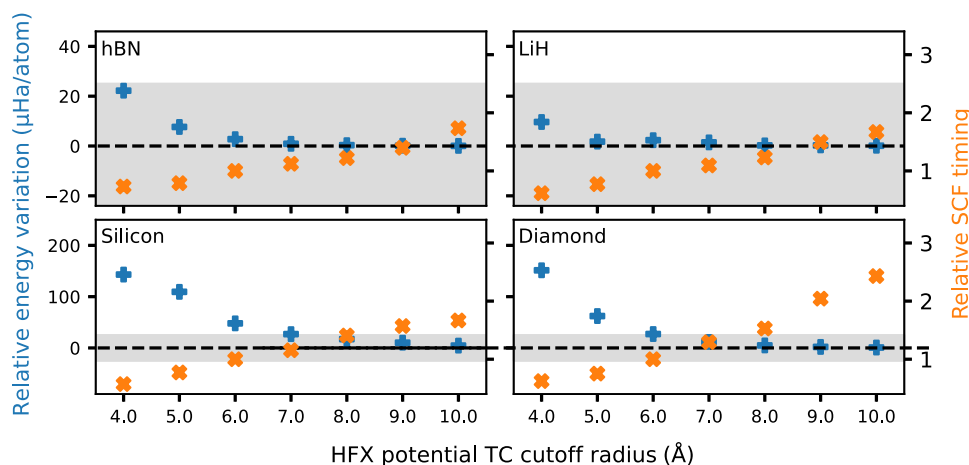


FIG. 3. Relative HF energy errors and SCF timings measured as a function of the HFX potential TC truncation radius. The reference relative energy was taken to be that calculated with a 12 \AA cutoff radius. The reference timing was set to the $R_C = 6 \text{ \AA}$ calculation as it is considered converged. The basis set employed for these tests was pob-TZVP-rev2. The number of atoms per unit cell is 2, 8, 2, and 2, and the k-point mesh is $15 \times 15 \times 1$, $12 \times 12 \times 12$, $12 \times 12 \times 12$, and $12 \times 12 \times 12$ for hBN, LiH, silicon, and diamond, respectively. Note the different y-axis error limits for the upper and lower sub-figures. The shaded area corresponds to an error of $\pm 25 \mu\text{Ha}/\text{atom}$.

C. Auxiliary density matrix method

The results presented in this section demonstrate the performance of the ADMM approximation in combination with the RI-HFXk method. Table II outlines the relative errors and speedups achieved with various combinations of basis sets for 2D hBN and bulk silicon. A consistent trend is observed, in which the most substantial speedups, reaching up to 700×, are associated with larger errors, typically an order of magnitude higher than the atom-specific RI-induced errors. This observation aligns with the findings of Rebolini and co-workers in molecular calculations.⁸¹ Note that the cFIT3⁵⁴ auxiliary basis for bulk silicon leads to particularly high errors. In the spirit of ADMM, high-quality triple-zeta valence basis sets (ccGRB-T and TZVP-MOLOPT-HYB-GTH⁸²) were chosen as the primary basis for this test. Given the strong dependence of the HFX cost on basis set size and diffuseness, significant benefits are expected from ADMM in this context. Furthermore, ADMM facilitates **k**-point HFX calculations with diffuse basis sets, provided that the selected auxiliary basis is less diffuse. The related accuracy and speedup measurements are not reported here because the reference RI-HFXk calculations become computationally intractable. Finally, ADMM can also be used with all-electron basis sets, as illustrated with 6-31G*^{83–85} in the table. There are not many ADMM optimized all-electron basis sets available, except for the ones from Kumar and co-workers,⁸⁶ which are too diffuse for efficient solid-state calculations. Instead, we show that the CRYSTAL¹⁶ pob family of basis sets can be efficiently used as auxiliary basis sets because of their limited diffuseness. Note that while 6-31G* is of double zeta quality, speedups and low errors are still achieved with the pob-TZVP-rev2 basis for silicon. The pob family can only be used as auxiliary basis sets for all-electron calculations. The results presented in Table II provide insights into selected few basis sets and systems, but the general trends are assumed to be universal. All calculations were done in the ADMMS flavor.

Additional tests were conducted in order to confirm the accuracy of RI-HFXk nuclear gradients and stress tensors, with and

TABLE II. ADMM HF relative energy errors (μ Ha/atom) and speedups for 2D hBN and bulk silicon. Both systems are simulated with a two atoms unit cell and a **k**-point mesh. Note that TZVP-HYB is a short notation for the TZVP-MOLOPT-HYB-GTH valence basis set.

