Fast periodic Gaussian density fitting by range separation
Hong-Zhou Ye1 and Timothy C. Berkelbach1, 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.1,10-16 The downside of atom-centered orbitals is the introduction of four-index electron repulsion integrals (ERIs), with $O(N_k^2 n_{AO}^2)$ storage and $O(N_k^2 n_{AO}^2)$ 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) method 17 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 (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.21-24 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 sets 25 with $n_{aux} \approx 3 n_{AO}$ have made GDF a great success in both mean-field 26-28 and correlated calculations.29-33 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.34-41 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.36-38,41,43,44 The locality could arise from the system itself,43 an explicit use of a local metric other than the Coulomb operator,51 or a combination of both.36 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 37,45 or in real space using a multipole expansion. 39,40,42,46,47 The global GDF with the Coulomb metric is generally considered more accurate but less computationally efficient than the local one.48-52

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 the range-separated Kohn-Sham (RSJK) method.53-57 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 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 Sun 60 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-
ing a basis of \( n_{AO} \) symmetry-adapted atomic orbitals (AOs)
\[
\phi_k^\mu(r) = \sum_m e^{i k \cdot m} \phi_m^\mu(r)
\]  
(1)
where \( k \) is a crystal momentum in the first Brillouin zone, \( m \) is a lattice translation vector, and \( \phi_m^\mu(r) = \phi_m^\mu(r - m) \). An analogous equation holds for the \( n_{aux} \) auxiliary atom-centered Gaussian basis \( \varphi_k^\mu(r) \). The ERIs are the Coulomb repulsion between pair densities
\[
(\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4}) = \int_{\Omega} dr_1 \int_{\Omega} dr_2 \rho_{\mu\nu}^{k_1 k_2}(r_1) \rho_{\mu\nu}^{k_3 k_4}(r_2) r_{12}^{-1}
\]  
(2)
where \( \Omega \) is the unit cell volume, \( \rho_{\mu\nu}^{k_1 k_2}(r) = \phi_m^\mu(\mathbf{r})\phi_n^\nu(\mathbf{r}) \), and the four crystal momenta satisfy \((k_1 - k_2 + k_3 - k_4) \cdot m = 0\) for all \( m \). In GDF, the pair densities are approximated by an auxiliary expansion
\[
\rho_{\mu\nu}^{k_1 k_2}(r) \approx \sum_{Q} \hat{d}_{Q}^{k_1 k_2} \varphi_{Q}^{k_1}(r),
\]  
(3)
with \( k_{12} = -k_1 + k_2 \). Minimizing the fitting error in some metric \( \rho(r_{12}) \) leads to a linear equation for \( \hat{d}_{Q}^{k_1 k_2} \)
\[
\sum_{Q} \rho_{Q}^{k_{12}} \rho_{Q}^{k_{12}} = \chi_{Q}^{k_{12}},
\]  
(4)
where the two- and three-center metric integrals are
\[
\rho_{Q}^{k_{1}} = \left( \chi_{P}(|w|\chi_{Q}^{k_{1}}) \right),
\]  
(5)
\[
\rho_{Q}^{k_{12}} = \left( \chi_{P}(|w|\chi_{Q}^{k_{12}}) \right).
\]  
(6)
Once \( \{\hat{d}_{Q}^{k_1 k_2}\} \) are determined, the ERIs can be easily recovered,
\[
(\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4}) \approx \sum_{Q} \hat{d}_{PQ}^{k_{12}} \left( \chi_{P}^{k_{12}} | w | \chi_{Q}^{k_{12}} \right) \hat{d}_{Q}^{k_{12}},
\]  
(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 PySCF aims to address this challenge by introducing a Gaussian charge basis \( \xi_{P}^{k} \) to remove the charge and multipole of the auxiliary basis. This splits Eq. (6) into two parts
\[
\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4} = (\chi_{P}^{k_1} | w | \chi_{P}^{k_2}) + (\xi_{P}^{k_1} | w | \chi_{P}^{k_2}).
\]  
(8)
The Gaussian exponents of \( \{\xi_{P}^{k}\} \) 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).

