

Fast periodic Gaussian density fitting by range separation

Hong-Zhou Ye¹ and Timothy C. Berkelbach^{1,2, a)}

¹⁾Department of Chemistry, Columbia University, New York, New York 10027, USA

²⁾Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

We present an efficient implementation of periodic Gaussian density fitting (GDF) using the Coulomb metric. The three-center integrals are divided into two parts by range-separating the Coulomb kernel, with the short-range part evaluated in real space and the long-range part in reciprocal space. With a few algorithmic optimizations, we show that this new method – which we call range-separated GDF (RSGDF) – scales sublinearly to linearly with the number of k -points for small to medium-sized k -point meshes that are commonly used in periodic calculations with electron correlation. Numerical results on a few three-dimensional solids show about 10-fold speedups over the previously developed GDF with little precision loss. The error introduced by RSGDF is about $10^{-5} E_h$ in the converged Hartree-Fock energy with default auxiliary basis sets and can be systematically reduced by increasing the size of the auxiliary basis with little extra work.

Introduction. For one-electron basis sets in periodic electronic structure calculations, translational symmetry-adapted atom-centered Gaussian functions^{1–4} are an alternative to the historically prevalent plane waves.^{5–9} Using Gaussian basis functions provides a more compact representation of orbitals, allows natural access to all-electron calculations without pseudopotentials, and facilitates the adaptation of accurate quantum chemistry methods for solids.^{1,10–17} The downside of atom-centered orbitals is the introduction of four-index electron repulsion integrals (ERIs), with $O(N_k^3 n_{AO}^4)$ storage and $O(N_k^2 n_{AO}^4)$ CPU costs for Hartree-Fock (HF) calculations, where N_k is the number of k -points sampled in the Brillouin zone and n_{AO} is the number of atomic orbitals in the unit cell. Moreover, the direct real-space evaluation of ERIs requires an expensive triple lattice summation. The Gaussian and plane wave (GPW) method¹⁸ reduces the scaling of the storage to $O(N_k^2 n_{AO}^2 N_{PW})$ and the HF cost to $O(N_k^2 n_{AO}^2 N_{PW} \ln N_{PW})$ by evaluating the ERIs entirely in reciprocal space using an auxiliary PW basis of size N_{PW} . However, doing so necessitates a pseudopotential and hence precludes all-electron calculations. In addition, a large PW basis may be needed if the basis set contains relatively compact orbitals.

Another way to reduce the cost of manipulating the ERIs is with Gaussian density fitting^{19–21} (GDF). In GDF, the orbital pair densities used to evaluate the ERIs are expanded in a second, auxiliary Gaussian basis of size n_{aux} , from which the four-center ERIs can be approximated using two- and three-center integrals evaluated with some metric function.^{22–25} The number of the latter integrals scales as $O(N_k^2 n_{AO}^2 n_{aux})$, which is much lower than that of the ERIs if n_{aux} is not too big. For molecules, highly optimized auxiliary basis sets²⁶ with $n_{aux} \approx 3n_{AO}$ have made GDF a great success in both mean-field^{27–29} and correlated calculations.^{30–34} We note that the GPW treatment of ERIs can also be understood as a PW density fitting where $N_{PW} \gg n_{AO}$.

The application of GDF to periodic systems has been a relatively recent effort.^{10,35–44} The main challenge is the high computational cost of evaluating the three-center integrals in real

space if the Coulomb metric is used. There are two classes of periodic GDF schemes. The first class exploits locality to limit the auxiliary expansion based on the proximity to the target pair density.^{10,37–39,42,44,45} The locality could arise from the system itself,⁴⁴ an explicit use of a local metric other than the Coulomb operator,⁴² or the use of Poisson-type orbitals.^{10,37,38} The other class insists on a global, Coulomb metric-based GDF and accelerates the integral evaluation by calculating the slowly convergent, long-range part separately, e.g., in reciprocal space using a PW basis^{15,46} or in real space using a multipole expansion.^{40,41,43,47,48} The global GDF with the Coulomb metric is generally considered more accurate but less computationally efficient than the local one.^{49–53}

Here, we introduce an efficient implementation of a global, Coulomb metric-based GDF for periodic systems. We use the error function to range-separate the Coulomb metric integrals, evaluating the short-range part in real space and the long-range part in reciprocal space, similar in spirit to Refs. 46, 54, and 55. With a few algorithmic developments, we show that the new scheme – which we call range-separated Gaussian density fitting (RSGDF) – scales sublinearly to linearly with N_k for small to medium-sized k -point meshes that are commonly used in periodic calculations with electron correlation.^{14,16,56–59} Numerical tests on three simple three-dimensional solids demonstrate that RSGDF accelerates previous implementations of GDF by an order of magnitude with negligible precision loss in the computed energies. We also show that the accuracy of the HF^{60,61} energy computed using RSGDF can be systematically improved with little extra computational effort by increasing the size of the auxiliary basis; we achieve accuracies on the order of $10^{-6} E_h$ per atom with speedups of one to two orders of magnitude compared to reference GPW calculations.

While finalizing this work, a preprint by Sun⁶² reported a similar range-separation idea to accelerate the direct computation of the four-center Coulomb and the exchange integrals for periodic HF calculations, i.e. without density fitting. Therefore, we will also compare our RSGDF to this new method (referred to as RSJK henceforth) in terms of accuracy and computational cost.

