JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

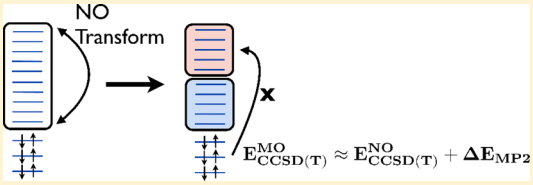
Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals

A. Eugene DePrince, III and C. David Sherrill*

Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States

Supporting Information

ABSTRACT: We assess the accuracy of basis set truncations based on natural orbitals determined by second-order perturbation theory for computing noncovalent interaction energies with coupled cluster through perturbative triples [CCSD(T)]. We consider two methods for truncation: (i) the usual frozen natural orbital approach (FNO) in which the basis set truncation occurs before the iterative CCSD computation [FNO CCSD(T)] and (ii) an approach in which the truncation occurs only for the perturbative triples contribution [CCSD+FNO(T)]. The errors incurred are comparable for both methods and are small enough for the methods to be used for benchmark-quality studies of noncovalent interactions. For the FNO CCSD(T) method with a modest natural orbital occupation tolerance of 10^{-5}, the mean absolute error in the interaction energies for the S22 data set in an aug-cc-pVDZ basis set is only 0.012 kcal mol^{-1} versus canonical CCSD(T) values. The same method exhibits a mean absolute error of 0.020 kcal mol^{-1} for the S11 data set in the aug-cc-pVTZ basis set versus canonical CCSD(T) values.



1. INTRODUCTION

For small chemical systems, the most accurate yet practical description of its electronic structure is usually obtained from coupled cluster (CC) methods,[1] specifically, CC with single, double, and perturbative triple excitations [CCSD(T)], which has come to be known as the "gold standard" in quantum chemistry.[2] Unfortunately, the unfavorable computational scaling of CCSD(T) prevents its routine application to many larger systems of modern chemical and biological interest. A particularly challenging problem in quantum chemistry is the accurate description of noncovalent interactions that arise in DNA base stacking, protein folding, or protein−ligand binding. The usual strategy in computing high-quality interaction energies is a focal point analysis[3,4] combining second-order perturbation theory (MP2) at the complete basis set limit (CBS) with a correlation correction, $\delta^{CCSD(T)}_{MP2}$, derived from CCSD(T) and MP2 in a modest basis set:

$$E^{CBS}_{CCSD(T)} \approx E^{CBS}_{MP2} + \delta^{CCSD(T)}_{MP2} \quad (1)$$

$$\delta^{CCSD(T)}_{MP2} = E_{CCSD(T)} - E_{MP2} \quad (2)$$

High-accuracy benchmark data sets, which are fundamental in assessing the utility of novel theories and approximations, are constructed in this way with $\delta^{CCSD(T)}_{MP2}$ computed in a basis of at least double-ζ-quality, augmented by diffuse functions. When possible, $\delta^{CCSD(T)}_{MP2}$ is more accurately determined in a basis set of triple-ζ quality,[5−8] again augmented by diffuse functions.

The improved accuracy of many recent approximate methods[9−13] necessitates benchmarks of ever increasing quality, but obtaining $\delta^{CCSD(T)}_{MP2}$ in larger basis sets is often impractical due to the steep fourth-power scaling of CCSD(T) with respect to the size of the space of unoccupied (virtual) orbitals. One possible solution is to adopt explicitly correlated CCSD(T) techniques, which typically achieve high accuracy with only augmented double-ζ basis sets. The popular F12a and F12b[14,15] approximations have been applied to the S22[16] test set of van der Waals dimers. Marchetti and Werner demonstrated a mean absolute error (MAE) of less than 0.2 kcal mol^{-1} for CCSD(T**)-F12a/aug-cc-pVDZ. A more recent study[7] testing against the revised S22B interaction energies[6] finds MAEs of less than 0.1 kcal mol^{-1} for CCSD(T**)-F12a and CCSD(T**)-F12b in an aug-cc-pVDZ basis set. For a given complex, F12a and F12b approximations tend to perform differently depending on the binding type; by applying a binding-motif-dependent switching function to mix the F12a and F12b energies, the dispersion-weighted (DW) CCSD(T**)-F12 method reduces the MAE considerably.[7] Despite these successes, questions remain regarding the utility of CBS-extrapolated CCSD(T**)-F12a/F12b interaction energies and the quality of DW-CCSD(T**)-F12 energies for systems that differ from the training set, S22B.