Primary/aux basis	hBN ($13 \times 13 \times 1$)		Silicon ($8 \times 8 \times 8$)	
	Error	Speedup	Error	Speedup
ccGRB-T/admm-tzp	−200.1	32	−24.4	62
ccGRB-T/admm-tz2p	−128.1	8	7.9	12
ccGRB-T/cFIT3	−121.1	68	454.9	233
ccGRB-T/pFIT3	−81.8	16	−7.9	39
TZVP-HYB/admm-tzp	−172.5	18	−16.9	182
TZVP-HYB/admm-tz2p	−122.9	4	12.1	36
TZVP-HYB/cFIT3	−75.6	37	424.2	703
TZVP-HYB/pFIT3	−55.1	9	−49.2	115
6-31G*/3-21G	−68.1	3	484.7	9
6-31G*/pob-DZVP-rev2	−33.6	4	−60.7	8
6-31G*/pob-TZVP-rev2	−9.8	1	9.1	7

TABLE III. Comparison of cell parameters (in Å) obtained for various crystals as a result of a cell optimization procedure. Γ -point and **k**-point calculations were conducted with an exact supercell to **k**-point mesh equivalence. The pob-TZVP-rev2 basis set was used, with pob-DZVP-rev2 as an auxiliary basis for ADMM. All calculations were done with the PBE0⁸⁷ hybrid functional. Note that LiH uses an eight atoms unit cell but all others have two.

	Graphene $11 \times 11 \times 1$	hBN $11 \times 11 \times 1$	Silicon $5 \times 5 \times 5$	LiH $5 \times 5 \times 5$
Γ -4center	2.449	2.496	3.867	4.038
RI-HFXk	2.450	2.496	3.862	4.037
ADMM- Γ -4 center	2.450	2.497	3.868	4.037
ADMM-RI-HFXk	2.450	2.497	3.863	4.037

without ADMM. Cell optimization calculations were conducted for multiple crystals, namely, graphene, hBN, silicon, and LiH, with Γ -point supercells and equivalent **k**-point meshes. Results are presented in Table III. The maximum difference in the converged cell parameter is 0.005 Å (0.1%) between the four-center Γ -point supercell approach and RI-HFXk, while it is only 0.001 Å between ADMM and non-ADMM runs. Two important lessons can be taken from this experiment: The new RI-HFXk method is consistent with the well-established Γ -point implementation of CP2K and ADMM nuclear gradients and stress tensors are accurate.

D. Performance

1. RI-HFXk vs Γ 4-center

The main interest of **k**-point approaches is to reach the thermodynamic limit more efficiently than with Γ -point supercell methods. It is shown in Figs. 4 and 5, where the cost of equivalent supercells and **k**-point meshes calculations are reported for 2D hBN and bulk silicon. Regardless of the system and the basis set used, RI-HFXk calculations are much more efficient. Moreover, their cost remains constant with respect to the **k**-point mesh, whereas Γ -point computation scales linearly with a supercell size at best. For both methods, this behavior is only possible because the cut-off radius of the TC HFX potential is kept constant. The cost of Γ -point ccGRB-D calculations with silicon increases rapidly, illustrating the difficulty in running hybrid DFT simulations of solids with diffuse basis sets. In contrast, the RI-HFXk calculations with the same setting are very well behaved, notably because of the local nature of our atom-specific RI scheme. We point to the fact that the y axis of both figures have very different bounds, demonstrating how much more costly the simulations of dense bulk systems are. These results show the Γ -point four-center method in its most favorable light, where enough resources are always allocated to precompute and store all ERIs at the first SCF step. Note that a series of Γ -point calculations were conducted with different amounts of resources, leading to slight variations in parallel efficiency, which are clearly visible in the ccGRB-D curve of Fig. 4. For both figures discussed here, the cost of RI-HFXk is constant with respect to the **k**-point mesh. This happens because the cost of building the real-space exact exchange matrices dominates in this regime. However, eventually performing a diagonalization at each **k**-point becomes the bottleneck of the calculation, leading to linear scaling with the number of **k**-points. For bulk silicon, this regime transition starts

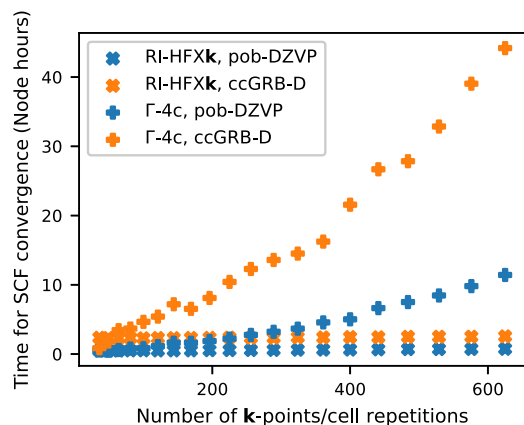


FIG. 4. Cost comparison of the Γ four-center and RI-HFXk methods for increasing supercell size and \mathbf{k} -point density, respectively. The system is 2D hBN with two atoms per unit cell, simulated with the truncated Coulomb operator ($R_C = 6 \text{ \AA}$) as the HFX potential. Calculations were run on Cray XC50 GPU nodes on the Piz Daint supercomputer.

