PDG: A Composite Method Based on the Resolution of the Identity

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	jp-2021-06186m.R3
Manuscript Type:	Special Issue Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Pham, Buu; Iowa State University of Science and Technology, Chemistry Datta, Dipayan; Iowa State University of Science and Technology, Department of Chemistry and Ames Laboratory Gordon, Mark; Iowa State University, Chemistry

PDG: A Composite Method Based on the Resolution of the Identity

Buu Q. Pham,[1,2] Dipayan Datta,[1,2] Mark S. Gordon[1,2,*]

[1]Ames Laboratory,
[2]Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA

Abstract

The Gaussian-3 (G3) composite approach for thermochemical properties is revisited in the light of enhanced computational efficiency and reduced memory costs by applying the resolution-of-the-identity (RI) approximation for two-electron repulsion integrals (ERIs) to the computationally demanding component methods in the G3 model: the energy and gradient computations via the second-order Møller-Plesset perturbation theory (MP2) and the energy computations using the coupled-cluster singles-doubles method augmented with non-iterative triples corrections [CCSD(T)]. Efficient implementation of the RI-based methods is achieved by employing a hybrid distributed/shared memory model based on MPI and OpenMP. The new variant of the G3 composite approach based on the RI approximation is termed the RI-G3 scheme, or alternatively the PDG method. The accuracy of the new RI-G3/PDG scheme is compared to the "standard" G3 composite approach that employs the memory-expensive four-center ERIs in the MP2 and CCSD(T) calculations. Taking the computation of the heats of formation of the closed-shell molecules in the G3/99 test set as a test case, it is demonstrated that the RI approximation introduces negligible changes to the mean absolute errors relative to the standard G3 model (less than 0.1 kcal/mol), while the standard deviations remain unaltered. The efficiency and memory requirements for the RI-MP2 and RI-CCSD(T) methods are compared to the standard MP2 and CCSD(T) approaches, respectively. The hybrid MPI/OpenMP based RI-MP2 energy plus gradient computation is found to attain a 7.5x speedup over the standard MP2 calculations. For the most demanding CCSD(T) calculations, the application of the RI approximation is found to nearly halve the memory demand, confer about a 4-5x speedup for the CCSD iterations, and reduce the computational time for the compute-intensive triples correction step by several hours.

* E-mail: mark@si.chem.msg.iastate.edu

1. Introduction

High level *ab initio* electronic structure theory (e.g., coupled cluster methods with large basis sets) can provide accurate thermodynamic properties, which are essential in the chemical sciences and industry. However, due to prohibitive computational costs, their applicability, especially with appropriate basis sets, is usually restricted to small molecular systems. Fortunately, several researchers, notably Lipscomb and Schlegel noted that, if one chooses a reasonable starting point (e.g., Hartree-Fock with a double zeta plus polarization basis set), subsequent improvements in the level of theory and in the atomic basis set are nearly independent of each other.[1,2] This recognition led to the concept of "composite methods", with which one can extrapolate to a desired accurate result. For example, define a desired high level (HL) of theory and a desired large basis set (LB) whose combination HL/LB cannot practicably be applied to molecules of interest. Then, define a low level (LL) of theory and a small basis set (SB) whose combination is tractable. If improvement in the level of theory and the improvement in the basis set are independent of each other, then one could approximate a HL/LB calculation as:

$$\mathrm{E[HL/LB]} \cong \mathrm{E[LL/LB]} + \{\mathrm{E[HL/SB]} - \mathrm{E[LL/SB]}\} \qquad (1)$$

The term in curly brackets in Eq. (*1*) is a level of theory correction obtained using the small basis. This correction is then added to the separate calculation using the low level of theory with the desired large basis. This composite approach is the principle for several methods, including the Gn methods of Pople and co-workers,[3] the Wn methods of Martin et al.,[4–6] and the ccCA methods of Wilson and co-workers.[7–11] These various methods differ in the detailed definitions of HL, LL, SB, LB, but the basic idea is the same, and the predicted thermodynamic properties, such as heats of formation, is frequently within "chemical accuracy", variously defined as 1-2 kcal/mol.

The use of composite methods to predict energetic properties of molecules increases the sizes of systems that are amenable to realistic computations, but such calculations are still

limited by the CPU and memory demands of the chosen levels of theory and basis sets. Memory requirements are a particular bottleneck for most high levels of theory. One approach to reduce the computational effort as well as the memory footprint of correlated electronic structure methods is to apply the resolution of the identity (RI) approximation.[12–15] In the RI approximation one employs an auxilary basis set to reduce time-consuming and memory-demanding 4-index 2-electron integrals to 3-index 2-electron integrals. Recently, the authors have derived and implemented in GAMESS[16] computationally efficient second-order perturbation theory (MP2) energies[17] and gradients[18] and coupled cluster with singles, doubles, and non-iterative triples [CCSD(T)] energies[19] within the RI *ansatz*. In this work, the RI-MP2 and RI-CCSD(T) methods in GAMESS are used in a reformulation of the Gaussian-3 (G3) method.[20–23] The accuracy of this new RI-G3 method, also called the PDG method, is essentially the same as that of the original method, while the computational efficiency and memory footprint are greatly improved.

The current work focuses on the modified G3 method,[24] in which HL is CCSD(T), LL is MP2, LB is G3L[20,25] and SB is 6-31G(d).[26,27] The G3 method uses MP2/6-31G(d) geometries and HF/6-31G(d) zero-point energy (ZPE) corrections and predicts heats of formation that are in agreement with experimental values with an average absolute error of ~1-2 kcal/mol.[28] In more detail, in the G3 approach, the geometry of a molecular system is first optimized at the HF/6-31G(d) level of theory followed by Hessian calculations for the ZPE correction. The geometry is further refined at the MP2/6-31G(d) level of theory, which is used for spin-orbit and single point energy calculations at the CCSD(T)/6-31G(d) and MP2/G3L levels of theory.

During the last decade, there have been large changes in computer hardware and programming models. After a period of boosting floating point operations per second (FLOPS) and the memory of individual processors, most computer vendors are now putting multiple processing units together in a shared memory pool to form a compute node, which dramatically increases FLOP counts while avoiding high heat due to operations of individual processors at high frequency. A drawback of this approach is that a computer code based on the popular distributed memory model will not automatically get faster after every cycle of hardware updates. They might even get slower due to the limitation of the memory resource per

compute process and/or reduction of the CPU frequency of individual processing units. Therefore, employing a hybrid programming model, in which shared memory for on-node parallelism and distributed memory for inter-node communication, is a natural approach to boost application performance.[17,29] In combination with the reduction of the memory footprint, enabling node shared memory using lightweight threading processes can potentially maximize local computation and speed up calculations.