A second avenue to higher-quality benchmarks would be to obtain $\delta^{CCSD(T)}_{MP2}$ in larger basis sets that are pruned to remove high-energy, unimportant virtual orbitals, thereby making such CCSD(T) computations possible. However, the dynamical correlation energy converges slowly within the virtual space; simply discarding high-energy canonical orbitals can lead to very large errors. Several other orbital representations have emerged as compact alternatives to canonical Hartree−Fock

Received: September 7, 2012
Published: November 26, 2012

   dx.doi.org/10.1021/ct300780u | J. Chem. Theory Comput. 2013, 9, 293−299

virtual orbitals. The simplest procedure truncates the virtual space according to the population of the MP2 natural orbitals, which are the eigenfunctions of the first-order one-particle density matrix (OPDM).[17−20] Within the context of CC theory, these frozen natural orbital (FNO) techniques have been shown to allow a 20%−60% reduction in the virtual space for a triple-ζ-quality basis set while introducing errors in CCSD(T) geometric parameters on the order of only a few tenths of a picometer.[20] Similar FNO methodologies have been applied to the iterative CCSDT and noniterative CCSDT(Q) equations[21] and to a variety of equation-of-motion (EOM)-CC methods.[22] They have also been useful in earlier work on reducing the cost of higher-order excitations in configuration interaction (CI).[23−25] Recent work has also revived interest in pair natural orbital techniques for electron correlation methods.[26,27] For a more detailed history of the use of FNOs in CC and CI methods, the reader is directed to ref 22.

A more sophisticated approach to obtain a compact virtual orbital space involves the optimization of the second-order Hylleraas functional,[28] which results in the so-called optimized virtual orbital space (OVOS).[29−31] The iterative procedure to obtain the OVOS is very closely related to the FNO procedure described above; FNOs can be obtained from the first iteration of an OVOS procedure before any optimization. As with FNOs, OVOS truncations are known to reliably reduce the cost of CCSD(T) computations without introducing significant error.[21]

Several studies have investigated the utility of OVOS and FNO representations of the virtual orbital space when treating noncovalent interactions. Pitonak et al.[32] have used OVOS CCSD(T) methods to report the interaction energy for several configurations of the benzene dimer up to an aug-cc-pVQZ basis set. OVOS CCSD(T) interaction energies agree with available quadratic configuration interaction [QCISD(T)] energies[8] to 0.03 kcal mol^{-1} in an aug-cc-pVQZ basis set. Although CCSD(T) and QCISD(T) are not strictly comparable, they happen to agree within 0.004 kcal mol^{-1} for the benzene dimer in the smaller aug-cc-pVTZ basis,[8] and thus the very good agreement between OVOS CCSD(T) and QCISD(T) for the aug-cc-pVQZ basis suggests that the OVOS approximation introduces very little error for the benzene dimer interaction energy. Dedíková et al.[33] also examined the applicability of the OVOS CCSD(T) method to the computation of noncovalent interaction energies. For four small hydrogen-bonded and stacked dimers, OVOS techniques were used to truncate the virtual space by as much as 50% while introducing errors in the interaction energy no larger than 0.1 kcal mol^{-1}. FNOs have been utilized within the context of symmetry adapted perturbation theory (SAPT) studies of noncovalent interactions,[34] where they significantly reduce the computational cost of the triples contribution to dispersion. For the S22 test set described by an aug-cc-pVDZ basis set, the SAPT FNO techniques allow for a 50% reduction in the size of the virtual space utilized by the triples contribution, leading to speedups of 15−20 times for that term. By scaling the triples contribution, the errors over the entire test set are on the order of only a few hundreths of a kilocalorie per mole. Finally, Kraus et al.[35] developed a method to control the accuracy of counterpoise-corrected interaction energies within the context of the OVOS CCSD(T) method. However, this study limited the application of OVOS truncations to monomer computations in dimer basis sets, as would be relevant for counterpoise computations.[36] For large molecules with low (C_1) symmetry, the cost of counterpoise-corrected interaction energies is dominated by the dimer computation; hence, the utility of FNO or OVOS techniques must be determined for the dimer computations also.