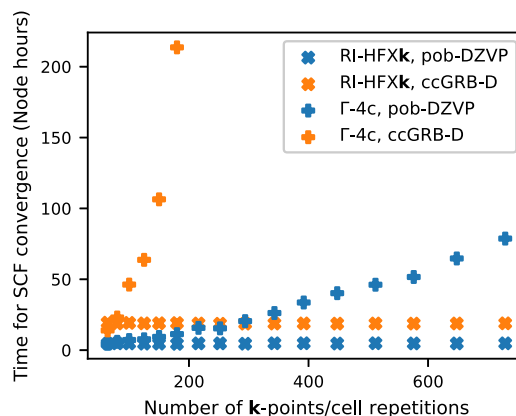


FIG. 5. Cost comparison of the Γ four-center and RI-HFXk methods for increasing supercell size and \mathbf{k} -point density, respectively. The system is bulk silicon with two atoms per unit cell, simulated with the truncated Coulomb operator ($R_C = 5 \text{ \AA}$) as the HFX potential. Calculations were run on Cray XC50 GPU nodes on the Piz Daint supercomputer.

at around 800–1000 \mathbf{k} -points, as illustrated in the supplementary material. Additional tests showcasing the convergence of the total energies as a function of the number of \mathbf{k} -points, as well as acceleration of individual SCF steps due to increasing $\Delta\mathbf{P}$ sparsity, are also available.

2. Strong scaling

As discussed in Sec. II E, the most expensive step of the algorithm is the triple loop over cell vector \mathbf{b} and i, j atom pairs. Because this represents a large amount of small matrix/tensor contractions, we took the decision to split the work among independent workers, each calculating a fraction of the \mathbf{b} and i, j atomic blocks. This comes at the cost of some data replication and communication overhead. There are two ways that this bottleneck can be accelerated: by increasing the number of workers (MPI subgroups) and by increasing the resources allocated to each worker. Figure 6 shows the strong

scaling performance of the RI-HFXk method for both approaches independently. The measured speedups and parallel efficiencies for the pob-TZVP-rev2 basis are reported in Table IV. The strong scaling performance is high across the board, especially when increasing the number of independent MPI subgroups. We expect better efficiencies with respect to the worker size with larger simulation cells and/or larger basis sets because the size of individual tensors would increase. Note that calculations using the ccGRB-D basis set are significantly more expensive than those using pob-TZVP-rev2 because of diffuseness rather than basis size. In both the subfigures, the first data point of the ccGRB-D series is missing. This is due to memory limitations on a single node as the ccGRB-D basis requires more than twice the number of periodic cells to be considered (697 vs 321). While running with a single MPI subgroup on one node is not possible, two workers on two nodes is possible. This happens because not all data are copied to each subgroup. Note that while the $\mathbf{b}, i,$

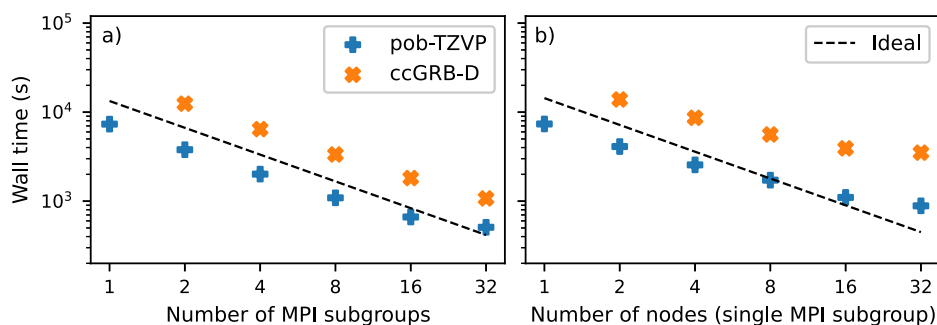


FIG. 6. Strong scaling performance for five SCF steps of a bulk silicon two-atom cell ($8 \times 8 \times 8$ \mathbf{k} -point mesh). In subfigure (a), the number of independent MPI subgroups is increased, while their size (1 Cray XC50 GPU node) is kept constant. In subfigure (b), calculations are run with a single MPI subgroup, which gradually increases in size. The HFX potential is the truncated Coulomb operator with a cutoff radius of $R_C = 5 \text{ \AA}$. Note that the first ccGRB-D point is missing on both plots due to memory limitations on a single node.

TABLE IV. Strong scaling speedups and parallel efficiencies corresponding to the pob-TZVP-rev2 curves shown in Fig. 6, for the two atoms primitive cell of bulk silicon.

	Number of workers		Size of workers		
	Speedup	Efficiency	Speedup	Efficiency	
1	1.00	1.00	1	1.00	1.00
2	1.93	0.97	2	1.79	0.90
4	3.65	0.91	4	2.88	0.72
8	6.73	0.84	8	4.28	0.53
16	11.05	0.69	16	6.69	0.42
32	14.43	0.45	32	8.32	0.26

and j triple loop is considerably accelerated with additional MPI subgroups, the overall increase in resources also speeds up the rest of the calculation, notably the evaluation of the ERIs and the contraction with the density matrices in Eq. (13). Additional tests covering various systems and basis sets, as well as measures of GPU acceleration, are available in the supplementary material. GPU runs were measured to be 1.3–2.4 times faster than the CPU only calculations on the same computer. The GPUs used in this work date back to 2016, and greater acceleration is expected on more modern hardware. A previous study using the state-of-the-art LUMI-G supercomputer⁷⁰ suggests that GPU acceleration could double.