Recently, the MPI/OpenMP hybrid distributed/shared memory parallel programming model has been applied to the RI-MP2 energy[17] and gradient[18] and RI-CCSD(T) energy[19] and implemented in the GAMESS[30] suite of electronic structure programs. These new algorithms have been demonstrated to significantly increase the performance and the applicability domain of these electron correlation methods. In addition, the HF energy and gradient methods in GAMESS have also been threaded, thereby attaining significant speedups.[29,31]

In this work, the G3 method for closed-shell molecular systems is restructured to incorporate the RI approximation and the hybrid distributed/shared memory parallel programming models MPI/OpenMP. The accuracy of the resulting MPI/OpenMP RI-G3 method, also called the PDG (Pham-Datta-Gordon or Pretty Darn Good) method, is tested against experimental heats of formation of closed-shell molecules in the G3/99 test set.[32] The computational efficiency of the PDG method is compared with the existing MPI-based G3 method previously implemented in GAMESS.[33] Although the current implementation is limited to closed shell molecules, the extension to open shell species is straightforward and will be accomplished in the future.

2. RI approximation

In the RI approximation using a Coulomb metric,[12–15] 4-index 2-electron repulsion integrals (4-2ERIs), defined in equation (*2*) below

$$(\mu\nu\,|\,\lambda\sigma) = \iint dr_1 dr_2 \phi_\mu^*(r_1)\phi_\nu(r_1) r_{12}^{-1} \phi_\lambda^*(r_2)\phi_\sigma(r_2) \qquad (2)$$

are approximated by the product of the 3-index matrix B

$$(\mu\nu \mid \lambda\sigma) \approx \sum_{P}^{AUX} B_{\mu\nu}^{P} B_{\lambda\sigma}^{P} \tag{3}$$

In equation (*2*), $\{\phi_\mu(r)\}$ are atomic orbital (AO) basis functions. The superscript AUX in equation (*3*) is the number of auxiliary basis functions. The matrix B is formed by contracting the matrix Ω with the 3-index 2-electron repulsion integrals:

$$B_{\mu\nu}^{P} = \sum_{R}^{AUX} \Omega_{PR}(R \mid \mu\nu) \tag{4}$$

The 3-index 2-electron repulsion integrals (3-2ERIs) are formulated from one AO and two auxiliary basis functions $\{\alpha_R(r)\}$ as follows:

$$(R \mid \mu\nu) = \iint dr_1 dr_2 \alpha_R(r_1) r_{12}^{-1} \phi_\mu^*(r_2) \phi_\nu(r_2) \tag{5}$$

The matrix Ω in equation (*4*) is obtained by decomposing the inverse of the matrix V (equation (*6*)), which is a matrix of 2-index 2-electron repulsion integrals defined in equation (*7*).

$$V^{-1} = \Omega^T \Omega \tag{6}$$

$$V_{PQ} = \iint dr_1 dr_2 \alpha_P(r_1) r_{12}^{-1} \alpha_Q(r_2) \tag{7}$$

The superscript T in equation (*6*) stands for the matrix transpose operation. The matrix decomposition can be carried out using the Cholesky[34] or the eigen-decomposition method.

The RI approximation thus requires an auxiliary basis set and introduces a 3-dimensional array B (equations (*3*) and (*4*)). The size of array B is ~N*N*AUX, where N is the number of AOs in the atomic basis set. The number of auxiliary basis functions (AUX) in optimized auxiliary basis sets[15,35] is $\sim\gamma$*N, in which γ is usually ~1-10. Integral storage (if necessary) in the RI approximation is thus $\sim O(N^3)$ with a small prefactor compared with $\sim O(N^4)$ when regular 4-2ERIs are used.

3. MPI/OpenMP RI-G3/PDG method

The family of G3 methods have been well documented in the literature.[23,36,37] In the GAMESS version QCISD(T) is replaced by CCSD(T), as in Ref. 24. All other components remain the same as in the original G3 method. In this paper, unless otherwise noted, the G3 method refers to the GAMESS version.

In the RI-G3 method, the MP2/6-31G(d) gradient, MP2/G3L and CCSD(T)/6-31(d) single point energy calculations are, respectively, substituted by the RI-MP2/6-31G(d)//AUX1 gradient, and the RI-MP2/G3L//AUX2 and CCSD(T)/6-31G(d)//AUX3 single point energy calculations. AUX1, AUX2 and AUX3 are auxiliary basis sets associated with the RI approximation in each step of the G3 calculation. In the next section, the effect of auxiliary basis sets on the accuracy of RI-G3 method will be examined using closed-shell molecules in the G3/99 testset.

Bottlenecks in the RI-G3 method, including the HF/6-31G(d) gradient,[31] RI-MP2/6-31G(d) gradient,[18] RI-CCSD(T)/6-31G(d)[19] and MP2/G3L[17] single point energy calculations have been parallelized using the hybrid distributed/shared memory models based on MPI and OpenMP API. In practice, one MPI rank is created on each compute node or socket, which then spawns a team of threads for the actual computation. Threads are lightweight processes that can be created and destroyed with small overhead. Furthermore, an essential property of threads in a team is that they can naturallly share memory with negligible communication overhead. This subsequently minimizes the memory footprint compared with the distributed model that needs to replicate most data structures. The performance of MPI/OpenMP for the RI-MP2 energy and gradient and the RI-CCSD(T) energy have been implemented and demonstrated previously.[17–19] These methods are interfaced with the G3 driver for the MPI/OpenMP RI-G3 method, whose accuracy and performance are discussed in next sections.

4. Results and discussion

4.1 RI-G3 accuracy

An important aspect of the current study is to assess how the accuracy of the G3 composite approach is affected when using the RI approximation for 4-2ERIs. For a thorough assessment, three independent sets of calculations were performed, which applied the RI approximation to different components of the G3 method. The first variant is termed the G3(RIMP2) method, which substitutes the standard MP2/6-31G(d) gradient and MP2/G3L energy with the RI-MP2/6-31G(d)//AUX1 gradient and RI-MP2/G3L//AUX2 energy, while still using standard CCSD(T) energies. The second variant substitutes only the standard CCSD(T)/6-31G(d) energy with the RI-CCSD(T)/6-31G(d)//AUX3 energy and is termed G3(RICC). The third variant introduces the RI approximation to both MP2 and CCSD(T) calculations and is termed the RI-G3 method. All three methods are collectively referred to as PDG. The accuracies of the three RI-based G3 methods were evaluated on the basis of the heats of formation of the closed-shell molecules in the G3/99 testset. In addition, their accuracies were also compared with the accuracy of the G3 method based entirely on standard 4-2ERIs, which is dubbed in the following the standard G3 scheme. All three methods G3(RIMP2), G3(RICC) and RI-G3 have been interfaced with the GAMESS G3 driver.

