UC Berkeley
UC Berkeley Previously Published Works

Title

An energy decomposition analysis for second-order Møller–Plesset perturbation theory based on absolutely localized molecular orbitals

Permalink

Journal

ISSN

Authors

Thirman, Jonathan
Head-Gordon, Martin

Publication Date

DOI

Peer reviewed

Fast, Accurate Evaluation of Exact Exchange: the occ-RI-K Algorithm

Samuel Manzer,[1] Paul R. Horn,[1] Narbe Mardirossian,[1] and Martin Head-Gordon[1, a)]

Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, University of California, Berkeley, and, Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720, USA.

Construction of the exact exchange matrix, **K**, is typically the rate-determining step in hybrid density functional theory, and therefore new approaches with increased efficiency are highly desirable. We present a framework with potential for greatly improved efficiency by computing a compressed exchange matrix that yields the exact exchange energy, gradient, and DIIS error vector. The compressed exchange matrix is constructed with one index in the compact molecular orbital basis, and the other index in the full atomic orbital basis. To illustrate the advantages, we present a practical algorithm that uses this framework in conjunction with the resolution of the identity (RI) approximation. We demonstrate that convergence using this occ-RI-K method in combination with the DIIS algorithm is well-behaved, that the accuracy of computed energetics is excellent (identical to conventional RI-K), and that significant speedups can be obtained over existing integral-direct and RI-K methods. For a 4400 basis function $C_{68}H_{22}$ hydrogen-terminated graphene fragment, our algorithm yields a 14x speedup over the conventional algorithm, and a speedup of 3.3x over RI-K.

I. INTRODUCTION

The widespread adoption of Kohn-Sham density functional theory[1,2] (DFT) has allowed for electronic structure calculations of reasonable accuracy to be performed on systems of unprecedented size.[3–5] Perdew's famous "Jacob's Ladder" analogy elegantly captures the different components of the ideal high-accuracy density functional.[6] A crucial component of this hierarchy are the so-called "hybrid functionals", containing a weighted contribution from the exact exchange energy. The importance of including exact exchange was initially recognized by Becke[7,8], and exact exchange has since become a core component of a huge number of widely used modern density functionals[9–13].

Unfortunately, exact exchange requires the computation of the numerous and expensive four-center two-electron repulsion integrals (ERIs):

$$(\mu\lambda|\nu\sigma) = \int\int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \phi_\mu^*(\boldsymbol{r}_1)\phi_\lambda(\boldsymbol{r}_1)\boldsymbol{r}_{12}^{-1}\phi_\nu(\boldsymbol{r}_2)^*\phi_\sigma(\boldsymbol{r}_2) \tag{1}$$

to construct the exchange matrix, **K**:

$$K_{\mu\nu} = \sum_{\lambda\sigma}(\mu\lambda|\nu\sigma)P_{\lambda\sigma} \tag{2}$$

This is in contrast to highly efficient linear scaling evaluation of the classical Coulomb interactions, where four center ERIs can be rigorously avoided by multipole expansions of long-range interactions[14], and rigorous reformulation of the short-range interactions (the "J-matrix engine"[15,16]). Naive construction of the full two-electron

integral tensor scales as the fourth power of molecular size. However, it was demonstrated by Almlöf early in the development of the direct SCF method that the number of numerically significant two-electron integrals scales asymptotically only quadratically with molecular size, as there are linear number of overlapping pairs of Gaussian atomic orbitals.[17] This locality is most commonly exploited via the use of Schwarz-inequality based integral bounds,[18,19] though tighter bounds on numerically significant integrals have since been further developed.[20–22]

In addition, the density matrix, $\rho(\mathbf{r}, \mathbf{r}')$ itself decays exponentially with distance, $|\mathbf{r} - \mathbf{r}'|$, in the position representation for systems with a non-negligible HOMO-LUMO gap.[3,23,24] For electronic structure theory based on localized atomic orbital (AO) basis functions, this translates into sparsity of the density matrix in the AO representation, as well as potential localization of the molecular orbitals (MOs). This sparsity is basis-set dependent and has been studied empirically.[25,26] For appropriate insulating systems and compact basis sets, significant sparsity is present and may be exploited to formulate efficient conditionally linear-scaling integral-driven algorithms.[27–29] Such methods make hybrid DFT calculations possible on very large insulating systems, provided the basis set is compact.

However, the situation worsens dramatically in larger basis sets for two reasons. First, the number of significant two-electron integrals scales as the fourth power of the mean number of basis functions on a single atomic center, leading to a tremendous pre-factor increase. Second, as larger basis sets contain additional diffuse functions, the condition number of the overlap matrix becomes larger, leading to decreased sparsity as well, and thus loss of effectiveness of linear scaling. The combination of these two factors makes large basis exact exchange calculations almost prohibitively more expensive than small basis set calculations. Unfortunately, hybrid functionals *must* be used with basis sets of at least triple-

zeta quality to obtain adequately converged thermodynamic quantities.[30,31] In smaller basis sets there is no accuracy advantage to hybrid functionals. For the same reason, it has been shown that in order to obtain relatively basis-set independent functional parameters, density functionals should be trained (and thus used) with triple-zeta or larger basis sets.[32] Functionals trained close to the basis set limit yield significantly inferior performance in smaller basis sets[13,33,34] This paper attempts to make progress on the challenging problem of constructing the K matrix in these large basis sets.