3. Weak scaling

A given calculation can be accelerated by allocating more resources to it, both by increasing the number and the size of the MPI subgroups working on the \mathbf{b} , i , and j loop of Eq. (14). On the other hand, the computational effort grows together with the simulation unit cell size, according to the $\mathcal{O}(N_{\text{atom}}^3 N_{\text{img}}^2)$ analysis performed in Sec. II E. In Fig. 7, we demonstrate a good weak scaling performance by keeping constant time to solution for 2D hBN, bulk silicon, and diamond systems of increasing sizes. The CPU time necessary to complete these calculations grows significantly following a linear trend on the log–log plot. Polynomial fitting yields the effective scaling of the method with the system

size, measured at $\mathcal{O}(N^{1.5})$, $\mathcal{O}(N^{1.7})$, and $\mathcal{O}(N^{1.6})$ for hBN, silicon, and diamond, respectively. While these tests were conducted with the overlap RI metric to keep the resource consumption low, the observed trends are not expected to be affected by that choice. As the cell size of the test systems increased, so did both the number of MPI subgroups and their sizes. The largest calculation—the 32 atoms diamond cell—was run on 192 nodes with 16 MPI subgroups. Note the significant augmentation of required resources when going from 2D hBN to bulk silicon and, finally, to diamond. This again highlights the difficulty of running HFX calculation of dense solids. Additional analysis, available in the supplementary material, shows that the method scales slightly more favorably than the predicted $\mathcal{O}(N_{\text{atom}}^3 N_{\text{img}}^2)$. We attribute this to sparsity in the tensor/matrix contractions.

IV. CONCLUSION

In this work, we have reported the development of the RI-HFXk method and its implementation into the CP2K software package. This novel approach enables accurate and efficient periodic Hartree–Fock exchange calculations with \mathbf{k} -point sampling. Our results are in excellent agreement with equivalent Γ -point supercell calculations based on a well-established implementation, for relative energies and nuclear gradients. To enhance the efficiency, both methods leverage the truncated Coulomb operator with a fixed, converged value.

We adopt a real-space approach that ensures a computational cost scaling, at worst, linearly with the number of \mathbf{k} -points when diagonalization dominates. The primary computational bottleneck occurs during the construction of real-space exact exchange matrices. In this context, our local atom-specific RI scheme demonstrates a notable advantage over more conventional two-electron four-center approaches. Indeed, its modest memory requirements allow for the storage and replication of pre-calculated quantities, leading to a very efficient parallel implementation.

Due to the employment of RI, the computational effort is largely dominated by linear algebra. This is efficiently handled by the use of the DBM and DBT libraries,⁷⁰ designed around block sparsity. Besides the inherent sparsity coming from atom-centered

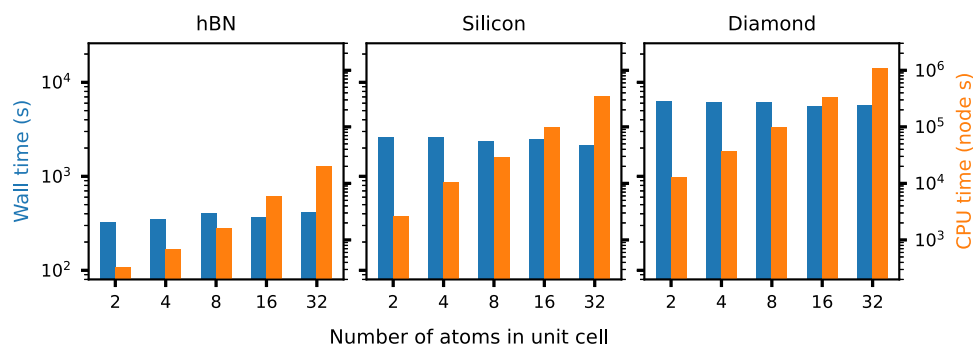


FIG. 7. Weak scaling performance of the RI-HFXk method with unit cells of increasing sizes, for various crystalline systems. The blue bars correspond to the wall time in seconds necessary for five SCF steps, while the orange bars correspond to the CPU time in node seconds (Cray XC50 GPU nodes on the Piz Daint supercomputer). The HFX potential is the truncated Coulomb operator with cutoff radii of 6.0, 5.0, and 5.0 Å for hBN, silicon, and diamond, respectively. Calculations were performed with the overlap RI metric and the pob-DZVP-rev2 basis set.

Gaussian basis functions, we leverage the linearity of the exact exchange with respect to the density matrix, leading to decreasing costs as the SCF converges. Overall, our implementation exhibits a good strong and weak scaling performance, as well as some level of GPU acceleration. This makes the method highly suitable for high-performance computing, offering rapid time to solution. The effective scaling of the method with system size was measured at $\mathcal{O}(N^{1.5-1.7})$.

The auxiliary density matrix method can be seamlessly combined with RI-HFXk. Significant speedups are achieved in calculations using large basis sets, with reasonable accuracy for relative energies. Assessments of cell optimization runs also demonstrate the high quality of ADMM-derived forces and stress tensors. Additionally, the use of ADMM makes calculations with highly diffuse basis sets possible, provided the suitable selection of a less diffuse auxiliary basis. This is of significant importance as real-space approaches to HFX encounter substantial computational challenges with basis set diffuseness.

This fully open-source implementation opens the door to a new range of calculations with the CP2K software package. High quality hybrid DFT calculations of solids near the thermodynamic limit are now possible at a limited computational cost. Furthermore, the RI-HFXk method can be used as a starting point for higher level methods, such as periodic GW with \mathbf{k} -point sampling,⁸⁸ and holds potential for future periodic post-HF developments.