Figures 1a-c show the heats of formation of closed-shell molecules of the G3/99 test set from experiments; standard G3; G3(RIMP2) with AUX1/AUX2 = cc-pVDZ-RI/cc-pVTZ-RI; G3(RICC) with AUX3 = cc-pVDZ-RI; and full RI-G3 with AUX1/AUX2/AUX3 to be cc-pVDZ-RI/cc-pVTZ-RI/cc-pVDZ-RI. Full details of the heats of formation for other auxiliary basis set combinations are listed in Table S1 in the Supporting Information.

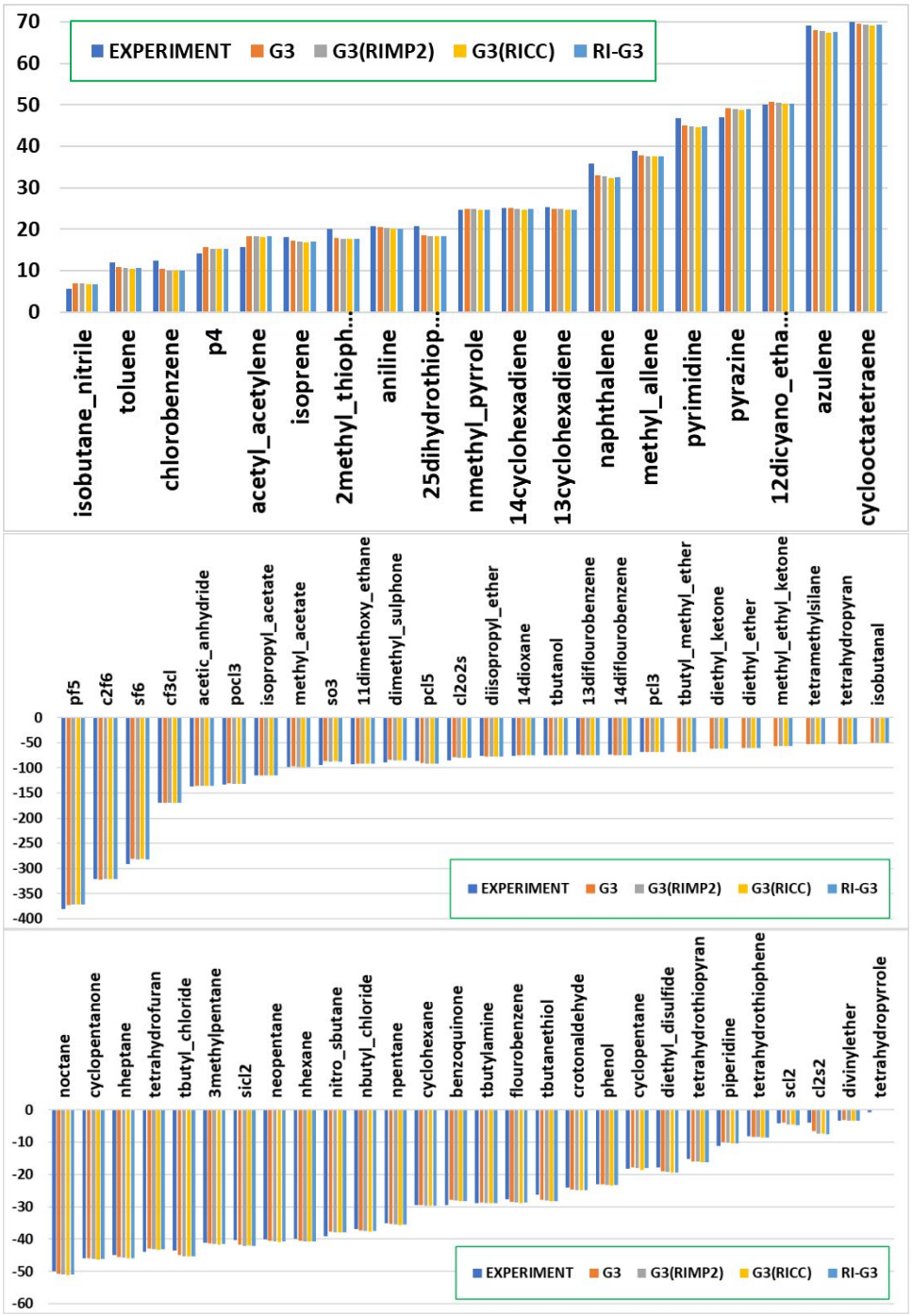

Figure 1. Heats of formation of closed-shell molecules in the G3/99 testset from experimental measurement, standard G3, G3(RIMP2), G3(RICC) and RI-G3 computation. The cc-pVDZ-RI and cc-pVTZ-RI auxiliary basis sets are used for the RI-MP2 gradients and energy, the cc-pVDZ-RI auxiliary basis set is used for RI-CCSD(T) energy component when the RI approximation is utilized in the composite methods. (a) Molecules with positive heats of formation; (b) the first half of molecules with negative heats of formation; (c) the second half of molecules with negative heats of formation.