In RSGDF, we range-separate the Coulomb operator using the error function
\[
\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4} = w^{SR}(r_{12}; \omega) + w^{LR}(r_{12}; \omega),
\]  
(9)
where
\[
\rho_{\mu\nu}^{SR}(r_{12}; \omega) = \text{erf}(\omega r_{12})/r_{12},
\]  
(9a)
\[
\rho_{\mu\nu}^{LR}(r_{12}; \omega) = \text{erf}(\omega r_{12})/r_{12},
\]  
(9b)
so that Eq. (6) is split into a short-range (SR) part and a long-range (LR) part with \( \omega \) controlling their relative weights. We evaluate the LR integrals in reciprocal space,
\[
(\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4})^{LR} = 4\pi \sum_{G} N_{\mu\nu}^{G} e^{-|G + k_{12}|^2 \omega^2} \chi_{P}^{k_{12} | | 2G}(G) - \chi_{P}^{k_{12} | | 2G}(G),
\]  
(10)
where \( \chi \) and \( \rho \) are the Fourier transform of \( \chi \) and \( \rho \) and the primed summation indicates \( G \neq 0 \) for \( k_1 = k_2 \). The \( G = 0 \) component is canceled by similar terms arising from the electron-nuclear and nuclear-nuclear energy. 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. We will return to this point later. The cost of this step is therefore dominated by the AFT of the orbital pair densities
\[
\tilde{\rho}_{\mu\nu}^{k_1 k_2}(G) = \sum_{n_{AO}} \text{e}^{-i k_{1} \cdot m} \int dr \phi_{m}^{\mu}(r) \phi_{m}^{\nu}(r) e^{-i(k_{1} + 2G) \cdot r}. \]  
(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 \( O(N_{\mu\nu}^{AO})N_{PW}^{2}N_{AO}^{2} \) and \( O(N_{\mu\nu}^{AO})N_{PW}^{2}N_{AO}^{2} \), respectively. Note that the number of unique crystal momentum pair differences grows linearly with \( N_{\mu\nu}^{AO} \) and \( N_{AO}^{2} \) can be estimated from the orbital overlap.

The SR part can be easily evaluated in real space by lattice summation
\[
(\rho_{\mu\nu}^{k_1 k_2 | | k_3 k_4})^{SR} = \sum_{mn} \text{e}^{-i k_{1} \cdot m} e^{-i k_{2} \cdot n} (\rho_{\mu\nu}^{m n})^{SR}_{\omega} - \frac{\pi}{\Omega_{AO}^2} \chi_{P}^{k_{1}} \chi_{P}^{k_{2}} \delta_{k_{1}, k_{2}}.
\]  
(12)
where
\[
(\rho_{\mu\nu}^{m n})^{SR}_{\omega} = \int_{\Omega} dr_1 \int_{\Omega} dr_2 \text{e}^{0}(r_1) w^{SR}(r_{12}; \omega) \times \phi_{m}^{\mu}(r_2) \phi_{n}^{\nu}(r_2),
\]  
(13a)
the summation range \( N_{\text{cell}} \) scales as \( O(\omega^{-3}) \) for three-dimensional solids because \( \omega^{-1} \) is the decay length of the SR potential. The second term in Eq. (12) cancels the \( G = 0 \) 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 \( \epsilon \) (see Supplementary Material for a detailed derivation). Therefore, the costs scale as \( O(N_{\text{cell}}n_{\text{aux}}n_{AO}^{2}) \) for the evaluation of real-space integrals in Eq. (13a) and \( O((N_{k}^{2}N_{\text{cell}} + N_{k}^{2}n_{\text{aux}}^{2}n_{AO}^{2}) \) for the double phase factor contraction in Eq. (12).

Since \( N_{\text{AFT}} \) and \( N_{\text{cell}} \) 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-Pack\(^{64} \)) mesh that includes the \( \Gamma \) point, then the phase factors satisfy \( e^{i\mathbf{k} \cdot \mathbf{r}_{\text{m}}} = e^{i\mathbf{k} \cdot \mathbf{r}_{\text{m}} + M} = e^{i\mathbf{k} \cdot \mathbf{r}_{\text{m}}} \), where \( \mathbf{r}_{\text{m}} \) 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

\[
\sum_{\mathbf{m} \in \Omega} \sum_{\mathbf{n} \in \Omega} e^{i\mathbf{k}_{1} \cdot \mathbf{r}_{\text{m}}} e^{i\mathbf{k}_{2} \cdot \mathbf{r}_{\text{m}}} \langle \mathbf{V}_{P\mu\nu} \rangle_{\omega}^{\text{SR}},
\]

where the phase factor contraction now costs \( O(N_{k}^{3}n_{\text{aux}}n_{AO}^{2}) \); a similar treatment for Eq. (11) gives \( O(N_{k}^{2}N_{\text{cell}}n_{\text{aux}}^{2}n_{AO}^{2}) \) 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 \( N_{\text{cell}} \) or \( N_{\text{AFT}} \)). The remaining cost-determining steps are the real-space integral evaluations in Eqs. (11) and (13a), which as a reminder scale as \( O(N_{k}^{2}N_{\text{cell}}n_{\text{aux}}^{2}n_{AO}^{2}) \) and \( O(N_{\text{cell}}n_{\text{aux}}n_{AO}^{2}) \), 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 \( \alpha_{\text{cut}} \) 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.\(^{63} \) 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 basis\(^{43} \) without range separation.