SUPPLEMENTARY MATERIAL

See the supplementary material for details on the RI bump function used in this work. Additional figures covering nuclear gradient accuracy, ϵ_{PGF} convergence tests, scaling with respect to \mathbf{k} -point number, total energy convergence as a function of \mathbf{k} -point number, an illustrative band structure calculation, and strong scaling/GPU acceleration are also available. An illustrative CP2K input file is included as well.

ACKNOWLEDGMENTS

This work was funded by the Swiss Platform for Advanced Scientific Computing (PASC) as part of the 2021–2024 “Ab Initio Molecular Dynamics at the Exa-Scale” project. Computing time on Piz Daint was provided by the CSCS Swiss National Supercomputing Centre under Account Nos. d110 and c18. The authors would like to thank Tiziano Müller for the fruitful discussions on ADMM.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Augustin Bussy: Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Jürg Hutter:** Funding acquisition (equal);

Project administration (equal); Resources (equal); Software (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study, namely, CP2K input and output files, are openly available in the Materials Cloud at <https://doi.org/10.24435/materialscloud:rg-dn>.⁸⁹

REFERENCES

- 1 J. Harris, “Adiabatic-connection approach to Kohn–Sham theory,” *Phys. Rev. A* **29**, 1648 (1984).
- 2 A. D. Becke, “A new mixing of Hartree–Fock and local density-functional theories,” *J. Chem. Phys.* **98**, 1372–1377 (1993).
- 3 J. P. Perdew, M. Ernzerhof, and K. Burke, “Rationale for mixing exact exchange with density functional approximations,” *J. Chem. Phys.* **105**, 9982–9985 (1996).
- 4 L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, “A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions,” *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).
- 5 N. Mardirossian and M. Head-Gordon, “Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals,” *Mol. Phys.* **115**, 2315–2372 (2017).
- 6 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, “Density functional theory is straying from the path toward the exact functional,” *Science* **355**, 49–52 (2017).
- 7 S. Dohm, A. Hansen, M. Steinmetz, S. Grimme, and M. P. Checinski, “Comprehensive thermochemical benchmark set of realistic closed-shell metal organic reactions,” *J. Chem. Theory Comput.* **14**, 2596–2608 (2018).
- 8 B. Chan, P. M. Gill, and M. Kimura, “Assessment of DFT methods for transition metals with the TMC151 compilation of data sets and comparison with accuracies for main-group chemistry,” *J. Chem. Theory Comput.* **15**, 3610–3622 (2019).
- 9 J. Muscat, A. Wander, and N. Harrison, “On the prediction of band gaps from hybrid functional theory,” *Chem. Phys. Lett.* **342**, 397–401 (2001).
- 10 D. Munoz, N. Harrison, and F. Illas, “Electronic and magnetic structure of LaMnO₃ from hybrid periodic density-functional theory,” *Phys. Rev. B* **69**, 085115 (2004).
- 11 E. R. Batista, J. Heyd, R. G. Hennig, B. P. Uberuaga, R. L. Martin, G. E. Scuseria, C. Umrigar, and J. W. Wilkins, “Comparison of screened hybrid density functional theory to diffusion Monte Carlo in calculations of total energies of silicon phases and defects,” *Phys. Rev. B* **74**, 121102 (2006).
- 12 Y. Wang, S. de Gironcoli, N. S. Hush, and J. R. Reimers, “Successful a priori modeling of CO adsorption on Pt(111) using periodic hybrid density functional theory,” *J. Am. Chem. Soc.* **129**, 10402–10407 (2007).
- 13 M. Betzinger, C. Friedrich, and S. Blügel, “Hybrid functionals within the all-electron FLAPW method: Implementation and applications of PBE0,” *Phys. Rev. B* **81**, 195117 (2010).
- 14 H.-P. Komsa and A. Pasquarello, “Assessing the accuracy of hybrid functionals in the determination of defect levels: Application to the As antisite in GaAs,” *Phys. Rev. B* **84**, 075207 (2011).
- 15 T. M. Henderson, J. Paier, and G. E. Scuseria, “Accurate treatment of solids with the HSE screened hybrid,” *Phys. Status Solidi B* **248**, 767–774 (2011).
- 16 R. Dovesi, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. Bush, P. D’arco, Y. Noël, M. Rérat, P. Carbonnière *et al.*, “The CRYSTAL code, 1976–2020 and beyond, a long story,” *J. Chem. Phys.* **152**, 204111 (2020).
- 17 S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig *et al.*, “TURBOMOLE: Modular program suite for *ab initio* quantum-chemical and condensed-matter simulations,” *J. Chem. Phys.* **152**, 184107 (2020).
- 18 Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng *et al.*, “Advances in molecular quantum chemistry contained in the Q-Chem 4 program package,” *Mol. Phys.* **113**, 184–215 (2015).