In this paper, we perform systematic studies of the accuracy of two approximate CCSD(T) methods based on frozen natural orbitals by considering 22 van der Waals dimers. We have chosen to use FNOs (as opposed to OVOS orbitals) in our basis set truncation due to their conceptual simplicity and ease of implementation. Furthermore, it has been demonstrated that the solutions obtained when utilizing FNO and OVOS truncations differ only slightly.[17,21] The first method studied herein is what one might intuitively associate with FNO CCSD(T). That is, the basis set is truncated before the iterative CCSD portion of the algorithm. Second, we isolate the effects of basis set truncation on the (T) contribution to the energy by solving the CCSD equations in the full basis set and exploiting the FNO truncation techniques only for the perturbative (T) contribution. For very large systems in large basis sets, the evaluation of (T) should dominate the computation, and thus the cost of both methods should be comparable. In practice, however, for the systems often studied to produce benchmark-quality results for noncovalent interactions, a significant amount of time is usually spent in the iterative solution to the CCSD equations, so the first approach, if sufficiently accurate, will be more desirable. Basis set truncation before the CCSD iterations has the added benefit that the two-electron integrals need not be available for the full NO basis, reducing the cost of (i) the AO−NO integral transformation, (ii) any out-of-core sorting of the two-electron integrals prior to the CC procedure, and (iii) storage of the largest (v^4) block of integrals, either on disk or in the core.

2. METHODS

2.1. Basis Set Truncations. The basis set truncations utilized herein are based on the natural orbitals of second-order perturbation theory (MP2). For a closed-shell restricted Hartree−Fock reference, the virtual−virtual block of the unrelaxed MP2 one-particle density matrix (OPDM) is given by

$$\gamma_{ab} = 2 \sum_{ijc} \frac{[2(ialjc) - (iclja)](ibljc)}{(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_c)(\varepsilon_i + \varepsilon_j - \varepsilon_b - \varepsilon_c)} \tag{3}$$

where i and j correspond to spatial orbitals that are occupied in the reference function, and a, b, and c correspond to spatial orbitals that are unoccupied in the reference function. The symbol ε_i represents the energy of orbital i, and $(ialjc)$ etc. represent two-electron integrals in chemists' notation. This block of the OPDM is diagonalized to obtain the virtual MP2 natural orbitals. The virtual natural orbital space is truncated according to the eigenvalues of the MP2 OPDM which correspond to the occupations of the NOs. Those orbitals with an occupation below some threshold are discarded. The Fock matrix is transformed to this truncated NO basis and then diagonalized to obtain semicanonical orbitals. These semicanonical orbitals are desirable for two reasons: (i) the standard formulas for the (T) contribution to the correlation energy assume that the Fock matrix is diagonal, and (ii) if the CCSD procedure is performed in the truncated NO basis, the convergence of the equations is fastest when the Fock matrix is diagonal.