Theory. We begin with a brief review of periodic GDF us-

^{a)}Electronic mail: tim.berkelbach@gmail.com

ing a basis of n_{AO} symmetry-adapted atomic orbitals (AOs)

$$\phi_{\mu}^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{m}} e^{i\mathbf{k}\cdot\mathbf{m}} \phi_{\mu}^{\mathbf{m}}(\mathbf{r}) \quad (1)$$

where \mathbf{k} is a crystal momentum in the first Brillouin zone, \mathbf{m} is a lattice translation vector, and $\phi_{\mu}^{\mathbf{m}}(\mathbf{r}) = \phi_{\mu}(\mathbf{r} - \mathbf{m})$. An analogous equation holds for the n_{aux} auxiliary atom-centered Gaussian basis $\chi_P^{\mathbf{k}}(\mathbf{r})$. The ERIs are the Coulomb repulsion between pair densities

$$(\rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} | \rho_{\lambda\sigma}^{\mathbf{k}_3\mathbf{k}_4}) = \int_{\Omega} d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}(\mathbf{r}_1) \rho_{\lambda\sigma}^{\mathbf{k}_3\mathbf{k}_4}(\mathbf{r}_2)}{r_{12}} \quad (2)$$

where Ω is the unit cell volume, $\rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}(\mathbf{r}) = \phi_{\mu}^{\mathbf{k}_1*}(\mathbf{r}) \phi_{\nu}^{\mathbf{k}_2}(\mathbf{r})$, and the four crystal momenta satisfy $(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{k}_4) \cdot \mathbf{m} = 0$ for all \mathbf{m} . In GDF, the pair densities are approximated by an auxiliary expansion

$$\rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}(\mathbf{r}) \approx \sum_Q^{n_{\text{aux}}} d_{Q\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} \chi_Q^{\mathbf{k}_{12}}(\mathbf{r}), \quad (3)$$

with $\mathbf{k}_{12} = -\mathbf{k}_1 + \mathbf{k}_2$. Minimizing the fitting error in some metric $w(r_{12})$ leads to a linear equation for $\mathbf{d}^{\mathbf{k}_1\mathbf{k}_2}$,

$$\sum_Q^{n_{\text{aux}}} J_{PQ}^{\mathbf{k}_{12}} d_{Q\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} = V_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}, \quad (4)$$

where the two- and three-center metric integrals are

$$J_{PQ}^{\mathbf{k}} = (\chi_P^{\mathbf{k}*} | w | \chi_Q^{\mathbf{k}}), \quad (5)$$

$$V_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} = (\chi_P^{\mathbf{k}_{12}*} | w | \rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}). \quad (6)$$

Once $\{\mathbf{d}^{\mathbf{k}_1\mathbf{k}_2}\}$ are determined, the ERIs can be easily recovered,

$$(\rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} | \rho_{\lambda\sigma}^{\mathbf{k}_3\mathbf{k}_4}) \approx \sum_{P,Q}^{n_{\text{aux}}} d_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} (\chi_P^{\mathbf{k}_{12}*} | \chi_Q^{\mathbf{k}_{34}}) d_{Q\lambda\sigma}^{\mathbf{k}_3\mathbf{k}_4}. \quad (7)$$

The fixed size of the auxiliary Gaussian basis is responsible for a DF error compared to a calculation without DF (throughout, we will call this the *accuracy*, to be contrasted with the *precision* with which the two- and three-center integrals are evaluated for a fixed auxiliary basis). Although in principle the same is true of GPW, the auxiliary PW basis is typically grown to achieve arbitrarily accurate results that are free of DF error.

The computational bottleneck of periodic GDF is due to the three-center integrals in Eq. (6), which, when using the long-ranged Coulomb metric $w(r_{12}) = r_{12}^{-1}$, are expensive to evaluate in real space or reciprocal space. The current implementation of periodic GDF in PySCF¹⁵ aims to address this challenge by introducing a Gaussian charge basis $\{\xi_P^{\mathbf{k}}\}$ to remove the charge and multipoles of the auxiliary basis. This splits Eq. (6) into two parts

$$V_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} = (\chi_P^{\mathbf{k}_{12}} - \xi_P^{\mathbf{k}_{12}} | w | \rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}) + (\xi_P^{\mathbf{k}_{12}} | w | \rho_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}). \quad (8)$$

The Gaussian exponents of $\{\xi_P^{\mathbf{k}}\}$ are optimized so that the first term in Eq. (8) can be evaluated in real space using a lattice summation and the second term in reciprocal space using Eq. (15). Although this yields an improvement over any attempt to evaluate the three-center integrals entirely in real or reciprocal space, the two separate summations can both be relatively slow to converge. Our new periodic RSGDF takes a different approach to evaluate the three-center integrals in Eq. (6). All techniques introduced below can be readily adapted to the evaluation of the two-center integrals in Eq. (5).

In RSGDF, we range-separate the Coulomb operator using the error function⁶³ $r_{12}^{-1} = w^{\text{SR}}(r_{12}; \omega) + w^{\text{LR}}(r_{12}; \omega)$,

$$w^{\text{SR}}(r_{12}; \omega) = \frac{\text{erfc}(\omega r_{12})}{r_{12}} \quad (9a)$$

$$w^{\text{LR}}(r_{12}; \omega) = \frac{\text{erf}(\omega r_{12})}{r_{12}} \quad (9b)$$

so that Eq. (6) is split into a short-range (SR) part and a long-range (LR) part with ω controlling their relative weights. We evaluate the LR integrals in reciprocal space,

$$(V_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2})_{\omega}^{\text{LR}} = 4\pi \sum_{\mathbf{G}}^{N_{\text{PW}}} \frac{e^{-|\mathbf{G} + \mathbf{k}_{12}|^2/4\omega^2}}{|\mathbf{G} + \mathbf{k}_{12}|^2} \tilde{\chi}_P^{\mathbf{k}_{12}}(-\mathbf{G}) \tilde{\rho}_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}(\mathbf{G}), \quad (10)$$