The error in the heat of formation (at 298K) for a molecule i in the G3/99 set obtained with various G3 composite approaches are defined in the following equations:

$$\delta_{G3}^{(i)} = E_{\mathrm{G3}}^{(i)} - E_{expt}^{(i)} \tag{8}$$

$$\delta_{G3(RIMP2)}^{(i)} = E_{\mathrm{G3(RIMP2)}}^{(i)} - E_{expt}^{(i)} \tag{9}$$

$$\delta_{G3(RICC)}^{(i)} = E_{\mathrm{G3(RICC)}}^{(i)} - E_{expt}^{(i)} \tag{10}$$

$$\delta_{RIG3}^{(i)} = E_{\mathrm{RIG3}}^{(i)} - E_{expt}^{(i)} \tag{11}$$

In equations (*8*)-(*11*), $E_{G3}^{(i)}$, $E_{G3(\mathrm{RIMP2})}^{(i)}$, $E_{G3(\mathrm{RICC})}^{(i)}$ and $E_{RIG3}^{(i)}$ denote the heats of formation of the molecule i calculated by the standard G3, G3(RIMP2), G3(RICC) and RI-G3 methods, respectively. $E_{expt}^{(i)}$ denotes the experimental result. Statistical measures of the accuracies of the various G3 methods are given in terms of the average absolute errors $\overline{\delta}$ and the standard deviations σ against the experimental values. For the G3(RIMP2) scheme, for example, the average absolute error $\overline{\delta}_{G3(RIMP2)}$ and the standard deviation $\sigma_{G3(RIMP2)}$, are defined as follows

$$\overline{\delta}_{G3(RIMP2)} = \frac{1}{n}\sum_{i}^{n}\left|\delta_{G3(RIMP2)}^{(i)}\right| \tag{12}$$

$$\sigma_{G3(RIMP2)} = \sqrt{\frac{\sum_{i}^{n}\left(\delta_{G3(RIMP2)}^{(i)} - \overline{\delta}_{G3(RIMP2)}\right)^{2}}{n}} \tag{13}$$

where n is the number of closed-shell molecules in the G3/99 set. Table 1 summarizes the average absolute errors and standard deviations for the G3 and RI-based G3 composite schemes.

Table 1. Average absolute error $\overline{\delta}$ and standard deviation (σ) of heats of formation of closed-shell molecules in the G3/99 test set calculated by standard G3, G3(RIMP2), G3(RICC) and the RI-

G3 methods against experimental values. Details of calculated heats of formation are shown in Table S1 of the supporting information. Here, CCD, CCT, ACCT stand for the cc-pvDZ-RI, cc-pVTZ-RI, aug-cc-pVTZ-RI auxiliary basis sets, respectively.

	G3	**G3(RIMP2)**			**G3(RICC)**			**RI-G3**		
AUX1	N/A	CCD	CCT	ACCT	N/A	N/A	N/A	CCD	CCT	ACCT
AUX2	N/A	CCT	CCT	ACCT	N/A	N/A	N/A	CCT	CCT	ACCT
AUX3	N/A	N/A	N/A	N/A	CCD	CCT	ACCT	CCD	CCT	ACCT
$\overline{\boldsymbol{\delta}}$	1.46	1.49	1.49	1.49	1.46	1.33	1.53	1.52	1.52	1.55
$\boldsymbol{\sigma}$	1.88	1.88	1.88	1.88	1.62	1.68	1.95	1.86	1.86	1.95

In G3(RIMP2), the MP2/6-31G(d) gradients and MP2/G3L single-point energies were approximated by the RI-MP2/6-31G(d)//AUX1 gradients and RI-MP2/G3L//AUX2 single-point energies, respectively. The accuracy of the RI approximation to 4-2ERIs depends on the size and the quality auxiliary basis set, which in turn should be chosen in a consistent manner along with the atomic basis set. Two different auxliary basis sets AUX1 and AUX2 were used in order to attain flexibility and higher accuracy in conjuction with the 6-31G(d) and G3L atomic basis sets, respectively. Table 1 indicates that the average absolute error and the standard deviation obtained with the G3(RIMP2) scheme are independent of the choice of the auxiliary basis set, and the average absolute error is only slightly different from that of the standard G3 approach (e.g., 1.49 vs. 1.46 kcal/mol), even though the standard devition remains unaffected (1.88 kcal/mol).

In the G3(MP2,RICC) method, the CCSD(T)/6-31G(d) single point energies were replaced with the RI-CCSD(T)/6-31G(d)//AUX3 energies. The average absolute errors and the standard deviations of the heats of formation obtained with the G3(MP2,RICC) scheme are slightly smaller than those obtained with the G3(RIMP2) scheme. For example, with the smallest cc-pVDZ-RI basis set, $\overline{\delta} = 1.46\ kcal/mol$ and $\sigma = 1.62\ kcal/mol$.

Table 1 indicates that the average absolute error and the standard devition obtained with the G3(RICC) scheme are a bit more sensitive to the choice of the auxliary basis set than the

G3(RIMP2) approach. In particular, both $\overline{\delta}$ and σ become sligtly worse when AUX3 is chosen to be aug-cc-pVTZ-RI. It should be noted that no auxiliary basis set specifically optimized for RI-CC calculations is available in the literature. It is a common practice to employ the auxliliary basis sets optimized for RI-MP2 calculations[38] also in RI-CC calculations. It has been previously reported[39] that these auxiliary basis sets provide reasonable accuracy in RI-CC calculations. However, the current results indicate that the RI-CCSD(T) energies are a bit more sensitive to the choice of the auxiliary basis set than the RI-MP2 energies.

Finally, the accuracy of the RI-G3 scheme is examined, in which the MP2/6-31G(d) gradients as well as MP2/G3L and CCSD(T)/6-31G(d) single-point energies were all replaced with the RI approximation: RI-MP2/6-31G(d)//AUX1 gradients, RI-MP2/G3L//AUX2 and RI-CCSD(T)/6-31G(d)//AUX3 single point energies, respectively. Table 1 indicates that the average absolute error and the standard deviation obtained with the RI-G3 scheme are nearly independent of the choice of the auxiliary basis sets except for slight differences when the aug-cc-pVTZ-RI basis is employed in all RI-MP2 and RI-CC calculations. When compared with the errors obtained with the standard G3 method, one finds that the average absolute error obtained with the RI-G3 scheme is slightly higher, while the standard deviation obtained with the latter is slightly smaller when using the cc-pVDZ-RI or cc-pVTZ-RI auxiliary basis sets. Overall, the errors obtained with the RI-G3 scheme are very close to those obtained with the standard G3 approach. The errors agree to within less than 0.1 kcal/mol, so no significant error is introduced by approximating the standard 4-2ERIs via the RI scheme. Furthermore, as demonstrated in the next section, the RI approximation confers significant memory savings and enhanced parallel efficiency to the G3 composite scheme, in particular, for the most expensive CCSD(T) computations.

In the next section, the computational efficiency of the RI-G3 method will be examined. For its small error as well as low computational cost, the cc-pVDZ-RI/cc-pVTZ-RI/cc-pVDZ-RI combination of AUX1/AUX2/AUX3 will be used to investigate the computational savings of the RI-G3 method.