\[
V_{P\mu\nu}^{k_{1}k_{2}} = 4\pi \sum_{G} \langle \mathbf{V}_{P\mu\nu} \rangle_{G} \frac{\chi_{k_{1}k_{2}}^{\dagger}(G)\chi_{k_{1}k_{2}}^{\dagger}(G)}{|G + k_{12}|^{2}}.
\]

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, \( N_{\text{PW}}, \omega_{\text{PW}}, \) and \( \alpha_{\text{cut}} \), which we discuss more below. Empirically we find that the optimal choice of \( N_{\text{PW}} \) 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 \( N_{\text{PW}} \) 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-pVQZ basis;\(^{65} \) for MgO and LiF, we use GTH pseudopotentials\(^{56,67} \) and the corresponding GTH-DZVP basis.\(^{17} \) We use the cc-pVQZ-jkfit basis\(^{68} \) and the even-tempered basis (ETB) generated with a
progression factor $\beta = 2.0$ for the auxiliary expansion of the cc-pVDZ and the GTH-DZVP bases, respectively.

We compare RSGDF to GDF,\textsuperscript{14} GPW (called FFTDF in PySCF),\textsuperscript{17} and RSJK,\textsuperscript{60} as implemented in PySCF. For RSJK, we use eq. (23) in ref. 60 to determine an appropriate $\omega$ 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 $\omega$ and $\alpha_{\text{cut}}$ that guarantee precision $\epsilon$ 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 $\omega$ and $\alpha_{\text{cut}}$, 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. 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 $O(N_k^2n_{\text{aux}}n_{\text{AO}}^2)$ 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 $\omega$ is verified in Fig. 2b, and the linear scaling of the LR time with $N_kN_{pw}$ 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 $\omega$ and $\alpha_{\text{cut}}$) 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 theory\textsuperscript{69} 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 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 \(O(N_e^2)\) 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 2\(s\) and 2\(p\) 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 \(10^{-7}\) 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. 60 for determining the optimal \(\omega\) 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-jkit basis for diamond and an ETB whose size is about 1.25 times larger (obtained by using a smaller \(\beta\)) 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.\(^{39,45}\) 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, which makes up at least half of the total cost of RSGDF, is dominated by the AFTs of the orbital pair densities, Eq. (11), whose cost is independent of \(n_{aux}\). 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 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 \(4^2\) \(k\)-points. These results demonstrate that RSGDF provides an extremely cost-effective approach to calculating accurate HF energies in periodic systems.

\textbf{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 implementation 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 αcut. 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, the ERI orbital transformation, and post-HF calculations. 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 for different choices of Npw and ω for MgO and LiF; (ii) optimal choices of Npw for Np from 1 to 5 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; (v) Nω-scaling of RSGDF and GDF for the timing shown in Fig. 3; (vi) values of the parameters needed by RSGDF, GDF, RSJK, and GWF; (vii) comparison of the MP2 correlation energies computed using RSGDF and GDF for diamond; (viii) 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.