Methods based on the resolution of the identity (RI) approximation (also frequently referred to as the "density fitting" or DF approximation) offer the best-established avenue to reducing the computational cost of large basis set DFT calculations. These methods expand the atomic orbital "primary basis" products as linear combinations of "auxiliary basis" functions:

$$\phi_\mu^*(\boldsymbol{r}_1)\phi_\lambda(\boldsymbol{r}_1) = \sum_Q C_Q^{\mu\lambda} \chi_Q(\boldsymbol{r}_1) \tag{3}$$

Highly efficient RI algorithms have been developed for the Coulomb problem[35,36], along with the auxiliary Coulomb fitting basis sets necessary to use such algorithms.[37–39] Turning to exchange, for which appropriate auxiliary basis sets are also available[40,41], the RI approximation results in a much more tractable expression for the K matrix in which only three and two-center quantities are present:

$$K_{\mu\nu} = \sum_{\lambda\sigma} \sum_{PQ} C_P^{\mu\lambda} C_Q^{\nu\sigma} \left(P|Q\right) P_{\lambda\sigma} \tag{4}$$

By reducing the dimensionality of the AO basis quantities, these methods attain significant improvements in performance for large basis sets. This comes at a formal cost; the subsequent recombination of these lower-dimensional quantities exhibits fourth-order scaling. However, the fact that the relevant quantities can be compressed in earlier steps by transformation to the MO basis allows for a significant advantage in the prefactor of the calculation.[40,42] It should be noted that while RI-based algorithms for constructing K are necessarily significantly more complex than those used for J, the underlying approximation is the same.

Extensive effort has been devoted to lowering the scaling of RI-SCF. These approaches may be broadly classified into two categories: local fit metrics[43–46] and local fit domains[42,47–50]. In the local fit metric approach, the fit coeffients are determined by minimizing the self-repulsion between the fitted and exact densities as measured by a rapidly decaying local analog of the Coulomb operator, rather than the more typical full Coulomb metric. Local fit domain approximations are applied by expanding a given primary basis product only with the auxiliary basis functions within a predefined spatial domain; they differ from local metrics in that locality is achieved explicitly

through a constraint rather than implicitly through the decay of the metric.

Two alternative approaches to the problem of large basis set K matrix calculations should be mentioned. The first is the Cholesky decomposition, which has been shown to be equivalent to RI expansion of the two-electron integrals in a customized auxiliary basis.[51] It may thus be regarded as an idealization of the RI approximation in some sense, and the algorithms based on CD and RI are often closely related.[52] A second related approach is the pseudospectral or "semi-numerical" approach introduced to quantum chemistry by Friesner[53] and substantially improved by Neese et al.[54] Pseudospectral methods apply numerical quadrature to one of the two electronic coordinates in the two-electron integral formula (1). In essence, this replaces a given two-electron integral with a large number of one-electron integrals resembling nuclear attraction integrals. When initially compared to the RI-K method, the pseudospectral approach was found to yield comparable speedups and accuracy,[55] while subsequent work has found superior speedups for large systems.[56]

While efforts to reduce the scaling of RI-K are necessary in the long run, in this work we focus on reducing the prefactor of the calculation. Combined with the already significant prefactor advantage of standard RI-K, we obtain a method that outperforms linear scaling integral evaluation and existing alternative RI methods for systems of up to 4400 basis functions in quadruple-zeta quality basis sets, and even outperforms conventional algorithms in small double-zeta basis sets for a surprisingly large size regime. We achieve this by introducing a framework in which one constructs a much smaller "economized" exchange matrix that nonetheless yields energies, orbital first variations, and DIIS error vectors that are exact. Within this framework, we then adapt the RI-K method to yield a new algorithm, termed the occ-RI-K method, which is intrinsically more efficient than RI-K. While our new approach can be combined with local RI approximations through the use of localized MOs, here we concentrate on the standard global Coulomb RI approximation. We demonstrate that the resulting algorithm is both faster and more accurate than our previous PARI-K implementation,[42] as well as much faster and just as accurate as RI-K itself.

II. THEORY

A. Economization of the K matrix

Using standard spin-orbital notation, the exchange energy, E_K, which enters a global hybrid functional in scaled form (i.e. $0.54E_K$ for M06-2X or $0.2E_K$ for B3LYP) can be written in terms of $\mathbf{K}$, defined in Eq.

2, as follows:

$$E_K = -\sum_{\mu\nu i} c_{\mu i} K_{\mu\nu} c_{\nu i} \tag{5}$$

where we have expanded the AO density matrix $P_{\mu\nu} = \sum_i c_{\mu i} c_{\nu i}$ in terms of the MO coefficients $c_{\mu i}$. Evidently to evaluate the energy, it is possible to simply evaluate the diagonal elements of the exchange matrix in the occupied block, K_{ii}, rather than the full matrix in the AO representation, $K_{\mu\nu}$.

However, more information is required to evaluate the orbital gradient of E_K, as needed for performing an SCF calculation. Variations of the MO coefficients away from our current set are given as a unitary transformation, U, which is parameterized in terms of the exponential of a skew-symmetric matrix composed of mixings, Δ_{ai}, between occupied orbitals, i, and empty orbitals, a.

$$U = \exp\left[\Delta - \Delta^T\right] \tag{6}$$

It is then straightforward to show[57] that

$$\frac{\partial E_K}{\partial \Delta_{ai}} = 2K_{ai} \tag{7}$$

Evidently the gradient of the exact exchange energy requires only the occupied-virtual block, K_{ia}, rather than the full matrix in the AO representation, $K_{\mu\nu}$. Together, the energy and the gradient require only $K_{i\nu}$ rather than the full $K_{\mu\nu}$.