where $\tilde{\chi}$ and $\tilde{\rho}$ are the Fourier transform of χ and ρ and the primed summation indicates $\mathbf{G} \neq \mathbf{0}$ for $\mathbf{k}_1 = \mathbf{k}_2$;¹⁴ the $\mathbf{G} = \mathbf{0}$ term contributes to finite-size errors and is handled on a case-by-case basis in the subsequent electronic structure calculations and not in the ERIs. A relatively small number of PWs are necessary for convergence due to the presence of the Gaussian damping factor. The analytical Fourier transform (AFT) is needed for compact auxiliary orbitals and pair densities, while the fast Fourier transform (FFT) can be used for diffuse ones;^{64,65} we will return to this point later. The cost of this step is therefore dominated by the AFT of the orbital pair densities

$$\tilde{\rho}_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2}(\mathbf{G}) = \sum_{\mathbf{m}}^{N_{\text{cell}}^{\text{AFT}}} e^{-i\mathbf{k}_2 \cdot \mathbf{m}} \int d\mathbf{r} \phi_{\mu}^{\mathbf{0}}(\mathbf{r}) \phi_{\nu}^{\mathbf{m}}(\mathbf{r}) e^{-i(\mathbf{k}_{12} + \mathbf{G}) \cdot \mathbf{r}}. \quad (11)$$

This AFT has two separate steps: the evaluation of the real-space integrals and the subsequent contraction of these integrals with phase factors, which scale as $O(N_k N_{\text{cell}}^{\text{AFT}} N_{\text{PW}} n_{\text{AO}}^2)$ and $O(N_k^2 N_{\text{cell}}^{\text{AFT}} N_{\text{PW}} n_{\text{AO}}^2)$, respectively. Note that the number of unique crystal momentum pair differences grows linearly with N_k and $N_{\text{cell}}^{\text{AFT}}$ can be estimated from the orbital overlap.

The SR part can be easily evaluated in real space by lattice summation

$$(V_{P\mu\nu}^{\mathbf{k}_1\mathbf{k}_2})_{\omega}^{\text{SR}} = \sum_{\mathbf{m}\mathbf{n}}^{N_{\text{cell}}} e^{-i\mathbf{k}_1 \cdot \mathbf{m}} e^{i\mathbf{k}_2 \cdot \mathbf{n}} (V_{P\mu\nu}^{\mathbf{0},\mathbf{m}\mathbf{n}})_{\omega}^{\text{SR}} - \frac{\pi}{\Omega \omega^2} S_P^{\mathbf{k}_{12}} S_{\mu\nu}^{\mathbf{k}_1\mathbf{k}_2} \delta_{\mathbf{k}_1, \mathbf{k}_2} \quad (12)$$

where

$$(V_{P\mu\nu}^{\mathbf{0},\mathbf{m}\mathbf{n}})_{\omega}^{\text{SR}} = \int_{\Omega} d\mathbf{r}_1 \int d\mathbf{r}_2 \chi_P^{\mathbf{0}}(\mathbf{r}_1) w^{\text{SR}}(r_{12}; \omega) \times \phi_{\mu}^{\mathbf{m}}(\mathbf{r}_2) \phi_{\nu}^{\mathbf{n}}(\mathbf{r}_2), \quad (13a)$$

$$S_P^{k_{12}} = \int_{\Omega} d\mathbf{r} \chi_P^{k_{12}}(\mathbf{r}), \quad (13b)$$

$$S_{\mu\nu}^{k_1 k_2} = \int_{\Omega} d\mathbf{r} \rho_{\mu\nu}^{k_1 k_2}(\mathbf{r}), \quad (13c)$$

and the summation range N_{cell} scales as $O(\omega^{-3})$ for three-dimensional solids because ω^{-1} is the decay length of the SR potential. The second term in Eq. (12) cancels the $\mathbf{G} = \mathbf{0}$ component of the first term. With proper integral screening, only $O(N_{\text{cell}})$ terms contribute significantly to the double lattice summation to achieve a finite precision ϵ (see Supplementary Material for a detailed derivation). Therefore, the costs scale as $O(N_{\text{cell}} n_{\text{aux}} n_{\text{AO}}^2)$ for the evaluation of real-space integrals in Eq. (13a) and $O((N_k^2 N_{\text{cell}} + N_k N_{\text{cell}}^2) n_{\text{aux}} n_{\text{AO}}^2)$ for the double phase factor contraction in Eq. (12).

Since $N_{\text{cell}}^{\text{AFT}}$ and N_{cell} can be as large as 10^4 , the phase factor contractions in both Eqs. (11) and (12) would account for most of the computational cost (except for very small k -point meshes). However, if the k -points are sampled from a uniform (e.g. Monkhorst-Pack⁶⁶) mesh that includes the Γ point, then the phase factors satisfy $e^{i\mathbf{k}\cdot\mathbf{m}} = e^{i\mathbf{k}\cdot(\tilde{\mathbf{m}}+M)} = e^{i\mathbf{k}\cdot\tilde{\mathbf{m}}}$, where $\tilde{\mathbf{m}}$ is inside the Born-von Karman supercell and M is a lattice translation vector of the Born-von Karman supercell. For example, the summation in Eq. (12) can then be rewritten as

$$\sum_{\tilde{\mathbf{m}}\tilde{\mathbf{n}}} e^{-i\mathbf{k}_1\cdot\tilde{\mathbf{m}}} e^{i\mathbf{k}_2\cdot\tilde{\mathbf{n}}} \sum_{\mathbf{m}\rightarrow\tilde{\mathbf{m}}, \mathbf{n}\rightarrow\tilde{\mathbf{n}}} (V_{P\mu\nu}^{0,mn})_{\omega}^{\text{SR}}, \quad (14)$$

where the phase factor contraction now costs $O(N_k^3 n_{\text{aux}} n_{\text{AO}}^2)$; a similar treatment for Eq. (11) gives $O(N_k^3 N_{\text{PW}} n_{\text{AO}}^2)$ cost for the phase factor contraction. This process significantly reduces the total cost of these contractions so that they are subdominant (at least for moderately sized k -point meshes where N_k is much smaller than N_{cell} or $N_{\text{cell}}^{\text{AFT}}$). The remain-