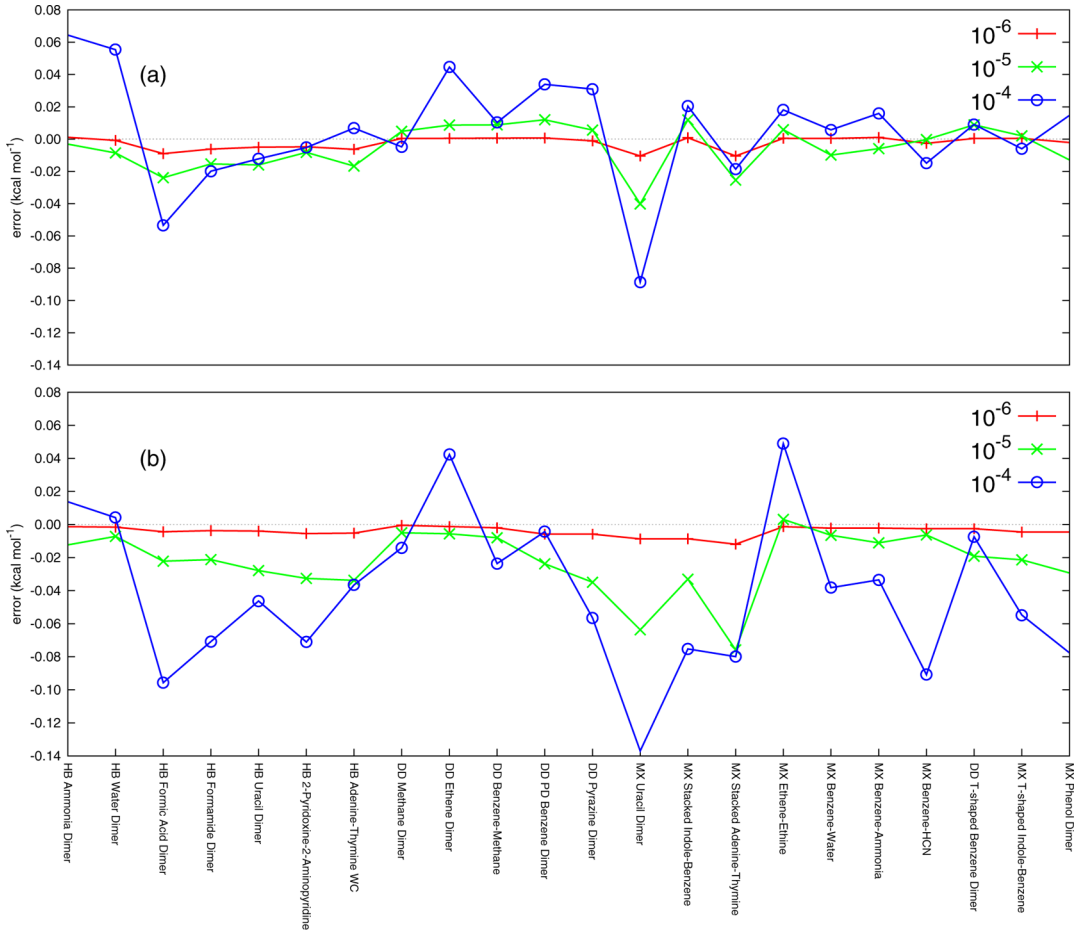


Figure 1. Errors in $\delta_{MP2}^{CCSD(T)}$ for (a) FNO CCSD(T) and (b) CCSD+FNO(T) methods relative to canonical CCSD(T) in the aug-cc-pVDZ basis set. All errors are given in kcal mol^{-1}.

For the FNO CCSD(T) method, the truncated NO basis is determined before any coupled-cluster computations are performed. The NOs are expressed in terms of atomic orbitals, and we perform the AO/NO transformation for all classes of integrals required by CCSD(T). Because the weakly occupied NOs are discarded, the FNO CCSD(T) energy exhibits some error compared to the canonical CCSD(T) result. The largest contribution to the missing correlation energy is approximated at second order, and we define a correction to the FNO CCSD(T) energy as the difference between the MP2 correlation energies in the full MO and truncated NO basis sets:

$$\Delta E_{MP2} = E_{MP2}^{MO} - E_{MP2}^{NO} \tag{4}$$

Thus, the correlation energy at the CCSD(T) level in the full basis set can be approximated as the sum of the correlation energy in the truncated basis plus this correction:

$$E_{CCSD(T)}^{MO} \approx E_{CCSD}^{NO} + E_{(T)}^{NO} + \Delta E_{MP2} \tag{5}$$