The Direct Inversion of the Iterative Subspace (DIIS) method introduced by Pulay is perhaps the most widely used convergence acceleration method.[58,59] For SCF convergence acceleration, the standard form of DIIS[59] accumulates a sequence of error vectors defined as the AO-basis representation of the commutator $[\hat{f}, \hat{\rho}]$. The commutator, which is proportional to the SCF orbital gradient (cf Eq. 7 for the exact exchange gradient), vanishes at SCF convergence. The AO basis commutator at the mth iteration is:

$$\mathbf{e}_m = \mathbf{F}_m \mathbf{P}_m \mathbf{S} - \mathbf{S} \mathbf{P}_m \mathbf{F}_m \tag{8}$$

where $\mathbf{S}$ is the AO overlap matrix. DIIS extrapolation to obtain a new candidate solution is performed by minimizing a weighted linear combination of error vectors, $\sum_m w_m \mathbf{e}_m$, subject to the normalization constraint, $\sum_m w_m = 1$. The extrapolated Fock matrix, $F_{extrap} = \sum_m w_m \mathbf{F}_m$ is then diagonalized to define the updated density matrix, $\mathbf{P}_{m+1}$, and thus $\mathbf{F}_{m+1}$, etc.

Since the Fock matrix appears in Eq. 8 only when multiplied by the density matrix, a sequence of DIIS iterates can be correctly evaluated using only the occupied row-block of the K matrix. This yields exactly the same set of DIIS coefficients as if the full K-matrix were used to evaluate the error vector. Hence construction of the full AO-basis K-matrix is potentially wasteful, as only the entries $K_{i\nu}$ are required to obtain an exact DIIS error vector

and exchange energy. In addition it is trivial to show that the occupied eigenvalues are exactly reproduced. *This is all that is typically required in a standard DFT calculation, and therefore exact exchange algorithms can and should be developed to exploit this fact.* There are two caveats to note. First, diagonalization steps that neglect the virtual-virtual part of K will not be identical with steps that include it. Second, the virtual eigenvalues will not be exactly reproduced, and thus a single full K-build is required for post-SCF purposes such as double hybrid density functional theory, or the random phase approximation. We note that the Eq. 8 suggests that all other components of the Fock matrix, not just the exchange matrix, may be utilized in half-transformed form; however, this would likely be a much more drastic approximation, as the virtual-virtual block of the Fock matrix is qualitatively correct when only the virtual-virtual contribution from K is neglected. Assessment of this more aggressive approach is beyond the scope of this work.

The compressed K-matrix, $K_{i\nu}$ can be employed to build an AO basis quantity as follows. We partition the exchange operator into terms involving projectors onto the occupied subspace (the projector is $\hat{\rho}$ of course) and the virtual subspace (the projector is $\hat{q} = (1 - \hat{\rho})$):

$$\hat{K} = (\hat{\rho} + \hat{q})\hat{K}(\hat{\rho} + \hat{q}) \tag{9}$$
$$= \hat{\rho}\hat{K}\hat{\rho} + \hat{q}\hat{K}\hat{\rho} + \hat{\rho}\hat{K}\hat{q} + \hat{q}\hat{K}\hat{q} \tag{10}$$

We neglect the final term in Eq. (10) and recombine the projection operators in the remaining terms to obtain a modified exact exchange operator:

$$\hat{K}' = \hat{\rho}\hat{K} + \hat{K}\hat{\rho} - \hat{\rho}\hat{K}\hat{\rho} \tag{11}$$

Transforming this expression into the AO basis yields our final working expression for $\hat{K}'$:

$$K'_{\mu\nu} = \sum_{\lambda i} S_{\mu\lambda} c_{\lambda i} K_{i\nu} + \sum_{i\lambda} K_{\mu i} c^*_{\lambda i} S_{\lambda\nu}$$
$$- \sum_{ij}\sum_{\lambda\sigma} S_{\mu\lambda} c_{\lambda i} K_{ij} c^*_{\sigma j} S_{\sigma\nu} \tag{12}$$

The second term is clearly just the transpose of the first, and the final term may be obtained easily from the common intermediate, $K_{i\nu}$. The basic idea of computing just $K_{i\nu}$, or, for that matter, just $F_{i\nu}$ is of course extremely simple and undoubtedly has been thought of before. It has been employed in the iterative updates of plane-wave DFT calculations[60], where the full Fock matrix is enormous and thus evaluating a narrow rectangular matrix is far preferable to an enormous square matrix. Indeed, the idea of constructing this economized version of K was mentioned briefly by Aquilante et al.,[61] but we have found no prior or subsequent discussion in the literature. The important point that has not yet been accomplished is to demonstrate how formation of an economized K matrix can be used for computational advantage with AO basis sets.

B. The RI Approximation

One approach to exploiting the gains in efficiency made possible by the economization of the K matrix, is to introduce the RI approximation. This approximation is applied by expanding a given pair of gaussians in terms of an atom-centered gaussian auxiliary basis set, as shown in Eq. (3). The expansion coefficients are typically determined by minimizing the self-repulsion of the residual between the exact gaussian product and its fitted expansion.[44] This formulation yields the following standard formulation of the coefficients in global Coulomb-metric RI:[44]

$$C_Q^{\mu\lambda} = \sum_P (\mu\lambda|P)(P|Q)^{-1} \tag{13}$$

Global Coulomb RI can be shown to be equivalent to the Dunlap functional for the two-electron integral, and thus yields error in the final integral quadratic in the expansion errors of the individual products.[62]

C. The occ-RI-K Algorithm

It is apparent from the above discussion that the problem of the efficient computation of the K matrix now reduces to the problem of computing $K_{i\nu}$. Application of the RI approximation and expressing the density matrix in the MO basis yields the following expression for the $K_{i\nu}$:

$$K_{i\nu} = \sum_{j\lambda\sigma Q} C_Q^{\mu\lambda}(Q|\nu\sigma) c_{\mu i} c_{\lambda j} c_{\sigma j} \tag{14}$$