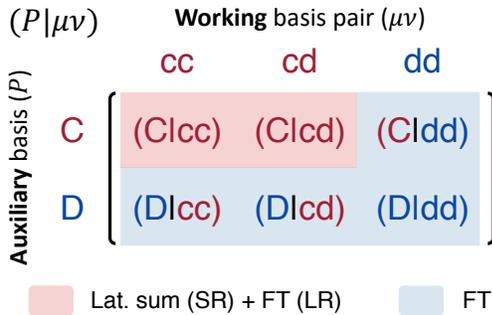


FIG. 1. Schematic illustration of how different components of the three-center integrals [Eq. (6)] are evaluated in RSGDF. Both the auxiliary and the AO bases are split into a compact (“C/c”) set and diffuse (“D/d”) set. The compact integrals (shaded in red) are range separated to yield short-range (SR) and long-range (LR) contributions that are evaluated in real space (by lattice summation) and in reciprocal space (by Fourier transforms), respectively; the diffuse integrals (shaded in blue) are evaluated entirely in reciprocal space using Fourier transforms. Note that the column corresponding to “dc”-type pair densities is similar to the “cd”-column and hence omitted for simplicity.

ing cost-determining steps are the real-space integral evaluations in Eqs. (11) and (13a), which as a reminder scale as $O(N_k N_{\text{cell}}^{\text{AFT}} N_{\text{PW}} n_{\text{AO}}^2)$ and $O(N_{\text{cell}} n_{\text{aux}} n_{\text{AO}}^2)$, respectively. Note that the cost of these expensive steps is no worse than linear in the number of auxiliary Gaussian basis functions.

The algorithm described so far yields significant performance improvements over existing periodic GDF schemes. We have identified an additional minor improvement, motivated by the observation that the real-space lattice summations needed for the SR part are slow to converge because of diffuse orbitals, i.e. those with small Gaussian exponents (recall that the FFT can be used in the LR part for diffuse orbitals). Therefore, we split both the auxiliary and the AO bases into a compact (“C/c”, upper case for auxiliary) and a diffuse (“D/d”) set based on a cutoff α^{cut} for the primitive Gaussian exponents. Similar ideas of compact and diffuse basis splitting have also been explored by Pulay and co-workers for molecular calculations.⁶⁵ This leads us to six types of three-center integrals as shown in Fig. 1. Four of the integral types have either a diffuse bra or a diffuse ket (shaded in blue; note that both “cc” and “cd” are compact) and can thus be readily evaluated in reciprocal space using a relatively small PW basis⁶⁵ *without range separation*,

$$V_{P\mu\nu}^{k_1 k_2} = 4\pi \sum_{\mathbf{G}} \frac{\tilde{\chi}_P^{k_{12}}(-\mathbf{G}) \tilde{\rho}_{\mu\nu}^{k_1 k_2}(\mathbf{G})}{|\mathbf{G} + \mathbf{k}_{12}|^2}. \quad (15)$$

To summarize, in RSGDF, we evaluate these four integral types directly in reciprocal space and the remaining two integral types with compact bra and ket using the range separation scheme defined above. Note that the expensive AFTs of the compact auxiliary orbitals and pair densities can be calculated once and used in the evaluation of both integral types. In practice, a large majority of orbitals are defined as compact and so we find that this separation of orbitals speeds up our calculations by a factor of two or less compared to a direct application of RSGDF for all orbitals.

While the above presentation suggests a computational scaling that is linear in N_k for typical mesh densities, the actual scaling is complicated by the choice of the parameters, N_{PW} , ω , and α^{cut} , which we discuss more below. Empirically we find that the optimal choice of N_{PW} scales as $O(N_k^{-1/2})$ (Fig. S3), and the overall cost of RSGDF scales roughly as $O(N_k^{0.8})$ for all the systems tested in this work (Fig. S6). However, this sublinear scaling only holds for small N_k , because N_{PW} eventually reaches a minimum value. Beyond that point, the AFTs needed for the LR part dominate the cost and we expect a linear scaling of RSGDF with N_k , at least for medium-sized k -point meshes.

Computational details. We implemented RSGDF as presented above in a local version of the PySCF software package.² We test its performance in terms of precision, accuracy, and computational efficiency using three simple three-dimensional solids: diamond, MgO, and LiF. For diamond, we perform all-electron calculations using the cc-pVDZ basis,⁶⁷ for MgO and LiF, we use GTH pseudopotentials^{68,69} and the corresponding GTH-DZVP basis.¹⁸ We use the cc-pVDZ-jkfit basis⁷⁰ and the even-tempered basis (ETB) generated with a