- ¹⁹V. Blum, M. Rossi, S. Kokott, and M. Scheffler, "The FHI-aims Code: All-electron, ab initio materials simulations towards the exascale," *arXiv:2208.12335* (2022).
- ²⁰Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui *et al.*, "Recent developments in the PySCF program package," *J. Chem. Phys.* **153**, 024109 (2020).
- ²¹M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Morokuma, A. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
- ²²C. Peng, C. A. Lewis, X. Wang, M. C. Clement, K. Pierce, V. Rishi, F. Pavošević, S. Slattery, J. Zhang, N. Teke *et al.*, "Massively parallel quantum chemistry: A high-performance research platform for electronic structure," *J. Chem. Phys.* **153**, 044120 (2020).
- ²³T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffrmann *et al.*, "CP2K: An electronic structure and molecular dynamics software package—Quickstep: Efficient and accurate electronic structure calculations," *J. Chem. Phys.* **152**, 194103 (2020).
- ²⁴M. Shishkin and G. Kresse, "Self-consistent GW calculations for semiconductors and insulators," *Phys. Rev. B* **75**, 235102 (2007).
- ²⁵F. Hüser, T. Olsen, and K. S. Thygesen, "Quasiparticle GW calculations for solids, molecules, and two-dimensional materials," *Phys. Rev. B* **87**, 235132 (2013).
- ²⁶J. Klimeš, M. Kaltak, and G. Kresse, "Predictive GW calculations using plane waves and pseudopotentials," *Phys. Rev. B* **90**, 075125 (2014).
- ²⁷X. Ren, P. Rinke, C. Joas, and M. Scheffler, "Random-phase approximation and its applications in computational chemistry and materials science," *J. Mater. Sci.* **47**, 7447–7471 (2012).
- ²⁸M. Kalkak, J. Klimes, and G. Kresse, "Low scaling algorithms for the random phase approximation: Imaginary time and Laplace transformations," *J. Chem. Theory Comput.* **10**, 2498–2507 (2014).
- ²⁹J. Wilhelm, P. Seewald, M. Del Ben, and J. Hutter, "Large-scale cubic-scaling random phase approximation correlation energy calculations using a Gaussian basis," *J. Chem. Theory Comput.* **12**, 5851–5859 (2016).
- ³⁰C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usyvat, "Periodic local MP2 method for the study of electronic correlation in crystals: Theory and preliminary applications," *J. Comput. Chem.* **29**, 2113–2124 (2008).
- ³¹A. Hermann and P. Schwerdtfeger, "Ground-state properties of crystalline ice from periodic Hartree–Fock calculations and a coupled-cluster-based many-body decomposition of the correlation energy," *Phys. Rev. Lett.* **101**, 183005 (2008).
- ³²M. Del Ben, J. Hutter, and J. VandeVondele, "Second-order Møller–Plesset perturbation theory in the condensed phase: An efficient and massively parallel Gaussian and plane waves approach," *J. Chem. Theory Comput.* **8**, 4177–4188 (2012).
- ³³X. Wang and T. C. Berkelbach, "Absorption spectra of solids from periodic equation-of-motion coupled-cluster theory," *J. Chem. Theory Comput.* **17**, 6387–6394 (2021).
- ³⁴C. Pisani and R. Dovesi, "Exact-exchange Hartree–Fock calculations for periodic systems. I. Illustration of the method," *Int. J. Quantum Chem.* **17**, 501–516 (1980).
- ³⁵R. Dovesi, C. Pisani, and C. Roetti, "Exact-exchange Hartree–Fock calculations for periodic systems. II. Results for graphite and hexagonal boron nitride," *Int. J. Quantum Chem.* **17**, 517–529 (1980).
- ³⁶R. Dovesi, C. Pisani, F. Ricca, and C. Roetti, "Exact-exchange Hartree–Fock calculations for periodic systems. III. Ground-state properties of diamond," *Phys. Rev. B* **22**, 5936 (1980).
- ³⁷R. Dovesi, C. Pisani, C. Roetti, and P. Dellarole, "Exact-exchange Hartree–Fock calculations for periodic systems. IV. Ground-state properties of cubic boron nitride," *Phys. Rev. B* **24**, 4170 (1981).
- ³⁸R. Dovesi, M. Causa, and G. Angonoa, "Exact-exchange Hartree–Fock calculations for periodic systems. V. Ground-state properties of silicon," *Phys. Rev. B* **24**, 4177 (1981).