1. R. 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).
2. Q. 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).
3. T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khalilullin, O. Schütt, F. Schüttmann, 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. Guion, S. Andermatt, N. Holmberg, G. K. Schenter, A. Henn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Gliöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, J. Chem. Phys. 152, 194103 (2020).
4. S. G. Balasubramanii, 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. Tapavica, D. P. Tew, C. van Wullen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu, J. Chem. Phys. 152, 184107 (2020).
5. J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C: Solid State Phys. 12, 4409 (1979).
6. R. Car and M. Parrinello, Phys. Rev. Lett. 55, 2471 (1985).
7. J. L. Martins and M. L. Cohen, Phys. Rev. B 37, 6134 (1988).
8. W. E. Pickett, Comput. Phys. Rep. 9, 115 (1989).
9. G. Kresse and J. Hutter, Comput. Mater. Sci. 6, 15 (1996).
10. A. F. Izmaylov and G. E. Scuseria, Phys. Chem. Chem. Phys. 10, 3421 (2008).
11. S. Hirata and T. Shimazaki, Phys. Rev. B 80, 055118 (2009).
12. L. Maschio, D. Usyvayt, M. Schütz, and B. Civalleri, J. Chem. Phys. 132, 134706 (2010).
13. J. McClain, Q. Sun, G. K.-L. Chan, and T. C. Berkelbach, J. Chem. Theory Comput. 13, 1209 (2017).
14. Q. Sun, T. C. Berkelbach, J. D. McClain, and G.-K.-L. Chan, J. Phys. Chem. C 147, 164117 (2019).
15. X. X. Wang and T. C. Berkelbach, J. Chem. Theory Comput. 16, 3905 (2020).
16. H. K. Buchholz and M. Stein, J. Chem. Comput. Chem. 39, 1335 (2018).
17. J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. 167, 103 (2005).
18. J. L. Whitten, J. Chem. Phys. 58, 4496 (1973).
19. B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. 71, 3396 (1979).
20. J. W. Mintmire and B. I. Dunlap, Phys. Rev. A 25, 88 (1982).
21. E. Baerends, D. Ellis, and P. Ros, Chem. Phys. 2, 41 (1973).
22. O. Vahtras, J. Almlof, and M. Feyereisen, Chem. Phys. Lett. 213, 514 (1993).
23. Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Natl. Acad. Sci. 102, 6692 (2005).
24. S. Reine, E. Tellgren, A. Krapp, T. Kjergaard, T. Helgaker, B. Jansik, S. Host, and P. Salek, J. Chem. Phys. 129, 104101 (2008).
25. J. G. Hill, Int. J. Quantum Chem. 113, 21 (2013).
26. A. Sodt, J. E. Subotnik, and M. Head-Gordon, J. Chem. Phys. 125, 194109 (2006).
27. A. Sodt and M. Head-Gordon, J. Chem. Phys. 128, 104106 (2008).
28. S. Manzer, P. R. Horn, N. Mardirossian, and M. Head-Gordon, J. Chem. Phys. 143, 024113 (2015).
29. H.-J. Werner, F. R. Manby, and J. P. Knowles, J. Chem. Phys. 118, 8149 (2003).
30. H. Eschuis, J. Yarkony, and F. Furchi, J. Chem. Phys. 132, 234114 (2010).
31. H.-J. Werner and M. Schütz, J. Chem. Phys. 135, 144116 (2011).
32. C. Riplinger and F. Neese, J. Chem. Phys. 138, 034106 (2013).
33. W. Györffy, T. Shiozaki, G. Knizia, and H.-J. Werner, J. Chem. Phys. 138, 104104 (2013).
34. V. Varga, Phys. Rev. B 71, 073103 (2005).
35. V. Varga, M. Milko, and J. Noga, J. Chem. Phys. 124, 034106 (2006).
36. L. Maschio, D. Usyvayt, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, Phys. Rev. B 76, 075101 (2007).
37. D. Usyvayt, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani, Phys. Rev. B 76, 075102 (2007).
38. C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usyvayt, J. Comput. Chem. 29, 2113 (2008).
39. A. M. Burow, M. Sierka, and F. Mohamed, J. Chem. Phys. 131, 214101 (2009).
40. R. Lazarski, A. M. Burow, and M. Sierka, J. Chem. Theory Comput. 11, 3029 (2015).
41. A. Luenser, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Theory Comput. 13, 1647 (2017).
42. M. M. J. Grundei and A. M. Burow, J. Chem. Theory Comput. 13, 1159 (2017).
43. X. Wang, C. A. Lewis, and E. F. Valeev, J. Chem. Phys. 153, 124116 (2020).
44. D. Usyvayt, L. Maschio, and M. Schütz, “Periodic local moller-plesset perturbation theory of second order for solids,” in Handbook of Solid State Science.
53. C. Møller and M. S. Plesset, Phys. Rev. 46, 618 (1934).
54. F. Weigend, Phys. Chem. Chem. Phys. 4, 4285 (2002).
55. C. Häser and M. S. Plesset, Phys. Rev. 46, 618 (1934).
56. P. M. Gill and R. D. Adamson, Chem. Phys. Lett. 261, 105 (1996).
57. L. Füsti-Molnar and P. Pulay, J. Chem. Phys. 117, 7827 (2002).
58. H. J. Monkhorst and J. D. Pack, Phys. Rev. B 13, 5188 (1976).
59. T. H. Dunning, J. Chem. Phys. 90, 1007 (1989).
60. S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B 54, 1703 (1996).
61. C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B 58, 3641 (1998).
62. F. Weigend, Phys. Chem. Chem. Phys. 4, 4285 (2002).
63. G. Schmitz and O. Christiansen, Chem. Phys. Lett. 701, 7 (2018).
64. L. N. Wirz, S. S. Reine, and T. B. Pedersen, J. Chem. Theory Comput. 13, 4897 (2017).
65. V. Varga, Int. J. Quantum Chem. 108, 1518 (2008).
66. A. Grüneis, J. Chem. Phys. 143, 102817 (2015).
Supplementary Material: Fast Periodic Gaussian Density Fitting By Range Separation

Hong-Zhou Ye† and Timothy C. Berkelbach*,†

†Department of Chemistry, Columbia University, New York, New York 10027, USA
‡Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

E-mail: tim.berkelbach@gmail.com

Contents

S1 Supplementary figures 2
S2 Supplementary tables 5
S3 Conditions for prescreening the double lattice summation 7
  S3.1 General considerations ........................................ 7
  S3.2 A bound for \( R_c \) ........................................ 9
  S3.3 A bound for \( R_{12} \) ........................................ 10
  S3.4 Summary .................................................... 12

Note: figures and equations appearing in the main text will be referred as “Fig. Mxxx” and “Eq. Mxxx” in this Supplementary Material document.
S1 Supplementary figures

**Figure S1:** Same plot as Fig. M1 for MgO/GTH-DZVP.

**Figure S2:** Same plot as Fig. M1 for LiF/GTH-DZVP.
Figure S3: Optimal choice of $N_{PW}$ (obtained from Fig. M1 and Figs. S1 and S2) plotted as a function of $N_k$. The estimated exponents are $-0.44$ (C), $-0.38$ (MgO), and $-0.43$ (LiF).