4.2. Computational efficiency and memory savings

In this section, the computational efficiencies and memory demands of the standard G3 and the RI-G3 composite schemes are compared. Since no RI approximation was applied to HF energy and gradient calculations, only the relative efficiencies of the RI-MP2 energy and gradient calculations and the RI-CCSD(T) energy calculations are assessed. Parallel implementations of both the MP2 and CCSD(T) methods with and without the RI approximation are available in GAMESS. The implementations of the standard MP2 and CCSD(T) methods both use a purely MPI based distributed memory parallel programming model. On the other hand, the recent RI-MP2 and RI-CCSD(T) implementations in GAMESS use a hybrid MPI/OpenMP parallelization model. Therefore, in addition to gaining efficiency from the use of the RI approximation, the recent RI-MP2 and RI-CCSD(T) codes derive additional efficiency over the standard MP2 and CCSD(T) codes from the hybrid MPI/OpenMP parallelization model. The latter is more efficient than the purely MPI based scheme due to the reduced network communication among MPI ranks and the memory demand due to data replication. The efficiencies of MPI/OpenMP and MPI codes are compared by running all calculations on identical computing resources (e.g., the number of nodes and cores per node).

The computations reported in this section were performed on a local group cluster Bolt and on an Iowa State University cluster Nova. The Bolt cluster has Haswell compute nodes, and each node is equipped with two 18-core Intel E5-2699V3 sockets with 2.3 GHz processor frequency, and 128 GB of memory per node. Each compute node of Nova consists of two 18-core Intel Skylake 6140 sockets with 2.3 GHz processor frequency and has 192 GB node memory capacity.

The first set of benchmark calculations was performed on polyaromatic hydrocarbons (PAH) consisting of 11-20 carbon atoms shown in Figure 2. These calculations were performed on the local cluster Bolt using one Haswell node. The GAMESS program runs an equal number of compute processes and data servers on each node. The compute processes perform the actual quantum-chemical computation, while the data servers are engaged in data communications between the compute processes.[40] For purely MPI-based MP2 calculations using standard 4-2ERIs, the number of MPI ranks (this number includes compute processes only) per node was chosen to be the number of physical cores on a CPU socket, i.e., 36. For the MPI/OpenMP based RI-MP2 calculations, on the other hand, one compute process was created on each CPU

socket, which spawned a team of threads equal to the number of physical cores on the socket, i.e., 18. Thus the number of compute processes per node was identical in both calculations. To avoid computational overlap, each MPI compute process was pinned to a socket by setting the environment variable MPI_PIN_DOMAIN=socket or omp.

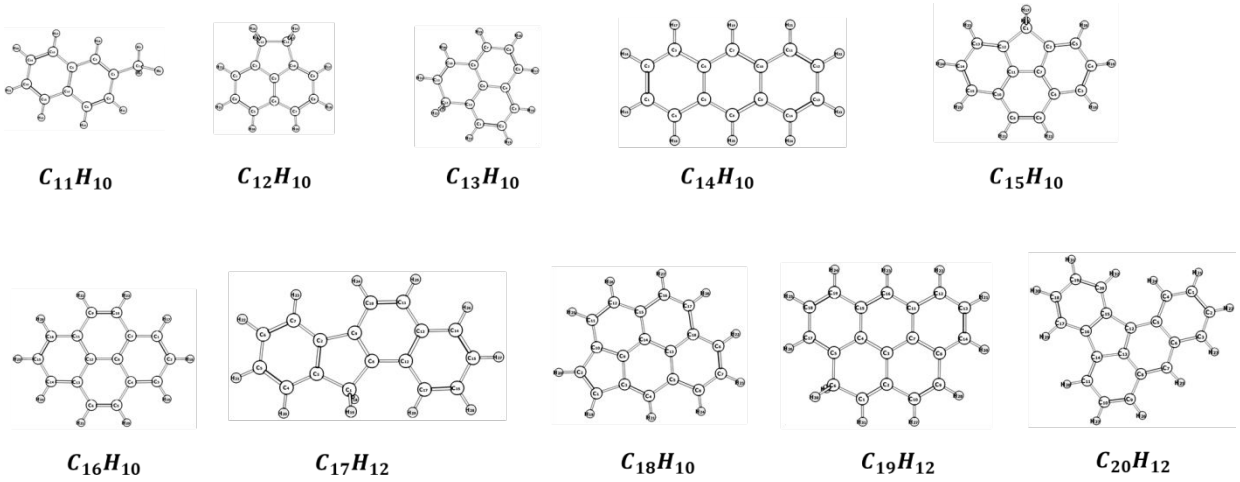

Figure 2. Polyaromatic hydrocarbons from 11-20 carbon atoms

The wall times for the MP2 energy and gradient calculations and the CCSD(T) energy calculations with and without the RI approximation are compared in Table 2. It is evident from Table 2 that the wall times for both the MP2 and CCSD(T) computations significantly decrease when the RI approximation is applied. For instance, for the largest $C_{20}H_{12}$ molecule, the total MP2 wall time for the energy and gradient parts reduces by a factor of ~7.5x. For the same molecule, the CCSD(T) wall time reduces by a factor of ~1.5. These results indicate the wall times for the RI-MP2 energy and gradient calculations are virtually negligible in comparison to the standard MP2 calculations on these systems.

Table 2. Wall times for the MP2 energy plus gradient calculations and RI-CCSD(T) energy calculations on polyaromatic hydrocarbons (Figure 2) with and without the RI approximation.

PAH	G3		RI-G3	
	MP2 (s)	CC (s)	RI-MP2 (s)	RI-CC (s)
$C_{11}H_{10}$	74.1	421.5	8.3	187.7
$C_{12}H_{10}$	95.9	606.0	11.7	314.5
$C_{13}H_{10}$	131.7	914.6	15.1	460.2
$C_{14}H_{10}$	150.7	1259.1	18.4	645.3
$C_{15}H_{10}$	208.0	1773.8	24.3	927.6
$C_{16}H_{10}$	243.4	2403.7	28.8	1274.0
$C_{17}H_{12}$	329.1	3732.9	40.2	2248.3
$C_{18}H_{12}$	380.2	4611.8	46.9	2737.7
$C_{19}H_{12}$	523.2	7130.9	60.7	4295.9
$C_{20}H_{12}$	540.1	8272.9	71.6	5574.7

The second set of benchmark calculations was performed on a peptide chain consisting of eight alanine molecules, $(Alanine)_8$, shown in Figure 3. The purpose of this benchmark is to compare the memory requirements and wall times for the most expensive component of the G3 composite scheme, i.e., the CCSD(T) calculation with and without the RI approximation. These calculations were performed with the 6-31G(d) atomic basis and the cc-pVDZ-RI auxiliary basis sets. Within these basis sets, the CCSD(T) calculations entail a total of 699 atomic basis functions, 3312 auxiliary basis functions, and 116 correlated occupied orbitals.