The procedure for the CCSD+FNO(T) method is slightly different but involves fewer approximations. The usual AO to MO integral transformation is performed before the CCSD iterations, and the CCSD equations are solved in the full MO basis. Upon convergence, the MP2 NOs are determined as outlined above, and all tensors required for the evaluation of the $E_{(T)}$ are transformed from the MO to NO basis. The required tensors are the T_2 and T_1 amplitudes, the two-electron integrals of the type $(oolov)$, $(ovlov)$, and $(ovlvv)$, and the orbital energies. The symbols o and v represent the occupied and virtual spaces, respectively. The $E_{(T)}$ contribution is evaluated in the truncated NO basis, and the total CCSD(T) energy is defined as

$$E_{CCSD(T)}^{MO} \approx E_{CCSD}^{MO} + E_{(T)}^{NO} \tag{6}$$

Note that eq 6 contains no correction for the basis set truncation. In principle, the value of $E_{(T)}^{MO}$ could be approximated by scaling $E_{(T)}^{NO}$ in a manner similar to that proposed in ref 34, but for our supermolecular computations, the scaled energies differed only slightly from the original values, and the corresponding interaction energies were not necessarily more accurate. Hence, the unscaled energy may be more reliable.

2.2. Computational Details. The frozen natural orbital procedures outlined above were implemented in the PSI4 electronic structure package,[37] which allows users to interface custom "plugins" to various libraries and solvers (e.g., SCF and four-index transformation routines). The present work uses a completely new CCSD(T) plugin interfaced to PSI4 in this way. All CCSD(T) computations were performed within the frozen core approximation. Both the FNO CCSD(T) and CCSD+FNO(T) methods were applied to the S22 data set in the aug-cc-pVDZ basis set and the subset of the 11 smallest systems (S11) in the aug-cc-pVTZ basis set.

dx.doi.org/10.1021/ct300780u | J. Chem. Theory Comput. 2013, 9, 293−299

Table 1. Mean Absolute Errors (kcal mol^{-1}) for Natural-Orbital-Based Estimates of the Coupled-Cluster Correction, $\delta_{MP2}^{CCSD(T)}$, versus Canonical CCSD(T) for the S22 Set Described by an aug-cc-pVDZ Basis Set[a]

	CCSD + FNO(T)			FNO CCSD(T)		
	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}
H-bonded	0.004 (0.005)	0.022 (0.034)	0.048 (0.096)	0.005 (0.009)	0.013 (0.024)	0.031 (0.064)
dispersion	0.006 (0.012)	0.031 (0.076)	0.054 (0.137)	0.003 (0.011)	0.015 (0.040)	0.032 (0.089)
mixed	0.003 (0.004)	0.014 (0.029)	0.050 (0.091)	0.001 (0.003)	0.007 (0.013)	0.012 (0.018)
total	0.004 (0.012)	0.023 (0.076)	0.051 (0.137)	0.003 (0.011)	0.012 (0.040)	0.025 (0.089)

[a]Errors within the FNO approximation are presented using different thresholds for the occupation numbers of the retained natural orbitals. Maximum errors are given in parentheses.