Figure S4: Absolute deviation between the converged HF energies computed using RSGDF and GDF for (a) diamond/cc-pVDZ, (b) MgO/GTH-DZVP, and (c) LiF/GTH-DZVP using $1^3$ to $5^3$ $k$-points.

Figure S5: Component timing data (SR: darker; LR: lighter) plotted as a function of $N_k$ for RSGDF (left), GDF (middle), and RSJK (right).
Figure S6: Same data for GDF and RSGDF as plotted in Fig. M3 with power-law fitting ($t = aN_k^\gamma$) where the exponent $\gamma$ is displayed in the figure. For RSGDF, the last four points are used, while for GDF, the last three points are used.
### S2 Supplementary tables

Table S1: Parameters for each method used in this work. RSGDF/RSJK: $\omega$ and $N_{PW}$. GDF: charge basis exponent $\alpha_{chg}$ and $N_{PW}$. GPW: $N_{PW}$.

#### RSGDF

| $N_k$ | $N_{k12}$ | Diamond | MgO | LiF |
|-------|-----------|---------|-----|-----|
|       | $\omega$ | $N_{PW}$ | $\omega$ | $N_{PW}$ | $\omega$ | $N_{PW}$ |
| $1^3$ | 1        | 0.792570 | 11$^3$ | 0.671035 | 11$^3$ | 0.707182 | 11$^3$ |
| $2^3$ | 14       | 0.494208 | 7$^3$ | 0.415922 | 7$^3$ | 0.494208 | 8$^3$ |
| $3^3$ | 63       | 0.476279 | 7$^3$ | 0.340426 | 6$^3$ | 0.429530 | 7$^3$ |
| $4^3$ | 172      | 0.379822 | 6$^3$ | 0.323641 | 6$^3$ | 0.403150 | 7$^3$ |
| $5^3$ | 365      | 0.294591 | 5$^3$ | 0.323641 | 6$^3$ | 0.264873 | 5$^3$ |

#### RSJK

| $N_k$ | Diamond | MgO | LiF |
|-------|---------|-----|-----|
|       | $\omega$ | $N_{PW}$ | $\omega$ | $N_{PW}$ | $\omega$ | $N_{PW}$ |
| $1^3$ | 5.664491 | 69$^3$ | 5.768690 | 83$^3$ | 3.700892 | 51$^3$ |
| $2^3$ | 1.888164 | 23$^3$ | 2.224073 | 32$^3$ | 2.902661 | 40$^3$ |
| $3^3$ | 1.149317 | 14$^3$ | 1.320543 | 19$^3$ | 1.741596 | 24$^3$ |
| $4^3$ | 0.820941 | 10$^3$ | 0.973032 | 14$^3$ | 1.233631 | 17$^3$ |
| $5^3$ | 0.656753 | 8$^3$ | 0.764525 | 11$^3$ | 1.015931 | 14$^3$ |
| $6^3$ | 0.574659 | 7$^3$ | 0.625521 | 9$^3$ | 0.798232 | 11$^3$ |
| $7^3$ | 0.492564 | 6$^3$ | 0.556018 | 8$^3$ | &nbsp; | &nbsp; |
| $8^3$ | 0.410470 | 5$^3$ | 0.486516 | 7$^3$ | &nbsp; | &nbsp; |
| $9^3$ | 0.410470 | 5$^3$ | 0.417014 | 6$^3$ | &nbsp; | &nbsp; |
| $10^3$ | 0.410470 | 5$^3$ | 0.347511 | 5$^3$ | &nbsp; | &nbsp; |

#### GDF

| $N_k$ | Diamond $\alpha_{chg}$ | MgO $\alpha_{chg}$ | LiF $\alpha_{chg}$ |
|-------|-------------------------|---------------------|--------------------|
|       | $N_{PW}$ | $N_{PW}$ | $N_{PW}$ | $N_{PW}$ | $N_{PW}$ | $N_{PW}$ |
| All   | 0.200000 | 7$^3$ | 0.200000 | 7$^3$ | 0.200000 | 7$^3$ |

#### GPW

| $N_k$ | C $N_{PW}$ | MgO $N_{PW}$ | LiF $N_{PW}$ |
|-------|-----------|-------------|-------------|
| All   | 31$^3$    | 83$^3$      | 51$^3$      |
Table S2: MP2 correlation energies of diamond/cc-pVDZ computed with the HF solution and ERIs calculated using RSGDF and GDF. For $N_k \leq 3^3$, all bands in the cc-pVDZ basis are used, while for $N_k = 4^3$, only the valence occupied bands and the first 10 conduction bands are correlated to fit the memory limit.