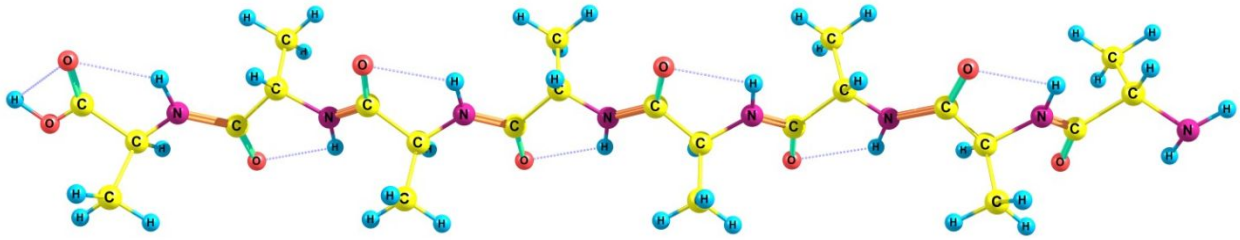

Figure 3. A polypeptide chain of eight alanine units

The CCSD(T) implementation in GAMESS based on standard 4-2ERIs[41] and the recent RI-CCSD(T) implementation[19] use different parallel models. In addition, the underlying equations are different in the two implementations in terms of the factorization schemes applied to the CCSD amplitude residuals. Hence, the sequential algorithms are quite different. Despite these differences, running both programs on an identical number of cores is a reasonable strategy for comparing their parallel efficiencies. The polypeptide calculations were performed on Nova using 4, 8, 12, and 16 compute nodes. For the RI-CCSD(T) calculations, one MPI rank per node was used, which was combined with nine OpenMP threads per MPI rank. The MPI ranks were pinned to CPU sockets as described above, while the OpenMP threads were bound to physical cores on that socket. For the standard CCSD(T) calculations, nine MPI ranks were used on every node. Thus, the total number of compute cores was identical in both calculations for each of 4, 8, 12, and 16 nodes.

The memory requirements for these calculations are compared in Table 3. The MPI based standard CCSD(T) program uses distributed storage for 4-2ERIs throughout the entire calculation.[41] On the other hand, the RI-CCSD(T) program adopts an integral-direct strategy, in which the 4-2ERIs are fully or partially (as three index tensors within loops) assembled on the fly. In the RI-CCSD part of the algorithm, the 4-2ERIs labeled with four virtual indices, i.e., (VV|VV), and those labeled with two occupied (O) and two virtual (V) indices, i.e., (VV|OO), are assembled in a distributed manner on a number of MPI ranks. Thus, in both standard CCSD(T) and RI-CCSD(T) computations, a component of the required memory is of distributed nature, which implies that this memory component scales with the number of compute nodes assigned to the calculation. In Table 3, these components are indicated as *distributed memory per node*, and they decrease proportionately to the increase in the number of nodes. The remaining memory components are constant for a given system size and for a given number of OpenMP threads or MPI ranks assigned per node.

Table 3 indicates that the RI approximation as well as the integral-direct strategy adopted for the integral assembly reduce the total memory requirement per node for the RI-CCSD(T) calculations by a factor of ~1.7 compared to the standard CCSD(T) calculations. Looking closely into individual memory components, one finds that both the shared memory required per node

and the distributed memory requirement are reduced nearly by a factor of two. However, the memory need per core (used by OpenMP threads) increases to some extent for the RI-CCSD(T) calculations. This memory-saving feature of the RI-CCSD(T) program is of great advantage for two reasons, (i) more parallel compute processes (e.g., more OpenMP threads per node) can be assigned for a given system size than the standard CCSD(T) code, thereby increasing the parallel efficiency, and (ii) larger molecular systems and/or larger atomic basis sets can be handled with the RI-CCSD(T) program than the system size that is usually tractable with the standard CCSD(T) program.

Table 3. Memory requirements (in GB) for the RI-CCSD(T) and standard CCSD(T) calculations on $(Alanine)_8$ using the 6-31G(d) atomic basis and the cc-pVDZ-RI auxiliary basis set.

Number of nodes	RI-CCSD(T)				Standard CCSD(T)			
	Memory per core (OpenMP thread)	Shared memory per node	Distributed memory per node	Total memory per node	Memory per core (MPI rank)	Shared memory per node	Distributed memory per node	Total memory per node
4	1.0	53.9	42.2	105.1	0.3	96.0	96.5	195.2
8	1.0	53.9	21.1	84.0	0.3	96.0	48.3	147.0
12	1.0	53.9	14.1	77.0	0.3	96.0	32.2	130.9
16	1.0	53.9	10.5	73.4	0.3	96.0	24.1	122.8

Table 4. Average wall times per CCSD iteration (in min) for the RI-CCSD(T) and the standard CCSD(T) calculations for $(Alanine)_8$ using the 6-31G(d) atomic basis and the cc-pVDZ-RI auxiliary basis.

Number of nodes	RI-CCSD iteration	Standard CCSD iteration	Speedup
4	35.43	131.33	3.7
8	18.87	82.29	4.4
12	15.17	67.41	4.4
16	12.13	60.82	5.0

The wall times per CCSD iteration for $(Alanine)_8$ on different numbers of nodes are compared in Table 4, and the wall times for the non-iterative triples correction are compared in Table 5. The speedup is defined as the time required for a standard CC run divided by an RI-CC

run. Figure 4 compares the individual speedups of the standard and the RI-CCSD(T) codes. Figure 4 indicates that both the standard CCSD(T) and the RI-CCSD(T) calculations speed up when the number of nodes is increased. As a result, the speedup values reported in Tables 4 and 5 remain nearly constant. Since the RI-CCSD iteration shows a better speedup than the standard CCSD iteration, as indicated by Figure 4a, the speedup of the RI-CCSD iteration relative to the standard CCSD iteration slightly increases as the number of nodes increases from 4 to 16. The speedup values reported in Table 4 indicate a significant time saving when using the RI approximation. The improved factorization scheme offered by the use of RI integrals,[19] which permits organizing the CCSD amplitude residuals into a compact quasi-linear form, largely contributes to this time saving. These results also indicate that the RI-CCSD algorithm is more amenable to efficiency enhancement with an increase in the number of parallel compute processes than the standard CCSD(T) algorithm. This stems from the highly memory–economic integral-direct strategy adopted in the RI-CCSD algorithm, which allows storing all requisite data in the node memory and completely bypasses the communication cost associated with fetching data from distributed memory that is done in the standard CCSD algorithm.