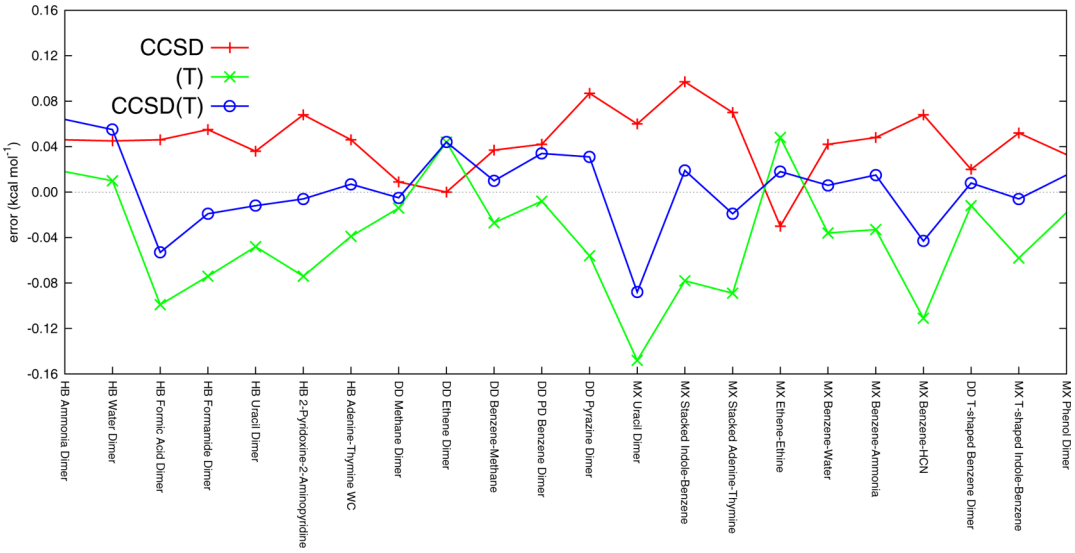


Figure 2. Errors in each component of $\delta_{MP2}^{CCSD(T)}$ for the FNO CCSD(T) method with an occupation tolerance of 10^{-4} relative to canonical CCSD(T) in the aug-cc-pVDZ basis set. All errors are given in kcal mol^{-1}.

3. RESULTS AND DISCUSSION

The FNO CCSD(T) and CCSD+FNO(T) methods were used to compute coupled-cluster corrections, $\delta_{MP2}^{CCSD(T)}$, eq 2, for the S22 set of molecules described by an aug-cc-pVDZ basis set. The errors in the approximate $\delta_{MP2}^{CCSD(T)}$ values compared to canonical CCSD(T) values using the full MO basis are presented in Figure 1. Dimers are labeled in the figure with HB, DD, or MX to denote hydrogen-bonded, dispersion-dominated, or mixed binding, respectively, according to SAPT2+3 analysis.[38] We present results using a wide range of orbital-occupancy tolerances for determining which virtual NOs can safely be discarded. The tightest tolerance used was 10^{-6}, which means that all virtual orbitals with occupancies less than 10^{-6} were discarded. Our most aggressive tolerance is 10^{-4}.

The errors incurred by the FNO approximations are negligible relative to the magnitude of $\delta_{MP2}^{CCSD(T)}$ for both schemes for even the most aggressive tolerance, 10^{-4} (a plot of the aug-cc-pVDZ delta corrections themselves can be found in Figure S1 of the Supporting Information). For example, the maximum error for the two schemes with a tolerance of 10^{-4} is found for the stacked uracil dimer: 0.089 [FNO CCSD(T)] and 0.137 [CCSD + FNO(T)] kcal mol^{-1}. These errors are only 7% and 10% of the canonical $\delta_{MP2}^{CCSD(T)}$ value of 1.262 kcal mol^{-1}. The MAE and maximum errors for each orbital occupation tolerance are given in Table 1. A tolerance of 10^{-6} yields essentially exact results, with MAEs of 0.003 and 0.004 and maximum errors of only 0.011 and 0.012 kcal mol^{-1} in the FNO CCSD(T) and CCSD + FNO(T) schemes, respectively. The MAEs for these two methods with a threshold of 10^{-5} across the whole S22 data set are 0.012 and 0.023 kcal mol^{-1}, which are quite reasonable for reliable high-accuracy studies. Importantly, the MAE and maximum errors do not depend heavily on the binding motif, implying that the FNO methods considered here should remain effective for interaction energies of other types of van der Waals dimers that may differ significantly from those found in the S22 set.

Perhaps the most interesting result here is the relative performance of FNO CCSD(T) and CCSD + FNO(T). One would suspect that, because the CCSD + FNO(T) method involves no approximations in the solution of the CCSD equations, the method should always outperform FNO CCSD(T). However, for each threshold chosen in Figure 1 and Table 1, FNO CCSD(T) consistently outperforms CCSD + FNO(T). Figure 2 provides the errors in each component of the FNO CCSD(T) method with an aggressive occupation threshold of 10^{-4}. We can see that the excellent performance of the method is due to a fortuitous cancellation of errors in the CCSD and (T) components of the energy. Unfortunately, there is no guarantee that such cancellations will occur for all systems in all basis sets.