| $N_k$ | $E^{(2)}_{\text{RSGDF}} / E_h$ | $E^{(2)}_{\text{GDF}} / E_h$ | $E^{(2)}_{\text{RSGDF}} - E^{(2)}_{\text{GDF}} / nE_h$ |
|-------|-------------------------------|-------------------------------|---------------------------------|
| 1$^3$ | $-0.1702783512$               | $-0.1702783506$               | $-0.58$                         |
| 2$^3$ | $-0.2444412960$               | $-0.2444412984$               | $2.40$                          |
| 3$^3$ | $-0.2642540235$               | $-0.2642540236$               | $0.13$                          |
| 4$^3$ | $-0.2015870651$               | $-0.2015870661$               | $1.03$                          |
S3 Conditions for prescreening the double lattice summation

Figure S7: Schematic illustration of the various vectors defined in the text. Note that without loss of generality, the auxiliary orbital $\chi_P$ is assumed to be located at the origin.

Here, we derive conditions for prescreening the double lattice summation in Eq. (M12), i.e., the short-range (SR) three-center Coulomb integrals. The basic idea is to analyze how each term decays in real space and then derive truncation conditions to achieve a desired precision $\epsilon$ in the computed integrals.

S3.1 General considerations

Eq. (M12) can be rewritten as (ignoring the phase factors)

$$V_{P_{\mu\nu}} = \sum_{m,n} v_{P_{\mu\nu}}(R_{11}^m, R_{22}^n; \omega)$$  \hspace{1cm} (S1)

where

$$v_{P_{\mu\nu}}(R_{11}^m, R_{22}^n; \omega) = \int_{\Omega} dr_1 \int dr_2 \chi_0^0(r_1) w_S^{SR}(r_{12}; \omega) \phi_{\mu}^{R_{11}^m}(r_2) \phi_{\nu}^{R_{22}^n}(r_2),$$  \hspace{1cm} (S2)

$R_{11}^m = \tau_1 + m$ and $R_{22}^n = \tau_2 + n$, with $\tau_1$ and $\tau_2$ the position of the two AOs in the reference unit cell (Fig. S7). For the simple case of all 1s-type primitive Gaussians (with exponents $\alpha_P$, $\alpha_{\mu}$, and $\alpha_{\nu}$ and coefficients $C_P$, $C_{\mu}$, and $C_{\nu}$), the integral in Eq. (S2) can be easily worked out. The result is

$$v_{P_{\mu\nu}}(R_{11}^m, R_{22}^n; \omega) = Ah(R_{12}^{m,n}) f(R_{c}^{m,n}; \omega)$$  \hspace{1cm} (S3)
where

$$A = \left( \frac{\pi}{4} \right)^3 \frac{C_p C_{\mu} C_{\nu}}{\alpha_p^{3/2} (\alpha_{\mu} + \alpha_\nu)^{3/2}},$$  \hspace{1cm} (S4)$$

$$h(R_{12}^{m,n}) = e^{-\beta(R_{12}^{m,n})^2},$$ \hspace{1cm} (S5)$$

with $\beta = \frac{\alpha_{\mu} \alpha_\nu}{\alpha_{\mu} + \alpha_\nu}$ and $R_{12}^{m,n} = R_m^m - R_n^n$, and

$$f(R_c^{m,n}; \omega) = \frac{1}{R_c^{m,n}} [\text{erf}(\eta_1 R_c^{m,n}) - \text{erf}(\eta_2 R_c^{m,n})]$$ \hspace{1cm} (S6)$$

with

$$\eta_1 = [\alpha_p^{-1} + (\alpha_{\mu} + \alpha_\nu)^{-1}]^{-1/2},$$ \hspace{1cm} (S7)$$

$$\eta_2 = (\eta_1^{-2} + \omega^{-2})^{-1/2},$$ \hspace{1cm} (S8)$$

and

$$R_c^{m,n} = \frac{\alpha_{\mu} R_1^m + \alpha_\nu R_2^n}{\alpha_{\mu} + \alpha_\nu}.$$ \hspace{1cm} (S9)$$

It can be shown that both $h$ and $f$ are positive-valued, monotonically decreasing functions.

Define two new lattice vectors,

$$n' = n - m$$ \hspace{1cm} (S10)$$

$$m' = R_c^{m,n} - \tau_{c,n'}$$ \hspace{1cm} (S11)$$

where $\tau_{c,n'}$ lies in the reference unit cell. Equation (S1) can be rewritten as

$$V_{\mu\nu} = A \sum_{n'} h(R_{12}^{0,n'}) \sum_m f(R_c^{m,m+n'}; \omega)$$

$$= A \sum_{n'} h(R_{12}^{0,n'}) \sum_{m'} f(|m' + \tau_{c,n'}|; \omega)$$ \hspace{1cm} (S12)$$

Equation (S12) changes the double lattice summation in Eq. (S1) from over the bare lattice vectors ($m$ and $n$) to over the center ($m'$) and the difference ($n'$) vectors. From now on, we will drop the prime in the vectors $m'$ and $n'$. 