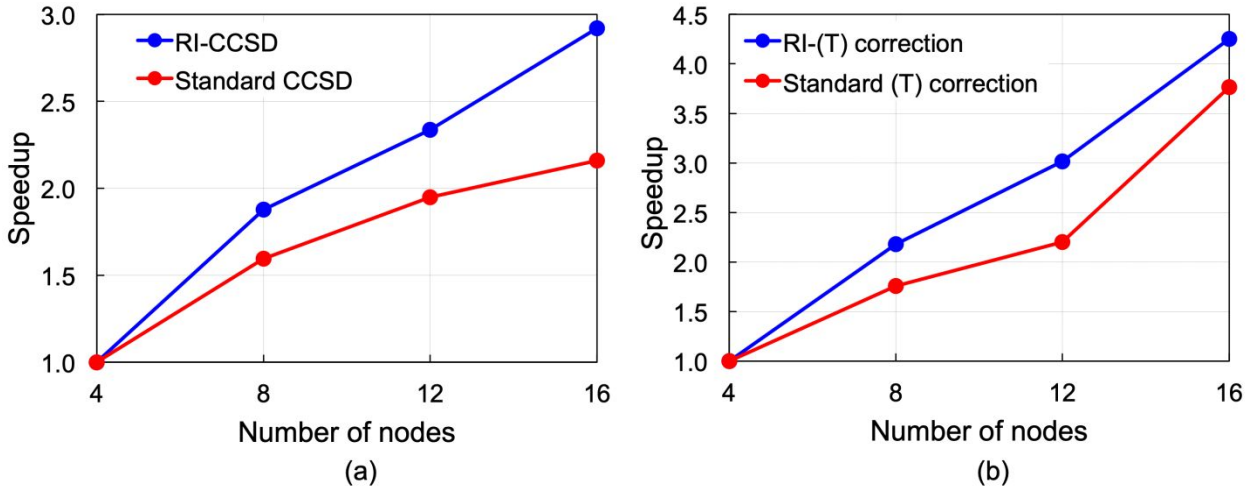

Figure 4. Speedups for (a) CCSD iteration and (b) (T) correction relative to the wall time for the four-node calculation.

One finds from Table 5 that the wall times for the (T) correction reduce by several hours when using the RI-CCSD(T) code in comparison to the standard CCSD(T) code. Furthermore, Figure 4b indicates that the (T) correction part of the RI-CCSD(T) code shows a better speedup than the standard CCSD(T) code. However, the time saving is not as significant in this case as for the CCSD iteration, which is indicated by smaller speedup values for (T) correction reported in Table 5. Unlike in the case of the CCSD iteration, one cannot explore the full advantages of using the RI approximation when designing the (T) correction algorithm. The (T) correction is the most compute-intensive step. Therefore, repeated on-the-fly integral assembly when using the RI approximation would further increase the computational task. For this reason, it is beneficial to preassemble the requisite four-center ERIs prior to computing the (T) correction to the energy. This preassembling in turn reduces the algorithmic difference between the RI-CCSD(T) and the standard CCSD(T) codes. However, it is important to note that the recent RI-CCSD(T) implementation[19] completely bypasses the repeated memory bandwidth bound index permutation operations in sharp contrast to the previously implemented MPI-based standard CCSD(T) code.[41] The observed time savings for the RI-CCSD(T) calculations are thus mainly associated with the efficiency gained by eliminating the index permutation operations.

Table 5. Wall times for the (T) correction (in hours) for the RI-CCSD(T) and the standard CCSD(T) calculations on $(Alanine)_8$ using the 6-31G(d) atomic basis and the cc-pVDZ-RI auxiliary basis.

Number of nodes	RI-CCSD(T)	CCSD(T)	Speedup
4	129.7	132.1	1.0
8	59.5	75.1	1.3
12	43.0	60.0	1.4
16	30.5	35.1	1.2

5. Concluding remarks

The MPI/OpenMP + RI motif can be used to enhance computational efficiency and reduce the memory demand of composite methods with no loss of accuracy. Introducing the RI approximation to the G3 method with a modest combination of auxiliary basis sets resulted in an increase in the average absolute error by less than 0.1 kcal/mol and an unaltered standard deviation in calculated heats of formation for closed-shell molecules of the G3/99 test set. The MPI/OpenMP + RI can speed up the G3 MP2 energy plus gradient component by ~7.5x in calculations of polyaromatic hydrocarbons of 11-20 carbon atoms. The same computational motif also reduces memory demands of the G3 CCSD(T) component by nearly a factor of 2, speeds up CCSD iterations by a factor of ~4.0-5.0, and decreases the wall time for the non-iterative triples correction step by several hours in calculations of a polypeptide chain of eight alanine units. It is anticipated a similar approach based on the RI and on a hybrid OpenMP/MPI approach would benefit other composite methods, such as the Wn and ccCA formulations.

Acknowledgements. This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Division of Chemical, Sciences, Geosciences, and Biological Sciences, under the Computational Chemical Sciences project in Ames Laboratory. Ames Laboratory is operated by Iowa State University under Contract DE-AC02-07CH11338.

Supporting information. This manuscript is accompanied by a supporting information file containing heats of formation calculated by G3, G3(RIMP2), G3(RICC) and RI-G3 using various combination of auxiliary basis sets for closed-shell molecules of the G3/99 test set.

ORCID

Buu Q. Pham: 0000-0001-7684-745X

Dipayan Datta: 0000-0003-0824-0837

Mark S. Gordon: 0000-0001-6893-553X

References

(1) Su, M. Der; Schlegel, H. B. Heats of Formation of Chlorosilanes (SiHmCln) Calculated by Ab Initio Molecular Orbital Methods. *J. Phys. Chem.* **1993**, *97*, 8732–8735.