Figure 3 illustrates the errors in the FNO CCSD(T) $\delta_{MP2}^{CCSD(T)}$ as compared to canonical CCSD(T) computed for the subset of the smallest 11 systems in the S22 set represented by an aug-cc-pVTZ basis set. As was observed in the aug-cc-pVDZ basis, $\delta_{MP2}^{CCSD(T)}$ is nearly perfectly represented in both FNO schemes when using a very tight, 10^{-6}, occupation threshold. Larger

dx.doi.org/10.1021/ct300780u | J. Chem. Theory Comput. 2013, 9, 293−299

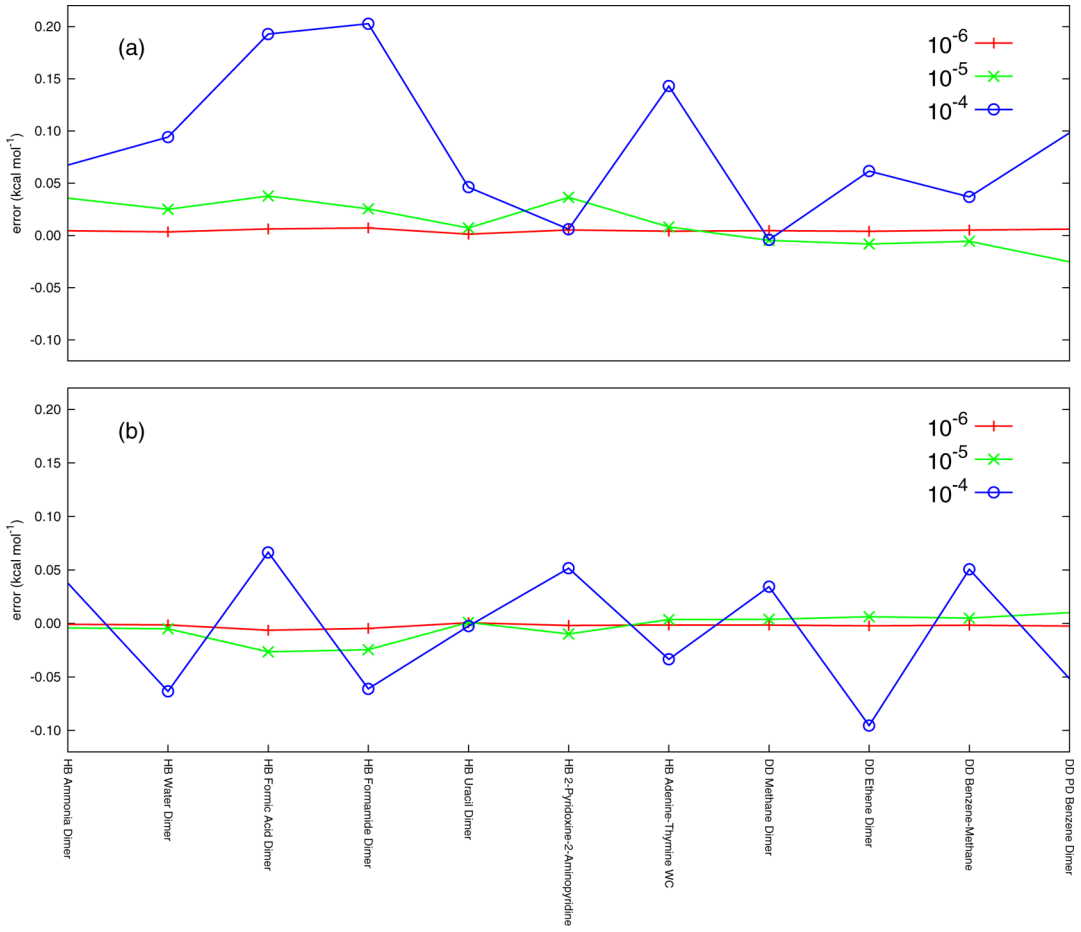


Figure 3. Errors in $\delta_{MP2}^{CCSD(T)}$ for (a) FNO CCSD(T) and (b) CCSD+FNO(T) methods relative to canonical CCSD(T) in the aug-cc-pVTZ basis set. All errors are given in kcal mol^{-1}.