8
S3.2 A bound for \( R_c \)

We first estimate a bound for truncating \( m \), or equivalently \( R_c \). Let \( S_h = \sum_n h(R_{12}^0, n) \). The error made by truncating the \( m \)-summation is

\[
\delta \approx AS_h \sum_{m > m_{cut}} f(m; \omega)
\approx AS_h \frac{4\pi}{\Omega} \int_{R_{cut}^c}^\infty dR \left[ \text{erf}(\eta_1 R) - \text{erf}(\eta_2 R) \right]
\approx AS_h \frac{4\pi}{\Omega} \frac{1}{\sqrt{\pi}} \int_{R_{cut}^c}^\infty dR \left( \eta_2^{-1} e^{-\eta_2^2 R^2} - \eta_1^{-1} e^{-\eta_1^2 R^2} \right)
\leq AS_h \frac{4\pi}{\Omega} \frac{1}{\sqrt{\pi}} R_{cut}^c \int_{R_{cut}^c}^\infty dR \left( \eta_2^{-3} e^{-(\eta_2 R_{cut}^c)^2} - \eta_1^{-3} e^{-(\eta_1 R_{cut}^c)^2} \right)
= AS_h \frac{4\pi}{\Omega} \frac{1}{\sqrt{\pi}} R_{cut}^c \frac{1}{2} \left( \eta_2^{-3} e^{-(\eta_2 R_{cut}^c)^2} - \eta_1^{-3} e^{-(\eta_1 R_{cut}^c)^2} \right)
\approx AS_h \frac{4\pi}{\Omega} \frac{1}{\sqrt{\pi}} \frac{1}{2} \frac{1}{\eta_2^3 R_{cut}^c} e^{-(\eta_2 R_{cut}^c)^2}
\]

where we have assumed that the final cutoff \( R_{cut}^c \) is not too small so that

1. in line 1, \( f(\{m + \tau_{c,n}\}; \omega) \approx f(m) \),
2. in line 2, the summation can be turned into an integral with \( (4\pi/\Omega)R^2 \) being the density of states,
3. in line 3, the leading-order asymptotic expansion of the integrand is valid (error \( \sim O(R^{-2}) \)),
4. in line 6, the \( \eta_1 \)-related term can be dropped.

These assumptions could break down if both the AOs and the auxiliary orbitals are very compact so that \( R_{cut}^c \) is small. However, for virtually all AO bases, even the most compact AOs have non-compact primitive Gaussians that control the long-range behavior and hence the magnitude of \( R_{cut}^c \). For this reason, the assumptions made above virtually hold for all AO bases. A bound for \( R_c \) is hence obtained by requiring

\[
AS_h \frac{4\pi}{\Omega} \frac{1}{\sqrt{\pi}} \frac{1}{2} \frac{1}{\eta_2^3 R_{cut}^c} e^{-(\eta_2 R_{cut}^c)^2} < \epsilon. \tag{S14}
\]
Note that for estimating this bound, $S_h$ does not need to be evaluated very accurately. In practice, we found that performing a small lattice summation

$$S_h \approx \sum_{n < n(\epsilon')} h(R_{12}^{0,n})$$  \hspace{1cm} (S15)$$
suffices, where $n(\epsilon')$ is determined such that $h(R_{12}^{0,n(\epsilon')}) < \epsilon'$. We found that $\epsilon' = 10^{-4}$ already leads to convergent results.

### S3.3 A bound for $R_{12}$

We then determine a bound for truncating $n$, or equivalently $R_{12}$. First, we note that since the final integral depends on the product of $h$ and $f$, a simple proposal like

$$h(R_{12}) < \epsilon$$  \hspace{1cm} (S16)$$
is not appropriate: it is too tight when $f(R_c; \omega) \ll 1$ (i.e., large $R_c$) and too loose when $f(R_c; \omega) \gg 1$ (i.e., small $R_c$).