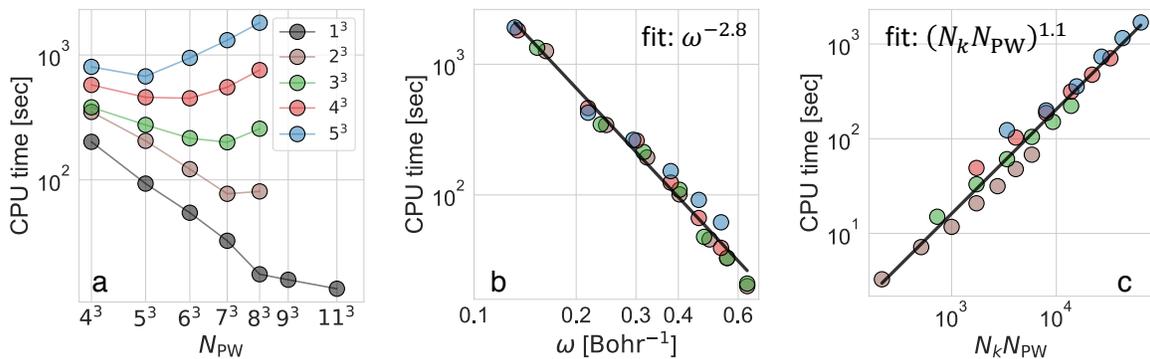


FIG. 2. Timing of RSGDF for diamond/cc-pVDZ using $N_k = 1^3$ to 5^3 . (a) Total DF initialization CPU time as a function of the number of PWs used to compute the LR integrals [Eqs. (10) and (15)]. (b) CPU time for computing the SR integrals [Eq. (12)] as a function of ω for all k -point meshes except for $N_k = 1^3$. (c) CPU time for computing the LR integrals [Eqs. (10) and (15)] as a function of $N_k N_{PW}$ for all k -point meshes except for $N_k = 1^3$. For (b) and (c), the black lines are power law fits to all data points leading to exponents as shown.

progression factor $\beta = 2.0$ for the auxiliary expansion of the cc-pVDZ and the GTH-DZVP bases, respectively.

We compare RSGDF to GDF,¹⁵ GPW (called FFTDF in PySCF),¹⁸ and RSJK,⁶² as implemented in PySCF. For RSJK, we use eq. (23) in ref. 62 to determine an appropriate ω for a given system and k -point mesh. For RSGDF, we manually test a range of N_{PW} and for each we determine the maximum ω and α^{cut} that guarantee precision ϵ in all integrals; future work will focus on the automated selection of these parameters. As a general trend, using a larger PW basis slows down the LR part by increasing the number of expensive real-space integrations to be performed in Eq. (11) but accelerates the SR part by allowing larger values for ω and α^{cut} ; a smaller PW basis has the opposite effect. Unless otherwise mentioned, all calculations are run with a target precision of $\epsilon = 10^{-8}$ a.u. for integral evaluation, which is the default setting for production-level periodic calculations in PySCF. The finite-size error of the HF exchange energy is corrected with a Madelung constant, which yields $O(N_k^{-1})$ convergence to the thermodynamic limit^{71–73} (Fig. S7); other possibilities exist^{73–76} but would require modification of the DF algorithm. All timing data reported below are the CPU time recorded using a single CPU core (Intel Xeon Gold 6126 2.6 GHz) with 16 GB of memory and 100 GB of disk space except for GPW which requires larger memory for $N_k \geq 4^3$. The current implementations of GDF and RSGDF are not integral-direct, meaning that we solve Eq. (4) only once and save the coefficients $\{\mathbf{d}^{k_1 k_2}\}$ to disk for later use; this step is called “DF initialization” below and requires $O(N_k^2 n_{aux} n_{AO}^2)$ disk space which limits our calculations to a maximum k -point mesh of $N_k = 5^3$ for all three systems. The other two methods, GPW and RSJK, are both implemented in an integral-direct manner and hence require little disk space. An integral-direct implementation of RSGDF tailored for specific applications will be presented in future work.

Results and discussion. We first verify our scaling analysis of the CPU cost of RSGDF. In Fig. 2a, we show the RSGDF initialization time as a function of N_{PW} for diamond using $N_k = 1^3$ to 5^3 k -points (recall that all calculations achieve

the same target precision). The optimal N_{PW} – identified as the minimum on each curve – indeed decreases with N_k as $O(N_k^{-1/2})$ (the fitted exponent is about -0.44 ; see Fig. S3). The inverse cubic dependence of the SR time on ω is verified in Fig. 2b, and the linear scaling of the LR time with $N_k N_{PW}$ is verified in Fig. 2c. Similar results are observed for the other two systems (Figs. S1 and S2), although the optimal values of N_{PW} for a given N_k vary slightly from system to system. We leave the automatic determination of the optimal N_{PW} to a future work as it requires a more careful calibration. In what follows, we will simply use the manually optimized values from Fig. 2a and Figs. S1a and S2a (summarized in Tab. S1).

The different choices of N_{PW} (and hence ω and α^{cut}) in RSGDF do not cause any inconsistency in the computed energies. As shown in Fig. S4, the converged RSGDF HF energies differ from the GDF results by less than $10^{-7} E_h$ for diamond and MgO and about $10^{-6} E_h$ for LiF for all data points shown in Fig. 2 and Figs. S1 and S2. We attribute the larger deviation observed for LiF to the linear dependency found in the auxiliary basis. Nonetheless, these deviations are acceptable as they are at least one order of magnitude smaller than the error introduced by DF itself (*vide infra*). Beyond the HF energy, we have also verified that the electron correlation energy of diamond computed with RSGDF using the second order Møller-Plesset perturbation theory⁷⁷ agrees with the GDF results to better than $10^{-8} E_h$ for all k -point meshes tested (Tab. S2). These observations confirm that the algorithmic developments in RSGDF cause negligible precision loss compared to the original implementation of GDF.

Next, we study the computational efficiency of RSGDF. In Fig. 3, we plot the per-SCF-cycle time as a function of N_k for computing the Coulomb and the exchange integrals in a HF calculation for all three systems using four different methods to handle the ERIs: GPW (green), RSJK (red), GDF (grey), and RSGDF (blue). Since the first two are implemented in an integral-direct fashion, we include the DF initialization time for GDF and RSGDF to enable a fair comparison.