In the following discussion of performance, we use the following notation for dimension quantities: N (number of primary basis functions), [NB2] (number of significant primary-basis function pairs, which is asymptotically linear in system size), X (number of auxiliary basis functions), and o (number of occupied orbitals). In order to avoid an expensive $O(X^2)$ step, we pretransform the three-center integrals prior to the formation of the Coulomb metric RI fit coefficients. For calculations run with realistic basis sets and thresholds, [NB2] $>> o^2$, and thus pretransforming the integrals renders the $O(X^2)$ step relatively unimportant. The resulting algorithm for the formation of $K_{i\nu}$ is shown in Table. I. With $K_{i\nu}$ in hand, the three terms in Eq. (12) can be formed trivially via fast matrix multiplications.

If the oNX storage requirement becomes prohibitive, the algorithm can be trivially modified to add an additional level of batching over the occupied orbital index j. This will then necessitate recomputation of the AO three-center integrals once per batch of j values. We have constructed such an implementation, but do not find it necessary to use for the timing experiments presented here. In practice, the disk I/O wall time cost is a negligible fraction of the algorithmic runtime.

We can anticipate the performance gains that can be achieved using occ-RI-K by comparing the three quartic steps in Table I against the two quartic steps of the standard RI-K method. In standard RI-K,[40] one forms $B_{\mu j}^Q + = (\mu j|P)(P|Q)^{-1/2}$ which requires oNX^2 operations, followed by $K_{\mu\nu} + = B_{j\mu}^Q B_{j\nu}^Q$ which requires oN^2X operations. The speedup given by occ-RI-K is therefore approximately:

$$s = \frac{N(X+N)}{o(X+2N)} \tag{15}$$

If we (conservatively) assume that $X \approx 2N$, then the approximate speedup is $s = \frac{3}{4}(N/o)$, suggesting that occ-RI-K can yield a speedup over RI-K which is 75% of the compression factor associated with calculating $K_{i\nu}$ rather than $K_{\mu\nu}$. For a basis set of cc-pVQZ size, this suggests an asymptotic speedup of roughly a factor of 15 (presuming the cubic steps are negligible, as in the very large molecule limit). The next step is to obtain real-world timings on the basis of an implementation for finite-size molecules, and compare against other methods for forming the K matrix.

III. RESULTS AND DISCUSSION

All calculations were performed with a development version of Q-Chem 4.2[63]

A. Accuracy

At SCF convergence, the energy errors associated with occ-RI-K are formally equivalent to those of RI-K. However, as RI-K has historically received only minimal accuracy benchmarking, primarily in the context of training auxiliary basis sets,[41] we wish to provide an expanded survey of the accuracy of this approximation. We also wish to provide an in-depth comparison to our recent implementation[42] of the (also sparsely benchmarked)[49,50] PARI-K algorithm. We now demonstrate that occ-RI-K yields superior accuracy to PARI-K, causing negligible errors with respect to unapproximated integral-direct algorithms for a wide variety of energetic properties. Our test suite is based on that of Mardirossian et al.,[34] and consists of a compilation of many separate sets of energetic reference data spanning thermochemistry and intermolecular interactions. Calculations performed at the B3LYP/def2-QZVPPD[8,64–68] level of theory yield the results summarized in Table II). While PARI-K yields impressive accuracy in its own right, the accuracy of occ-RI-K is clearly superior. The errors per electron in the absolute energies of all molecules in the test set are quite acceptable, with a mean of 0.7 cal mol^{-1} electron^{-1} and a maximum of 3.2 cal mol^{-1} electron^{-1}. This is significantly lower than the corresponding mean of 2.0 cal

Step	Operations	Scaling	Memory
At the start of first SCF iteration:			
$\quad$ Calculate $(P\|Q)^{-1}$	X^3	M^3	M^2
For each SCF iteration:			
$\quad$ Loop over atomic batches of auxiliary functions P:			
$\qquad$ Calculate $(\mu\lambda\|P)$ in batch	$[\text{NB2}]\,X$	M^2	M^1
$\qquad (\mu j\|P) = \sum_{\lambda} (\mu\lambda\|P)\, c_j^{\lambda}$	$o\,[\text{NB2}]\,X$	M^3	M^2
$\qquad$ Write $(\mu j\|P)$ to disk.	oNX	M^3	M^2 (M^3 disk)
$\qquad (ij\|P) = \sum_{\mu} (\mu j\|P)\, c_j^{\mu}$	o^2NX	M^4	M^2
$\qquad C_Q^{ij}{+}= \sum_{P} (ij\|P)\,(P\|Q)^{-1}$	o^2X^2	M^4	M^2
$\quad$ Loop over atomic batches of auxiliary functions Q:			
$\qquad$ Read $(\nu j\|Q)$ batch from disk.	oNX	M^3	M^2
$\qquad K_{i\nu}{+}= \sum_{Q} C_Q^{ij}\,(\nu j\|Q)$	o^2NX	M^4	M^2

TABLE I. The occ-RI-K algorithm, omitting the trivial final matrix multiplication steps. The second column gives the operation cost for each step in terms of N (number of primary basis functions), [NB2] (number of significant primary-basis function pairs, which is asymptotically linear in system size), X (number of auxiliary basis functions), and o (number of occupied orbitals). The computational cost scaling is given as a power of the size of the molecule, M, assuming a uniform choice of the AO and auxiliary basis sets. Memory demand is indicated in the same way.

mol^{-1} electron^{-1} and max of 11.8 cal mol^{-1} electron^{-1} for PARI-K. Note that in these tests the Coulomb interactions are evaluated by exact methods.