To that end, we determine here a bound for $R_{12}$ that is $R_c$-dependent. We first decouple the two summations

$$V_{P_{\mu \nu}} \leq A \sum_m f_{m}^{\max}(\omega) \sum_n h(R_{12}^{0,n})$$  \hspace{1cm} (S17)$$
where

$$f_{m}^{\max}(\omega) = \max_{\tau \in \Omega_0} f(|m + \tau|; \omega)$$  \hspace{1cm} (S18)$$
($m \in \Omega_0$ denotes all vectors in the reference unit cell). Now consider the error introduced by truncating the $n$-summation for $m$ in a finite shell $R \in R + \Delta R$

$$\delta_{R-R+\Delta R} \leq AN_{\text{cell}}(R, R + \Delta R) f(R; \omega) \sum_{n > n_{\text{cut}}} h(R_{12}^{0,n})$$  \hspace{1cm} (S19)$$
where $N_{\text{cell}}(R, R + \Delta R)$ is the number of cells in the range $R \sim R + \Delta R$, and the "$\leq$" comes from the
fact that $f$ is a monotonically decreasing function. We found that $\Delta R = 1 \text{ Bohr}$ and

$$N_{\text{cell}}(R, R + \Delta R) = \begin{cases} \frac{4\pi}{\Omega} \times 2(R + \Delta R)^2 \Delta R, & R > R_0 \\ N_{\text{cell}}^{\text{NF}}, & R \leq R_0 \end{cases}$$  \hfill (S20)

works well, where $N_{\text{cell}}^{\text{NF}}$ is the number of cells within a sphere with radius $3\Omega^{1/3}$, and $R_0$ is determined as the crossing point of the two pieces of Eq. (S20)

$$\frac{4\pi}{\Omega} \times 2(R_0 + \Delta R)^2 \Delta R = N_{\text{cell}}^{\text{NF}}.$$

(S21)

The motivation behind Eq. (S20) is that the continuous approximation of density of states works well for large $R$ but not for small $R$. The summation in Eq. (S19) can be approximated by an integral

$$\sum_{R > R_{\text{cut}}} h(R^0_n) \approx \frac{4\pi}{\Omega} \int_{R_{\text{cut}}^{\text{cut}}(R)}^{\infty} dR \left( R^2 e^{-\beta R^2} \approx \frac{4\pi}{\Omega} \frac{R_{\text{cut}}^{\text{cut}}(R)}{2\beta} e^{-\beta (R_{\text{cut}}^{\text{cut}}(R))^2} \right).$$  \hfill (S22)

Combining Eqs. (S19), (S20) and (S22), a bound of $R_{12}$ for $R < R_c < R + \Delta R$ is found by

$$AN_{\text{cell}}(R, R + \Delta R)f(R; \omega)\frac{4\pi}{\Omega} \frac{R_{\text{cut}}^{\text{cut}}(R)}{2\beta} e^{-\beta (R_{\text{cut}}^{\text{cut}}(R))^2} < \epsilon.$$  \hfill (S23)

However, when $R_{12}^{\text{cut}}$ found by Eq. (S23) is too small, the integral in Eq. (S22) is not a good approximation to the lattice summation in Eq. (S19). Thus, we use

$$R_{12}^{\text{cut}} = \max\{\text{Eq. (S23)}, R_{12}^{\text{cut, min}}\}$$  \hfill (S24)

where $R_{12}^{\text{cut, min}}$ is determined by

$$N_{\text{cell}}(R_{12}^{\text{cut, min}}, R_{12}^{\text{cut, min}} + \Delta R)\frac{R_{12}^{\text{cut, min}}}{2\beta} e^{-\beta (R_{12}^{\text{cut, min}})^2} = \sum_{R_{12}^{\text{cut, min}}} h(R^0_n),$$  \hfill (S25)

i.e., the minimal $R_{12}$ where the integral in Eq. (S22) remains a good approximation to the summa-
tion. Similar to \( S_h \) in Eq. (S14), the summation on the right-hand side of Eq. (S25) need not be evaluated very accurately and can be safely truncated using the \( n(\epsilon') \) determined for Eq. (S15).

**S3.4 Summary**

To summarize, if both the AOs and the auxiliary orbitals are 1\( s \)-type primitive Gaussians,

1. we first use Eq. (S14) to derive a bound for \( R_c \), denoted by \( R_c^{\text{cut}} \),

2. we then make a mesh for \( R_c \) with increment \( \Delta R \), i.e.,

\[
0, \Delta R, 2\Delta R, \cdots, R_c^{\text{cut}},
\]

(S26)

and use Eq. (S24) to determine a series of bounds for \( R_{12} \), denoted by \( \{R_{12}^{\text{cut}}(R_c)\} \), each for a window of \( R_c \).

With these two bounds, a term \( v_{P\mu\nu}(R_1^m, R_2^n; \omega) \) is not discarded only if

\[
R_c^{m,n} \leq R_c^{\text{cut}} \leq R_{12}^{m,n}
\]

(S27)

The bounds discussed above are determined for every AO shell pair \((\mu\nu)\) and auxiliary shell \((P)\). For primitive Gaussians with higher angular momentum, we found that treating them as 1\( s \) works very well (note that the normalization constant needs to be recalculated using the 1\( s \)-orbital normalization condition). For a contracted Gaussian orbital, the long-range behavior is determined solely by the most diffuse primitive Gaussian orbital. We can hence use the latter for determining the bounds.