We first compare GDF and RSGDF, which both compute the three-center integrals through a SR part in real space and

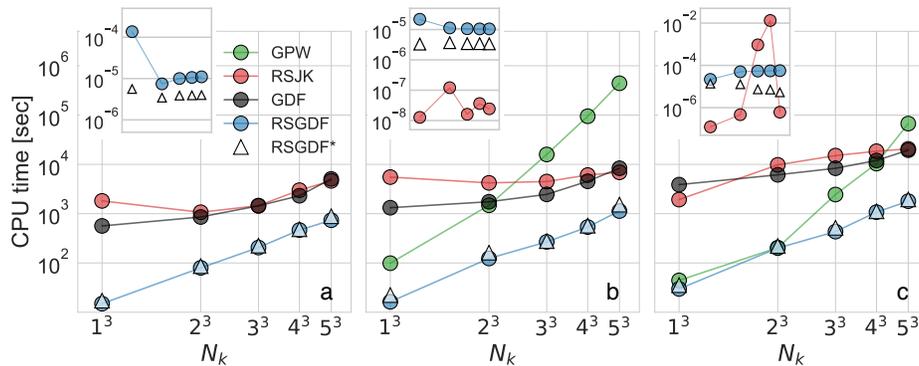


FIG. 3. CPU time (per SCF cycle) for computing the Coulomb and the exchange integrals in a HF calculation for (a) diamond/cc-pVDZ, (b) MgO/GTH-DZVP, and (c) LiF/GTH-DZVP using GPW (green), RSJK (red), GDF (grey), and RSGDF (blue) to handle the ERIs. For GDF and RSGDF, the DF initialization time is included. For RSGDF, results using a larger auxiliary basis are also included (RSGDF*, white triangles). Precise timing data are given in Table S3. Insets show the deviation of the RSGDF HF energies from RSJK for diamond (a) and deviations of the RSGDF and the RSJK HF energies from GPW for MgO (b) and LiF (c), where the x-axis is N_k and the y-axis is in E_h .

a LR part in reciprocal space. Although both methods exhibit a similar sublinear scaling with N_k at large N_k , only the GDF timings plateau at small N_k . This difference arises from the adjustment of the PW basis size for each N_k and for each system (Tab. S1), which ultimately balances the SR-LR cost in RSGDF as analyzed above (Fig. S5). More importantly, the algorithmic optimizations developed for RSGDF in this work significantly reduce its computational cost and lead to speedups of one to two orders of magnitude over the previous GDF for all three systems studied here.

We next compare RSGDF with the two other methods without DF error, i.e. GPW and RSJK. The GPW timing shows the characteristic $O(N_k^2)$ scaling of computing exact exchange starting from $N_k = 2^3$ and is 40 to 400 times slower than RSGDF for the largest N_k tested here. The very high cost of GPW for MgO is caused by the compact primitive Gaussians in the $2s$ and $2p$ orbitals of Mg, which require 83^3 PWs to reach the target precision of 10^{-8} (cf. 51^3 for LiF). We emphasize that lowering the precision requirement for GPW (hence lowering N_{PW}) only moderately reduces the cost due to the $O(N_{PW} \ln N_{PW})$ scaling of FFT. For example, using $\epsilon = 10^{-6}$ and 10^{-4} requires 75^3 and 66^3 PWs for MgO and reduces the cost by only factors of 1.4 and 2.1, respectively.

The RSJK timings are similar to those of GPW; although they show a slightly weaker dependence on N_k , the precise scaling is unclear. This peculiar N_k -dependence of RSJK arises from a significant SR-LR cost unbalance (Fig. S5) and suggests a breakdown of eq. (23) in ref. 62 for determining the optimal ω for RSJK. Despite this, RSJK still achieves computational efficiency similar to GDF and much higher than GPW for moderately sized k -point meshes, which is remarkable given that RSJK does not use DF.

Finally, we examine the accuracy of the HF energies computed by RSGDF, which is a combination of the DF error due to the auxiliary basis set and the precision error when calculating the matrices in Eqs. (5) and (6). The RSJK results are used as the benchmark for diamond and the GPW results are used as the benchmark for MgO and LiF. To probe the possibility of

achieving higher accuracy with DF, we also include results for RSGDF using a larger auxiliary basis (denoted by ‘‘RSGDF*’’ in Fig. 3); we use the cc-pVTZ-jkfit basis for diamond and an ETB whose size is about 1.25 times larger (obtained by using a smaller β) for MgO and LiF. The per-atom errors of the converged HF energies are plotted in the insets of Fig. 3.

With the default auxiliary basis, the error introduced by RSGDF is about $10^{-5} E_h$ for diamond and MgO and about $10^{-4} E_h$ for LiF; these errors are typical for DF-based HF calculations.^{40,46} Using the slightly larger auxiliary basis (RSGDF*) reduces the error by a factor of three or more and, most remarkably, requires little extra work as can be seen from the nearly identical timings of RSGDF and RSGDF* in Fig. 3. This is because the LR part is dominated by the AFTs of the orbital pair densities, Eq. (11), whose cost is independent of n_{aux} , while the SR part scales linearly with n_{aux} . By contrast, the accuracy of RSJK is in general very high ($10^{-7} E_h$ or less), but relatively large errors of about $10^{-2} E_h$ are also observed for certain k -point meshes (inset of Fig. 3c); the accuracy loss in the latter cases is likely due to an inaccurate integral screening, as tightening the precision to 10^{-10} reduces the error to about $10^{-5} E_h$ for the calculation of LiF using 4^3 k -points. These results demonstrate that RSGDF provides an extremely cost-effective approach to calculating accurate HF energies in periodic systems.