- ³⁹F. Gygi and A. Baldereschi, "Self-consistent Hartree–Fock and screened-exchange calculations in solids: Application to silicon," *Phys. Rev. B* **34**, 4405 (1986).
- ⁴⁰C. Tymczak, V. T. Weber, E. Schwegler, and M. Challacombe, "Linear scaling computation of the Fock matrix. VIII. Periodic boundaries for exact exchange at the Γ point," *J. Chem. Phys.* **122**, 124105 (2005).
- ⁴¹M. Guidon, J. Hutter, and J. VandeVondele, "Robust periodic Hartree–Fock exchange for large-scale simulations using Gaussian basis sets," *J. Chem. Theory Comput.* **5**, 3010–3021 (2009).
- ⁴²A. Irmrl, A. M. Burow, and F. Pauly, "Robust periodic Fock exchange with atom-centered Gaussian basis sets," *J. Chem. Theory Comput.* **14**, 4567–4580 (2018).
- ⁴³G. J. Martyna and M. E. Tuckerman, "A reciprocal space based method for treating long range interactions in *ab initio* and force-field-based calculations in clusters," *J. Chem. Phys.* **110**, 2810–2821 (1999).
- ⁴⁴J. Spencer and A. Alavi, "Efficient calculation of the exact exchange energy in periodic systems using a truncated Coulomb potential," *Phys. Rev. B* **77**, 193110 (2008).
- ⁴⁵R. Sundararaman and T. Arias, "Regularization of the Coulomb singularity in exact exchange by Wigner–Seitz truncated interactions: Towards chemical accuracy in nontrivial systems," *Phys. Rev. B* **87**, 165122 (2013).
- ⁴⁶B. I. Dunlap, J. Connolly, and J. Sabin, "On some approximations in applications of $X\alpha$ theory," *J. Chem. Phys.* **71**, 3396–3402 (1979).
- ⁴⁷O. Vahtras, J. Almlöf, and M. Feyereisen, "Integral approximations for LCAO-SCF calculations," *Chem. Phys. Lett.* **213**, 514–518 (1993).
- ⁴⁸S. V. Levchenko, X. Ren, J. Wierfink, R. Johanni, P. Rinke, V. Blum, and M. Scheffler, "Hybrid functionals for large periodic systems in an all-electron, numeric atom-centered basis framework," *Comput. Phys. Commun.* **192**, 60–69 (2015).
- ⁴⁹P. Lin, X. Ren, and L. He, "Accuracy of localized resolution of the identity in periodic hybrid functional calculations with numerical atomic orbitals," *J. Phys. Chem. Lett.* **11**, 3082–3088 (2020).
- ⁵⁰P. Lin, X. Ren, and L. He, "Efficient hybrid density functional calculations for large periodic systems using numerical atomic orbitals," *J. Chem. Theory Comput.* **17**, 222–239 (2020).
- ⁵¹X. Wang, C. A. Lewis, and E. F. Valeev, "Efficient evaluation of exact exchange for periodic systems via concentric atomic density fitting," *J. Chem. Phys.* **153**, 124116 (2020).
- ⁵²J. Lee, A. Rettig, X. Feng, E. Epifanovsky, and M. Head-Gordon, "Faster exact exchange for solids via occ-RI-K: Application to combinatorially optimized range-separated hybrid functionals for simple solids with pseudopotentials near the basis set limit," *J. Chem. Theory Comput.* **18**, 7336–7349 (2022).
- ⁵³A. Rettig, J. Lee, and M. Head-Gordon, "Even faster exact exchange for solids via tensor hypercontraction," *J. Chem. Theory Comput.* **19**, 5773–5784 (2023).
- ⁵⁴M. Guidon, J. Hutter, and J. VandeVondele, "Auxiliary density matrix methods for Hartree–Fock exchange calculations," *J. Chem. Theory Comput.* **6**, 2348–2364 (2010).
- ⁵⁵L. Lin, "Adaptively compressed exchange operator," *J. Chem. Theory Comput.* **12**, 2242–2249 (2016).
- ⁵⁶W. Hu, L. Lin, and C. Yang, "Interpolative separable density fitting decomposition for accelerating hybrid density functional calculations with applications to defects in silicon," *J. Chem. Theory Comput.* **13**, 5420–5431 (2017).
- ⁵⁷I. Carnimeo, S. Baroni, and P. Giannozzi, "Fast hybrid density-functional computations using plane-wave basis sets," *Electron. Struct.* **1**, 015009 (2019).
- ⁵⁸K. Wu, X. Qin, W. Hu, and J. Yang, "Low-rank approximations accelerated plane-wave hybrid functional calculations with k-point sampling," *J. Chem. Theory Comput.* **18**, 206–218 (2021).
- ⁵⁹X. Qin, W. Hu, and J. Yang, "Interpolative separable density fitting for accelerating two-electron integrals: A theoretical perspective," *J. Chem. Theory Comput.* **19**, 679–693 (2023).