B. Convergence

At no point in the course of SCF iteration does the occ-RI-K method explicitly construct the exact Fock matrix; rather, we construct a modified Fock matrix that yields the identical orbital gradient and DIIS error vectors, and thus converges to the same result as a calculation using the conventional RI-K method. However, the results of the diagonalization updates are not identical, as these depend upon the unoccupied eigenvalues which are not obtained exactly via occ-RI-K. So while the final result (presuming convergence) is not in doubt, the convergence properties of occ-RI-K are an open question. We empirically demonstrate that convergence using occ-RI-K is essentially identical to that of standard DIIS. Over the entire test set, the average number of iterations to convergence with and without occ-RI-K is identically 9.3 iterations. A histogram of the change in SCF iteration count for all calculations in the test set is shown in Fig. (1). Convergence for the vast majority of jobs is completely unaffected, while 752 jobs converge faster with occ-RI-K compared to 702 that converge slower. No job requires more than six additional iterations to converge. We can thus state definitively that occ-RI-K convergence is well-behaved, and on the basis of these statistics, we can simply assess performance based on timing a single K-build.
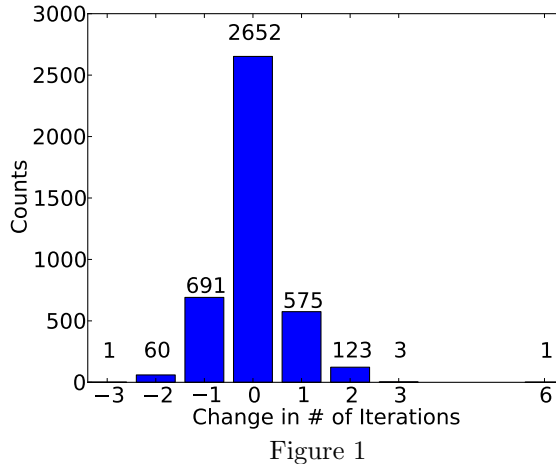

Figure 1

FIG. 1. Histogram of change in number of SCF iterations for the calculations in the test set between occ-RI-K and exact integral-direct K.

C. Timings

We now demonstrate that occ-RI-K outperforms our recently developed PARI-K algorithm and traditional RI-K. The timings for conventional integral evaluation are shown to reiterate the dramatic advantage of these methods for large basis sets. The systems chosen for comparison are n-by-6 hydrogen-terminated graphene sheets of successively higher depth; we prefer these "somewhat

Dataset	occ-RI-K Error	PARI-K Error
A21x12[69]	0.00	0.00
A24[70]	0.00	0.00
AE18[71],	0.01	0.01
AlkAtom19[72],	0.01	0.14
AlkIsomer11[72]	0.00	0.02
AlkIsod14[72]	0.00	0.03
Bauza30[73,74]	0.00	0.00
Butanediol65[75]	0.00	0.00
BzDC215[76]	0.00	0.00
CT20[77]	0.00	0.00
DIE60[78]	0.00	0.00
DS14[79]	0.00	0.01
EA13[80]	0.01	0.01
EIE22[81]	0.00	0.00
FmH2O10[82,83]	0.01	0.16
ACONF[84,85]	0.00	0.01
BHPERI[85,86]	0.00	0.00
CYCONF[85,87]	0.00	0.00
G21EA[85,88]	0.00	0.00
G21IP[85,88]	0.00	0.00
NBPRC[85]	0.00	0.02
WATER27[85,89]	0.00	0.03
NHTBH38[85,90]	0.01	0.02
HTBH38[85,91]	0.00	0.01
BH76RC[85]	0.00	0.00
DBH24[86]	0.00	0.01
H2O6Bind8[83,92]	0.00	0.05
HB15[93]	0.00	0.01
HSG[94,95]	0.00	0.00
HW30[96]	0.00	0.01
HW6Cl[82,83]	0.00	0.04
HW6F[82,83]	0.00	0.04
IP13[80]	0.00	0.00
NBC10[95]	0.01	0.01
NC15[97]	0.00	0.00
Pentane14[98]	0.00	0.01
RG10[99]	0.00	0.00
S22[95,100]	0.00	0.00
S66[101,102],	0.01	0.00
S66x8[101]	0.00	0.00
Shields38[103]	0.01	0.06
Styrene45[104]	0.00	0.01
SW49Rel345[105]	0.01	0.00
SW49Bind345[105]	0.00	0.02
SW49Rel6[105]	0.00	0.01
SW49Bind6[105]	0.00	0.04
TA13[106]	0.00	0.00
BDE99MR[107]	0.01	0.01
BDE99nonMR[107]	0.00	0.01
HAT707nonMR[107]	0.00	0.00
HAT707MR[107]	0.01	0.01
ISOMERIZATION20[107]	0.00	0.00
SN13[107]	0.00	0.00
TAE140nonMR[107]	0.00	0.01
TAE140MR[107]	0.01	0.02
X40[108]	0.00	0.00
XB18[109]	0.00	0.00
XB51[109]	0.00	0.01

TABLE II. occ-RI-K vs. PARI-K Accuracy comparison over a compilation of test sets, using the B3LYP/def2-QZVPPD level of the theory. Data in kcal/mol.
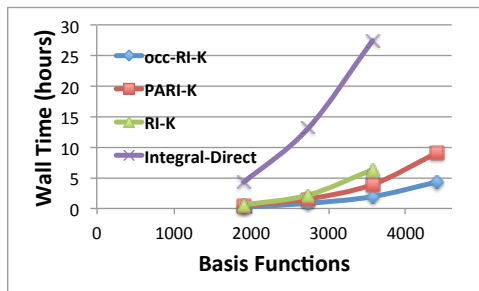

Figure 2

FIG. 2. Timings for the second K build of various depths of n-by-6 graphene sheets in the cc-pVQZ basis set using a single thread. An integral screening threshold of 10^{-10} was used for all calculations.

two-dimensional" benchmarks as their size can be increased in a linear fashion but they yield arguably more realistic AO locality than linear systems. The basis set used was the Dunning cc-pVQZ basis set.[110] The results of this timing experiment are shown in Fig. (2).