(2) Schlegel, H. B. Heats of Formation of Fluorine-Substituted Silylenes, Silyl Radicals, and

Silanes. *J. Phys. Chem.* **1984**, *88*, 6254–6258.
(3) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.
(4) Martin, J. M. L.; de Oliveira, G. Towards Standard Methods for Benchmark Quality Ab Initio Thermochemistry—W1 and W2 Theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.
(5) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. W3 Theory: Robust Computational Thermochemistry in the KJ/Mol Accuracy Range. *J. Chem. Phys.* **2004**, *120*, 4129–4141.
(6) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 Theory for Computational Thermochemistry: In Pursuit of Confident Sub-KJ/Mol Predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
(7) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The Correlation Consistent Composite Approach (CcCA): An Alternative to the Gaussian-n Methods. *J. Chem. Phys.* **2006**, *124*, 114104.
(8) DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the Intrinsic Error of the Correlation Consistent Composite Approach (CcCA). *Mol. Phys.* **2009**, *107*, 1107–1121.
(9) DeYonker, N. J.; Williams, T. G.; Imel, A. E.; Cundari, T. R.; Wilson, A. K. Accurate Thermochemistry for Transition Metal Complexes from First-Principles Calculations. *J. Chem. Phys.* **2009**, *131*, 24106.
(10) Prascher, B. P.; Lai, J. D.; Wilson, A. K. The Resolution of the Identity Approximation Applied to the Correlation Consistent Composite Approach. *J. Chem. Phys.* **2009**, *131*, 44130.
(11) Mintz, B.; Williams, T. G.; Howard, L.; Wilson, A. K. Computation of Potential Energy Surfaces with the Multireference Correlation Consistent Composite Approach. *J. Chem. Phys.* **2009**, *130*, 234104.
(12) Weigend, F.; Häser, M. RI-MP2: First Derivatives and Global Consistency. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
(13) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
(14) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral Approximations for LCAO-SCF Calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
(15) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
(16) Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; et al. Recent Developments in the General Atomic and Molecular Electronic Structure System. *J. Chem. Phys.* **2020**, *152*, 154102.
(17) Pham, B. Q.; Gordon, M. S. A Hybrid Distributed/Shared Memory Model For the RI-MP2 Method in the Fragment Molecular Orbital Framework. *J. Chem. Theory Comput.* **2019**, *15*, 5252–5258.
(18) Pham, B. Q.; Gordon, M. S. Development of the FMO/RI-MP2 Fully Analytic Gradient Using a Hybrid-Distributed/Shared Memory Programming Model. *J. Chem. Theory Comput.* **2020**, *16*, 1039–1054.

(19) Datta, D.; Gordon, M. S. A Massively Parallel Implementation of the CCSD(T) Method Using the Resolution-of-the-Identity Approximation and a Hybrid Distributed/Shared Memory Parallelization Model. *J. Chem. Theory Comput.* **2021**, *17*, 4799-4822.
(20) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764–7776.
(21) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gn Theory. *WIREs Comput. Mol. Sci.* **2011**, *1*, 810–825.
(22) Curtiss, L. A.; Raghavachari, K. Gaussian-3 and Related Methods for Accurate Thermochemistry. *Theor. Chem. Acc.* **2002**, *108*, 61–70.
(23) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Baboul, A. G.; Pople, J. A. Gaussian-3 Theory Using Coupled Cluster Energies. *Chem. Phys. Lett.* **1999**, *314*, 101–107.
(24) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Baboul, A. G.; Pople, J. A. Gaussian-3 Theory Using Coupled Cluster Energies. *Chem. Phys. Lett.* **1999**, *314*, 101–107.
(25) Mayhall, N. J.; Raghavachari, K.; Redfern, P. C.; Curtiss, L. A.; Rassolov, V. Toward Accurate Thermochemical Models for Transition Metals: G3Large Basis Sets for Atoms Sc-Zn. *J. Chem. Phys.* **2008**, *128*, 144122.
(26) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
(27) Gordon, M. S. The Isomers of Silacyclopropane. *Chem. Phys. Lett.* **1980**, *76*, 163–168.
(28) Weber, R.; Wilson, A. K. Do Composite Methods Achieve Their Target Accuracy? *Comput. Theor. Chem.* **2015**, *1072*, 58–62.
(29) Mironov, V.; Alexeev, Y.; Keipert, K.; D'mello, M.; Moskovsky, A.; Gordon, M. S. An Efficient MPI/OpenMP Parallelization of the Hartree-Fock Method for the Second Generation of Intel&Reg; Xeon Phi&Trade; Processor. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; SC '17; ACM: New York, NY, USA, 2017; pp 39:1--39:12.
(30) Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; et al. Recent Developments in the General Atomic and Molecular Electronic Structure System. *J. Chem. Phys.* **2020**, *152*, 154102.
(31) Mironov, V.; Alexeev, Y.; Fedorov, D. G. Multithreaded Parallelization of the Energy and Analytic Gradient in the Fragment Molecular Orbital Method. *Int. J. Quantum Chem.* **2019**, *119*, e25937.
(32) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and Density Functional Theories for a Larger Experimental Test Set. *J. Chem. Phys.* **2000**, *112*, 7374–7383.
(33) GAMESS doccumentation https://www.msg.chem.iastate.edu/gamess/documentation.html.
(34) Røeggen, I.; Wisløff-Nilssen, E. On the Beebe-Linderberg Two-Electron Integral Approximation. *Chem. Phys. Lett.* **1986**, *132*, 154–160.
(35) Hättig, C. Optimization of Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations: Core–Valence and Quintuple-ζ Basis Sets for H to Ar and QZVPP Basis Sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

(36) Curtiss, L. A.; Raghavachari, K. Gaussian-3 and Related Methods for Accurate Thermochemistry. *Theor. Chem. Acc.* **2002**, *108*, 61–70.
(37) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gn Theory. *WIREs Comput. Mol. Sci.* **2011**, *1*, 810–825.
(38) Weigend, F.; Köhn, A.; Hättig, C. Efficient Use of the Correlation Consistent Basis Sets in Resolution of the Identity MP2 Calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
(39) DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a T1-Transformed Hamiltonian. *J. Chem. Theory Comput.* **2013**, *9*, 2687–2696.
(40) Fletcher, G. D.; Schmidt, M. W.; Bode, B. M.; Gordon, M. S. The Distributed Data Interface in GAMESS. *Comput. Phys. Commun.* **2000**, *128*, 190–200.
(41) Olson, R. M.; Bentz, J. L.; Kendall, R. A.; Schmidt, M. W.; Gordon, M. S. A Novel Approach to Parallel Coupled Cluster Calculations: Combining Distributed and Shared Memory Techniques for Modern Cluster Based Systems. *J. Chem. Theory Comput.* **2007**, *3*, 1312–1328.

TOC Graphic