Conclusion. To summarize, we have presented an efficient scheme that uses range separation for Gaussian density fitting (RSGDF) for periodic systems. The computational scaling is analyzed to be sublinear with N_k for small k -point meshes and linear for medium-sized ones. With all-electron and pseudopotential-based numerical results on a few three-dimensional solids, we verified the scaling of RSGDF and showed that it achieves about 10-fold speedups over the previously developed GDF with little precision loss. The error introduced by RSGDF is about $10^{-5} E_h$ with default auxiliary basis sets and can be systematically reduced by increasing the size of the auxiliary basis with little extra work.

The primary purpose of the current integral-indirect imple-

mentation of RSGDF is to speed up Hartree-Fock (and hybrid density functional theory) calculations for a given Gaussian basis and auxiliary basis. Motivated by the excellent performance seen in these preliminary calculations, we are currently working on the automatic determination of the optimal N_{PW} , ω , and α^{cut} . Looking forward, the fast integral construction enabled by RSGDF encourages the development of integral-direct algorithms tailored to specific tasks such as the evaluation of exact exchange,⁴⁴ the ERI orbital transformation,^{37,38} and post-HF calculations.^{38,42} Such integral-direct methods would significantly reduce the high memory footprint currently required by post-HF calculations on periodic systems with Gaussian basis sets.

SUPPLEMENTARY MATERIAL

See the supplementary material for (i) RSGDF initialization time for different choices of N_{PW} and ω for MgO and LiF; (ii) optimal choices of N_{PW} for N_k from 1^3 to 5^3 and different systems; (iii) difference of the HF energies computed using RSGDF and GDF; (iv) SR and LR component time of RSGDF, GDF, and RSJK for varying size of k -point meshes; (v) N_k -scaling of RSGDF and GDF for the timing data shown in Fig. 3; (vi) values of the parameters needed by RSGDF, GDF, RSJK, and GPW; (vii) comparison of the MP2 correlation energies computed using RSGDF and GDF for diamond; (viii) CPU time for computing the Coulomb and exchange integrals per SCF cycle by RSGDF, GDF, RSJK, and GPW; (ix) **Details of the treatment of exchange divergence**; (x) derivations of the conditions for prescreening the double lattice summation in Eq. (12).

ACKNOWLEDGEMENTS

HY thanks Dr. Qiming Sun and Dr. Xiao Wang for helpful discussions. This work was supported by the National Science Foundation under Grant No. OAC-1931321. We acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010. The Flatiron Institute is a division of the Simons Foundation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

¹R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman, *Wiley Interdiscip. Rev. Comput. Mol. Sci* **8**, e1360 (2018).

- ²Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, *Wiley Interdiscip. Rev. Comput. Mol. Sci* **8**, e1340 (2018).
- ³T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöb, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, *J. Chem. Phys.* **152**, 194103 (2020).
- ⁴S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu, *J. Chem. Phys.* **152**, 184107 (2020).
- ⁵J. Ihm, A. Zunger, and M. L. Cohen, *J. Phys. C: Solid State Phys.* **12**, 4409 (1979).
- ⁶R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- ⁷J. L. Martins and M. L. Cohen, *Phys. Rev. B* **37**, 6134 (1988).
- ⁸W. E. Pickett, *Comput. Phys. Rep.* **9**, 115 (1989).
- ⁹G. Kresse and J. Furthmüller, *Comput. Mater. Sci.* **6**, 15 (1996).
- ¹⁰L. Maschio and D. Usvyat, *Phys. Rev. B* **78**, 073102 (2008).
- ¹¹A. F. Izmaylov and G. E. Scuseria, *Phys. Chem. Chem. Phys.* **10**, 3421 (2008).
- ¹²S. Hirata and T. Shimazaki, *Phys. Rev. B* **80**, 085118 (2009).
- ¹³L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri, *J. Chem. Phys.* **132**, 134706 (2010).
- ¹⁴J. McClain, Q. Sun, G. K.-L. Chan, and T. C. Berkelbach, *J. Chem. Theory Comput.* **13**, 1209 (2017).
- ¹⁵Q. Sun, T. C. Berkelbach, J. D. McClain, and G. K.-L. Chan, *J. Chem. Phys.* **147**, 164119 (2017).
- ¹⁶X. Wang and T. C. Berkelbach, *J. Chem. Theory Comput.* **16**, 3095 (2020).
- ¹⁷H. K. Buchholz and M. Stein, *J. Comput. Chem.* **39**, 1335 (2018).
- ¹⁸J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Commun.* **167**, 103 (2005).
- ¹⁹J. L. Whitten, *J. Chem. Phys.* **58**, 4496 (1973).
- ²⁰B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *J. Chem. Phys.* **71**, 3396 (1979).
- ²¹J. W. Mintmire and B. I. Dunlap, *Phys. Rev. A* **25**, 88 (1982).
- ²²E. Baerends, D. Ellis, and P. Ros, *Chem. Phys.* **2**, 41 (1973).
- ²³O. Vahtras, J. Almlöf, and M. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- ²⁴Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, *Proc. Natl. Acad. Sci.* **102**, 6692 (2005).
- ²⁵S. Reine, E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansik, S. Høst, and P. Salek, *J. Chem. Phys.* **129**, 104101 (2008).
- ²⁶J. G. Hill, *Int. J. Quantum Chem.* **113**, 21 (2013).
- ²⁷A. Sodt, J. E. Subotnik, and M. Head-Gordon, *J. Chem. Phys.* **125**, 194109 (2006).
- ²⁸A. Sodt and M. Head-Gordon, *J. Chem. Phys.* **128**, 104106 (2008).
- ²⁹S. Manzer, P. R. Horn, N. Mardirossian, and M. Head-Gordon, *J. Chem. Phys.* **143**, 024113 (2015).
- ³⁰H.-J. Werner, F. R. Manby, and P. J. Knowles, *J. Chem. Phys.* **118**, 8149 (2003).
- ³¹H. Eshuis, J. Yarkony, and F. Furche, *J. Chem. Phys.* **132**, 234114 (2010).
- ³²H.-J. Werner and M. Schütz, *J. Chem. Phys.* **135**, 144116 (2011).
- ³³C. Riplinger and F. Neese, *J. Chem. Phys.* **138**, 034106 (2013).
- ³⁴W. Györfy, T. Shiozaki, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **138**, 104104 (2013).
- ³⁵v. Varga, *Phys. Rev. B* **71**, 073103 (2005).
- ³⁶v. Varga, M. Milko, and J. Noga, *J. Chem. Phys.* **124**, 034106 (2006).
- ³⁷L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, *Phys. Rev. B* **76**, 075101 (2007).
- ³⁸D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani, *Phys. Rev. B* **76**, 075102 (2007).
- ³⁹C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat, *J. Comput. Chem.* **29**, 2113 (2008).