For the largest system treated with the conventional integral code and RI-K, occ-RI-K provides a 14× speedup, compared to 7× for PARI and 4× for RI. We note that the difference between this PARI-K speedup and that reported previously[42] for the same system/basis set is likely due to a looser screening threshold (10^{-10} here vs 10^{-12} in previous work), which accelerates the conventional code relative to PARI and occ-RI-K. The use of a different architecture (AMD Opteron 6376 vs. AMD Opteron 2376 previously) could also play a role.

For the largest system, occ-RI-K outperforms PARI-K by a factor of two. For the largest system treated with RI-K, occ-RI-K yields a speedup of 3.3x relative to RI-K. This factor is considerably smaller than our predicted speedup relative to RI-K of about 15 in cc-pVQZ. This reflects the fact that quartic steps do not entirely dominate the calculation for molecules of this size, contrary to what we assumed in the speedup analysis. As a result the speedups over RI-K will increase further with yet larger size molecules. We demonstrate this trend by considering two linear alkanes, $C_{30}H_{62}$ and $C_{70}H_{142}$, in the basis sets cc-pVQZ and cc-pVTZ, respectively. For the smaller alkane in the larger basis set, the computational cost is dominated by the lower scaling steps, as shown in Table III. When the system size is increased and the basis set size is reduced, the higher-scaling MO basis steps become much more significant, illustrating the trend of larger speedups with respect to RI-K with increasing system size.

We shall now demonstrate that occ-RI-K can be applied to quite large systems even in small basis sets, while still outperforming conventional SCF. While occ-RI-K has unfavorable asymptotic scaling compared to conventional SCF, in practice this shortcoming is drastically mitigated by two factors. First, the much higher floating point efficiency of standardized linear algebra libraries relative to AO integral evaluation codes coupled with the

Step	$C_{30}H_{62}$/cc-pVQZ Time (s)	% of Total	$C_{70}H_{142}$/cc-pVTZ Time (s)	% of Total
Calculate $(\mu\lambda|P)$	955	26	1090	8
$(\mu j|P) = \sum_\lambda (\mu\lambda|P)\, c_j^\lambda$	1686	46	4294	33
$(ij|P) = \sum_\mu (\mu j|P)\, c_j^\mu$	174	5	1419	11
$C_Q^{ij} = \sum_P (ij|P)\,(P|Q)^{-1}$	267	7	3483	27
$K_{i\nu} = \sum_{jQ} C_Q^{ij}\,(\nu j|Q)$	224	6	1524	12
Speedup rel. to RI-K	3.6x		5.0x	

TABLE III. Contribution of various steps to overall occ-RI-K walltime for two systems and basis sets. Disk reads/writes are included in overall times for various steps, though their contribution is minor. An integral threshold of 10^{-10} was employed in both calculations. The total number of basis functions were 3510 and 4088, respectively.
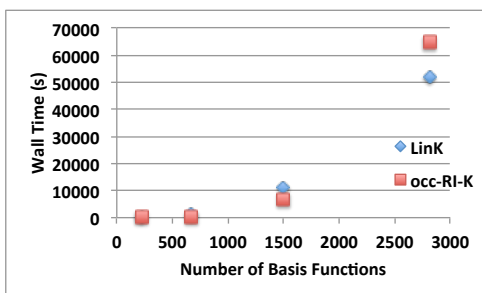

FIG. 3. Timings for the second K build of various size octahedral diamondoids in the cc-pVDZ basis set using a single thread. The points correspond to $n = 2, 3, 4, 5$ in the compound series $C_{(4n^3-n)/3}H_{4n^2}$ used by Strout and Scuseria.[111] An integral screening threshold of 10^{-10} was used for all calculations

replacement of extended primary basis function indices with more compact occupied oribtal indices yields a much lower prefactor for RI methods that is not overcome by lower-scaling methods until very long length scales. Second, the favorable scaling of conventional K construction is very difficult to attain in practice, as both more diffuse basis sets and tighter integral thresholds than those used in timing benchmark studies must often be used in practical applications.

As an illustration of these challenges, we compare the performance of occ-RI-K to that of LinK.[29] An integral threshold of 10^{-10} is often recommended[111,112] to converge SCF energies to 10^{-6}. The systems on which these comparisons are performed are the octahedral diamondoid model systems used by Strout and Scuseria to model SCF scaling in a previous publication,[111] treated using the cc-pVDZ basis set.[110] Due to the absence of a standard exchange auxiliary fitting basis for this basis set, the double zeta RI-MP2 auxiliary basis set of Weigend was used instead.[113] The results are shown in Fig. (3).

While occ-RI-K is eventually surpassed by LinK, the first system in which LinK performs better is $C_{165}H_{100}$.

Thus, even for this very small, compact basis set, a significant size regime exists where occ-RI-K is competitive. We thus assert that occ-RI-K is likely to accelerate a significant subset of everyday quantum chemistry calculations.