- ⁶⁰Q. Sun, “Exact exchange with range-separated algorithm for thermodynamic limit of periodic Hartree–Fock theory,” *J. Chem. Phys.* **159**, 024108 (2023).
- ⁶¹J. Heyd, G. E. Scuseria, and M. Ernzerhof, “Hybrid functionals based on a screened Coulomb potential,” *J. Chem. Phys.* **118**, 8207–8215 (2003).
- ⁶²M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, “Bulk liquid water at ambient temperature and pressure from MP2 theory,” *J. Phys. Chem. Lett.* **4**, 3753–3759 (2013).
- ⁶³A.-M. El-Sayed, M. B. Watkins, V. V. Afanas’ev, and A. L. Shluger, “Nature of intrinsic and extrinsic electron trapping in SiO₂,” *Phys. Rev. B* **89**, 125201 (2014).
- ⁶⁴C. Sprefaco and J. VandeVondele, “The nature of excess electrons in anatase and rutile from hybrid DFT and RPA,” *Phys. Chem. Chem. Phys.* **16**, 26144–26152 (2014).
- ⁶⁵W. Karim, C. Sprefaco, A. Kleibert, J. Gobrecht, J. VandeVondele, Y. Ekinici, and J. A. van Bokhoven, “Catalyst support effects on hydrogen spillover,” *Nature* **541**, 68–71 (2017).
- ⁶⁶J. Gaberle and A. L. Shluger, “Structure and properties of intrinsic and extrinsic defects in black phosphorus,” *Nanoscale* **10**, 19536–19546 (2018).
- ⁶⁷K. Konstantinou, F. C. Mocanu, T.-H. Lee, and S. R. Elliott, “Revealing the intrinsic nature of the mid-gap defects in amorphous Ge₂Sb₂Te₅,” *Nat. Commun.* **10**, 3065 (2019).
- ⁶⁸O. A. Dicks, J. Cottom, A. L. Shluger, and V. V. Afanas’ev, “The origin of negative charging in amorphous Al₂O₃ films: The role of native defects,” *Nanotechnology* **30**, 205201 (2019).
- ⁶⁹N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Cengage Learning, 2022).
- ⁷⁰A. Bussy, O. Schütt, and J. Hutter, “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.* **158**, 164109 (2023).
- ⁷¹A. Sodt, J. E. Subotnik, and M. Head-Gordon, “Linear scaling density fitting,” *J. Chem. Phys.* **125**, 194109 (2006).
- ⁷²B. I. Dunlap, “Robust and variational fitting,” *Phys. Chem. Chem. Phys.* **2**, 2113–2116 (2000).
- ⁷³D. Usvyat, L. Maschio, and M. Schütz, “Periodic local MP2 method employing orbital specific virtuals,” *J. Chem. Phys.* **143**, 102805 (2015).
- ⁷⁴P. Merlot, R. Izsák, A. Borgoo, T. Kjærgaard, T. Helgaker, and S. Reine, “Charge-constrained auxiliary-density-matrix methods for the Hartree–Fock exchange contribution,” *J. Chem. Phys.* **141**, 094104 (2014).
- ⁷⁵L. Maschio, “Direct inversion of the iterative subspace (DIIS) convergence accelerator for crystalline solids employing Gaussian basis sets,” *Theor. Chem. Acc.* **137**, 60 (2018).
- ⁷⁶E. F. Valeev, “Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions,” <http://libint.valeev.net/>, version 2.8.0, 2022.
- ⁷⁷G. L. Stoychev, A. A. Auer, and F. Neese, “Automatic generation of auxiliary basis sets,” *J. Chem. Theory Comput.* **13**, 554–562 (2017).
- ⁷⁸H. J. Monkhorst and J. D. Pack, “Special points for Brillouin-zone integrations,” *Phys. Rev. B* **13**, 5188 (1976).
- ⁷⁹J. Laun, D. Vilela Oliveira, and T. Bredow, “Consistent Gaussian basis sets of double- and triple-zeta valence with polarization quality of the fifth period for solid-state calculations,” *J. Comput. Chem.* **39**, 1285–1290 (2018).
- ⁸⁰S. Goedecker, M. Teter, and J. Hutter, “Separable dual-space Gaussian pseudopotentials,” *Phys. Rev. B* **54**, 1703 (1996).
- ⁸¹E. Rebolini, R. Izsák, S. S. Reine, T. Helgaker, and T. B. Pedersen, “Comparison of three efficient approximate exact-exchange algorithms: The chain-of-spheres algorithm, pair-atomic resolution-of-the-identity method, and auxiliary density matrix method,” *J. Chem. Theory Comput.* **12**, 3514–3522 (2016).
- ⁸²J. VandeVondele and J. Hutter, “Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases,” *J. Chem. Phys.* **127**, 114105 (2007).
- ⁸³J. D. Dill and J. A. Pople, “Self-consistent molecular orbital methods. XV. Extended Gaussian-type basis sets for lithium, beryllium, and boron,” *J. Chem. Phys.* **62**, 2921–2923 (1975).
- ⁸⁴W. J. Hehre, R. Ditchfield, and J. A. Pople, “Self—Consistent molecular orbital methods. XII. Further extensions of Gaussian—Type basis sets for use in molecular orbital studies of organic molecules,” *J. Chem. Phys.* **56**, 2257–2261 (1972).
- ⁸⁵M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, “Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements,” *J. Chem. Phys.* **77**, 3654–3665 (1982).
- ⁸⁶C. Kumar, H. Fliegl, F. Jensen, A. M. Teale, S. Reine, and T. Kjærgaard, “Accelerating Kohn–Sham response theory using density fitting and the auxiliary-density-matrix method,” *Int. J. Quantum Chem.* **118**, e25639 (2018).
- ⁸⁷C. Adamo and V. Barone, “Toward reliable density functional methods without adjustable parameters: The PBE0 model,” *J. Chem. Phys.* **110**, 6158–6170 (1999).
- ⁸⁸M. Graml, K. Zollner, D. Hernangómez-Pérez, P. E. F. Junior, and J. Wilhelm, “Low-scaling GW algorithm applied to twisted transition-metal dichalcogenide heterobilayers,” [arXiv:2306.16066](https://arxiv.org/abs/2306.16066) (2023).
- ⁸⁹A. Bussy and J. Hutter (2023). “Efficient periodic resolution-of-the-identity Hartree–Fock exchange method with k-point sampling and Gaussian basis sets,” Materials Cloud Archive 2023.182. <https://doi.org/10.24435/materialscloud:rg-dn>