- ⁴⁰A. M. Burow, M. Sierka, and F. Mohamed, *J. Chem. Phys.* **131**, 214101 (2009).
- ⁴¹R. Lazarski, A. M. Burow, and M. Sierka, *J. Chem. Theory Comput.* **11**, 3029 (2015).
- ⁴²A. Luenser, H. F. Schurkus, and C. Ochsenfeld, *J. Chem. Theory Comput.* **13**, 1647 (2017).
- ⁴³M. M. J. Grundei and A. M. Burow, *J. Chem. Theory Comput.* **13**, 1159 (2017).
- ⁴⁴X. Wang, C. A. Lewis, and E. F. Valeev, *J. Chem. Phys.* **153**, 124116 (2020).
- ⁴⁵D. Usvyat, L. Maschio, and M. Schütz, “Periodic local möller-plesset perturbation theory of second order for solids,” in *Handbook of Solid State Chemistry* (American Cancer Society, 2017) Chap. 3, pp. 59–86.
- ⁴⁶C. H. Patterson, *J. Chem. Phys.* **153**, 064107 (2020).
- ⁴⁷R. Lazarski, A. M. Burow, L. Grajciar, and M. Sierka, *J. Comput. Chem.* **37**, 2518 (2016).
- ⁴⁸M. Becker and M. Sierka, *J. Comput. Chem.* **40**, 2563 (2019).
- ⁴⁹v. Varga, *Int. J. Quantum Chem.* **108**, 1518 (2008).
- ⁵⁰v. Varga, *J. Math. Chem.* **49**, 1 (2011).
- ⁵¹P. Merlot, T. Kjærgaard, T. Helgaker, R. Lindh, F. Aquilante, S. Reine, and T. B. Pedersen, *J. Comput. Chem.* **34**, 1486 (2013).
- ⁵²G. Schmitz and O. Christiansen, *Chem. Phys. Lett.* **701**, 7 (2018).
- ⁵³L. N. Wirz, S. S. Reine, and T. B. Pedersen, *J. Chem. Theory Comput.* **13**, 4897 (2017).
- ⁵⁴T. Shimazaki, T. Kosugi, and T. Nakajima, *J. Phys. Soc. Japan* **83**, 054702 (2014).
- ⁵⁵S. Sharma and G. Beylkin, “Efficient evaluation of gaussian integrals in periodic systems,” (2020), [arXiv:2010.05400](https://arxiv.org/abs/2010.05400).
- ⁵⁶A. Grüneis, M. Marsman, and G. Kresse, *J. Chem. Phys.* **133**, 074107 (2010).
- ⁵⁷A. Grüneis, *J. Chem. Phys.* **143**, 102817 (2015).
- ⁵⁸F. Hummel, T. Gruber, and A. Grüneis, *Eur. Phys. J. B* **89**, 235 (2016).
- ⁵⁹W. Mayr-Schmölzer, J. Planer, J. Redinger, A. Grüneis, and F. Mittendorfer, *Phys. Rev. Research* **2**, 043361 (2020).
- ⁶⁰C. Pisani and R. Dovesi, *Int. J. Quantum Chem.* **17**, 501 (1980).
- ⁶¹R. Dovesi, R. Orlando, C. Roetti, C. Pisani, and V. Saunders, *phys. stat. sol. (b)* **217**, 63 (2000).
- ⁶²Q. Sun, “Exact exchange matrix of periodic hartree-fock theory for all-electron simulations,” (2020), [arXiv:2012.07929](https://arxiv.org/abs/2012.07929).
- ⁶³P. M. Gill and R. D. Adamson, *Chem. Phys. Lett.* **261**, 105 (1996).
- ⁶⁴L. Füsti-Molnar and P. Pulay, *J. Chem. Phys.* **116**, 7795 (2002).
- ⁶⁵L. Füsti-Molnar and P. Pulay, *J. Chem. Phys.* **117**, 7827 (2002).
- ⁶⁶H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **13**, 5188 (1976).
- ⁶⁷T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- ⁶⁸S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- ⁶⁹C. Hartwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).
- ⁷⁰F. Weigend, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).
- ⁷¹J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, *J. Chem. Phys.* **124**, 154709 (2006).
- ⁷²P. Broqvist, A. Alkauskas, and A. Pasquarello, *Phys. Rev. B* **80**, 085114 (2009).
- ⁷³R. Sundararaman and T. A. Arias, *Phys. Rev. B* **87**, 165122 (2013).
- ⁷⁴F. Gygi and A. Baldereschi, *Phys. Rev. B* **34**, 4405 (1986).
- ⁷⁵J. Spencer and A. Alavi, *Phys. Rev. B* **77**, 193110 (2008).
- ⁷⁶M. Guidon, J. Hutter, and J. VandeVondele, *J. Chem. Theory Comput.* **5**, 3010 (2009).
- ⁷⁷C. Möller and M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).