IV. CONCLUSIONS

The occ-RI-K method is a new scheme for building the K matrix partially in the MO basis that, combined with the RI approximation, leads to an extremely efficient algorithm that typically outperforms current alternatives. The occ-RI-K method yields extremely accurate energetics (identical to conventional RI-K) and well-behaved convergence (essentially unchanged relative to conventional methods). It yields speedups over conventional integrals and conventional RI-K that increase with the size of the AO basis set. Occ-RI-K thus helps to make larger basis set hybrid DFT calculations more feasible, which is desirable for improved accuracy with modern functionals. Whilst we have not treated the case explicitly here as it is a very simple generalization, we note that range-separated hybrid functionals evaluated using occ-RI-K yield an essentially equal speedup over RI-K. Finally, there are very intriguing prospects for employing the economized K framework in lower scaling (conditionally linear scaling) algorithms, using localized orbitals, which will be valuable for the largest molecules.

V. ACKNOWLEDGEMENTS

[1] P. Hohenberg, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
[4] S. F. Sousa, P. A. Fernandes, and M. J. a. Ramos, J. Phys. Chem. A **111**, 10439 (2007).
[5] E. Rudberg, E. H. Rubensson, and P. Saek, J. Chem. Theory Comput. **7**, 340 (2011).
[6] J. P. Perdew and K. Schmidt, AIP Conference Proceedings **577**, 1 (2001).
[7] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[8] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[9] Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **110**, 13126 (2006).
[10] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2007).
[11] J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).
[12] J. D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106 (2008).
[13] N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys. **16**, 9904 (2014).
[14] C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994).
[15] C. A. White and M. Head-Gordon, J. Chem. Phys. **104**, 2620 (1996).
[16] Y. Shao and M. Head-Gordon, Chem. Phys. Lett. **323**, 425 (2000).
[17] J. Almlöf, K. Faegri Jr., and K. Korsell, J. Comput. Chem. **3**, 385 (1982).
[18] D. Cremer and J. Gauss, J. Comput. Chem. **7**, 274 (1986).
[19] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[20] D. S. Lambrecht, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **123**, 184102 (2005).
[21] D. S. Lambrecht and C. Ochsenfeld, J. Chem. Phys. **123**, 184101 (2005).
[22] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, J. Chem. Phys. **136**, 144107 (2012).
[23] R. Baer and M. Head-Gordon, Phys. Rev. Lett. **79**, 3962 (1997).
[24] M. Benzi, P. Boito, and N. Razouk, SIAM Rev. **55**, 3 (2013).
[25] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[26] P. E. Maslen, C. Ochsenfeld, C. a. White, M. S. Lee, and M. Head-Gordon, J. Phys. Chem. A **102**, 2215 (1998).
[27] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[28] E. Schwegler and M. Challacombe, Theor. Chem. Acc. **104**, 344 (2000).
[29] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[30] N. X. Wang and A. K. Wilson, J. Chem. Phys. **121**, 7632 (2004).
[31] N. X. Wang and A. K. Wilson, Mol. Phys. **103**, 345 (2005).
[32] A. D. Boese, J. M. L. Martin, and N. C. Handy, J. Chem. Phys. **119**, 3005 (2003).
[33] J. D. Chai and M. Head-Gordon, J. Chem. Phys. **131** (2009), 10.1063/1.3244209.
[34] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **142**, 074111 (2015).
[35] M. Sierka, A. Hogekamp, and R. Ahlrichs, J. Chem. Phys. **118**, 9136 (2003).
[36] A. Sodt, J. E. Subotnik, and M. Head-Gordon, J. Chem. Phys. **125**, 194109 (2006).
[37] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[38] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).
[39] F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).
[40] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[41] F. Weigend, J. Comput. Chem. **29**, 167 (2008).
[42] S. F. Manzer, E. Epifanovsky, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 518 (2015).
[43] E. Baerends, D. Ellis, and P. Ros, Chem. Phys. **2**, 41 (1973).
[44] O. Vahtras, J. Almlöf, and M. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).
[45] Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Nat. Acad. Sci. **102**, 6692 (2005).
[46] S. Reine, E. Tellgren, A. Krapp, T. Kjaergaard, T. Helgaker, B. Jansik, S. Host, and P. Salek, J. Chem. Phys. **129**, 104101 (2008).
[47] A. Sodt and M. Head-Gordon, J. Chem. Phys. **128**, 104106 (2008).
[48] D. Mejía-Rodríguez and A. M. Köster, J. Chem. Phys. **141**, 124114 (2014).
[49] P. Merlot, T. Kjæ rgaard, T. Helgaker, R. Lindh, F. Aquilante, S. Reine, and T. B. Pedersen, J. Comput. Chem. **34**, 1486 (2013).
[50] D. S. Hollman, H. F. Schaefer, and E. F. Valeev, J. Chem. Phys. **140**, 064109 (2014).
[51] N. H. F. Beebe and J. Linderberg, Int. J. Quantum. Chem. **12**, 683 (1977).
[52] J. Boström, F. Aquilante, T. B. Pedersen, and R. Lindh, J. Chem. Theory Comput. **9**, 204 (2013).
[53] R. A. Friesner, Chem. Phys. Lett. **116**, 39 (1985).
[54] F. Neese, F. Wennmohs, A. Hansen, and U. Becker, Chem. Phys. **356**, 98 (2009).
[55] S. Kossmann and F. Neese, Chem. Phys. Lett. **481**, 240 (2009).
[56] P. Plessow and F. Weigend, J. Comput. Chem. **33**, 810 (2012).
[57] T. Van Voorhis and M. Head-Gordon, Mol. Phys. **100**, 1713 (2002).
[58] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[59] P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[60] V. Weber, J. Vandevondele, J. Hutter, and A. M. N. Niklasson, J. Chem. Phys. **128**, 084113 (2008).
[61] F. Aquilante, L. Boman, J. Boström, H. Koch, R. Lindh, A. S. D. E. Merás, and T. B. Pedersen, *Linear-Scaling Techniques in Computational Chemistry and Physics*, edited by R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer Netherlands, Dordrecht, 2011) pp. 301–343.
[62] B. Dunlap, THEOCHEM **529**, 37 (2000).
[63] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D.

Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W.-Z. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).

[64] C. Lee, W. Yang, and R. Parr, Phys. Rev. B **37**, 785 (1988).
[65] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[66] F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[67] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[68] D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010).
[69] J. Witte, M. Goldey, J. B. Neaton, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 1481 (2015).
[70] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **9**, 2151 (2013).
[71] S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. Fischer, Phys. Rev. A **47**, 3649 (1993).
[72] A. Karton, D. Gruzman, and J. M. L. Martin, J. Phys. Chem. A **113**, 8434 (2009).
[73] A. Bauzá, I. Alkorta, A. Frontera, and J. Elguero, J. Chem. Theory Comput. **9**, 5201 (2013).
[74] E. R. Johnson and G. A. Dilabio, J. Chem. Theory Comput. **10**, 5436 (2014).
[75] S. Kozuch, S. M. Bachrach, and J. M. L. Martin, J. Phys. Chem. A **118**, 293 (2014).
[76] D. L. Crittenden, J. Phys. Chem. A **113**, 1663 (2009).
[77] S. N. Steinmann, C. Piemontesi, A. Delachat, and C. Corminboeuf, J. Chem. Theory Comput. **8**, 1629 (2012).
[78] L.-J. Yu and A. Karton, Chem. Phys. **441**, 166 (2014).
[79] B. J. Mintz and J. M. Parks, J. Phys. Chem. A **116**, 1086 (2012).
[80] B. J. Lynch, Y. Zhao, and D. G. Truhlar, J. Phys. Chem. A **107**, 1384 (2003).
[81] L.-J. Yu, F. Sarrami, A. Karton, and R. J. OReilly, Mol. Phys. (2014), 10.1080/00268976.2014.986238.
[82] K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107 (2013).
[83] K. U. Lao and J. M. Herbert, J. Phys. Chem. A **119**, 235 (2015).
[84] D. Gruzman, A. Karton, and J. M. L. Martin, J. Phys. Chem. A **113**, 11974 (2009).
[85] L. Goerigk and S. Grimme, J. Chem. Theory Comput. **7**, 291 (2011).
[86] A. Karton, A. Tarnopolsky, J. F. Lamère, G. C. Schatz, and J. M. L. Martin, J. Phys. Chem. A **112**, 12868 (2008).
[87] J. J. Wilke, M. C. Lind, H. F. Schaefer, A. G. Császár, and W. D. Allen, J. Chem. Theory Comput. **5**, 1511 (2009).

[88] L. Curtiss and K. Raghavachari, J. Chem. Phys. **94**, 7221 (1991).
[89] V. S. Bryantsev, M. S. Diallo, A. C. T. Van Duin, and W. A. Goddard, J. Chem. Theory Comput. **5**, 1016 (2009).
[90] Y. Zhao, N. Gonzalez-Garcia, and D. G. Truhlar, J. Phys. Chem. A **109**, 2012 (2005).
[91] Y. Zhao, B. J. Lynch, and D. G. Truhlar, Phys. Chem. Chem. Phys. **7**, 43 (2005).
[92] Y. Chen and H. Li, J. Phys. Chem. A **114**, 11719 (2010).
[93] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **8**, 141 (2012).
[94] J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. Merz, J. Chem. Theory Comput. **7**, 790 (2011).
[95] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).
[96] K. L. Copeland and G. S. Tschumper, J. Chem. Theory Comput. **8**, 1646 (2012).
[97] D. G. A. Smith, P. Jankowski, M. Slawik, H. A. Witek, and K. Patkowski, J. Chem. Theory Comput. **10**, 3140 (2014).
[98] J. M. L. Martin, J. Phys. Chem. A **117**, 3118 (2013).
[99] K. T. Tang and J. P. Toennies, J. Chem. Phys. **118**, 4976 (2003).
[100] P. Jurecka, J. Sponer, J. Cerný, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[101] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[102] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).
[103] B. Temelso, K. A. Archer, and G. C. Shields, J. Phys. Chem. A **115**, 12034 (2011).
[104] A. Karton and J. M. Martin, Mol. Phys. **110**, 2477 (2012).
[105] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas, and M. Head-Gordon, J. Chem. Theory Comput. **9**, 1368 (2013).
[106] P. R. Tentscher and J. S. Arey, J. Chem. Theory Comput. **9**, 1568 (2013).
[107] A. Karton, S. Daon, and J. M. Martin, Chem. Phys. Lett. **510**, 165 (2011).
[108] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **8**, 4285 (2012).
[109] S. Kozuch and J. M. L. Martin, J. Chem. Theory Comput. **9**, 1918 (2013).
[110] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[111] D. L. Strout and G. E. Scuseria, J. Chem. Phys. **102**, 8448 (1995).
[112] F. Neese, *SOME THOUGHTS ON THE SCOPE OF LINEAR SCALING SELF-CONSISTENT FIELD ELECTRONIC STRUCTURE METHODS*, edited by R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer Netherlands, Dordrecht, 2011) pp. 227–261.
[113] F. Weigend, M. Haser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).