Table 2. Mean Absolute Errors (kcal mol^{-1}) for Natural-Orbital-Based Estimates of the Coupled-Cluster Correction, $\delta_{MP2}^{CCSD(T)}$, versus Canonical CCSD(T) for the 11 Smallest Members of the S22 Set Described by an aug-cc-pVTZ Basis Set[a]

	CCSD + FNO(T)			FNO CCSD(T)		
	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}
H-bonded	0.003 (0.006)	0.015 (0.027)	0.057 (0.067)	0.005 (0.007)	0.031 (0.038)	0.139 (0.203)
dispersion	0.001 (0.002)	0.005 (0.010)	0.029 (0.052)	0.004 (0.005)	0.017 (0.037)	0.065 (0.143)
mixed	0.002 (0.002)	0.006 (0.010)	0.058 (0.095)	0.005 (0.006)	0.011 (0.025)	0.050 (0.098)
total	0.004 (0.006)	0.009 (0.027)	0.050 (0.095)	0.005 (0.007)	0.020 (0.038)	0.087 (0.203)

[a]Errors within the FNO approximation are presented using different thresholds for the occupation numbers of the retained natural orbitals. Maximum errors are given in parentheses.

Table 3. Timings (in minutes) for FNO−CCSD(T) Computations of Counterpoise-Corrected Interaction Energies for Adenine−Thymine (WC) Described by an aug-cc-pVDZ Basis Set[a]

	adenine−thymine				adenine				thymine			
	full	10^{-6}	10^{-5}	10^{-4}	full	10^{-6}	10^{-5}	10^{-4}	full	10^{-6}	10^{-5}	10^{-4}
no. of virtual orbs	468	443	406	320	501	241	211	164	503	229	201	156
$(ov \mid ov)$ transformation		115	119	119		109	110	112		108	110	108
form NOs		0	0	0		0	0	0		0	0	0
AO/NO transformation	253	232	197	151	263	114	107	97	271	112	108	99
CC integral sort	202	171	122	53	218	11	5	2	222	8	4	2
CCSD iterations	1398	1197	895	435	557	29	20	10	541	21	15	8
(T)	2827	2337	1580	1061	514	43	24	9	460	29	18	6
total	4680	4052	2913	1819	1552	306	266	230	1494	278	255	223

[a]Computations were performed using all 16 cores of an AMD Opteron 6272 CPU with access to 64 GB of memory. The number of virtual orbitals included in each computation is also presented.

dx.doi.org/10.1021/ct300780u | J. Chem. Theory Comput. 2013, 9, 293−299

Table 4. Timings (in minutes) for FNO−CCSD(T) Computations of Counterpoise-Corrected Interaction Energies for Benzene−Methane (S22-10) Described by an aug-cc-pVTZ Basis Set[a]

	benzene−methane				benzene				methane			
	full	10^{-6}	10^{-5}	10^{-4}	full	10^{-6}	10^{-5}	10^{-4}	full	10^{-6}	10^{-5}	10^{-4}
no. of virtual orbs	526	457	360	202	531	349	277	156	547	114	83	43
(ov | ov) transformation		178	171	178		168	173	164		157	163	161
form NOs		0	0	0		0	0	0		0	0	0
AO/NO transformation	378	286	219	163	369	211	194	147	361	137	132	129
CC integral sort	293	159	72	4	317	61	30	2	306	0	0	0
CCSD iterations	411	273	120	7	494	79	13	2	332	0	0	0
(T)	281	147	58	6	153	26	11	1	19	0	0	0
total	1363	1043	640	358	1333	545	421	316	1018	294	295	290

[a]Computations were performed using all 16 cores of an AMD Opteron 6272 CPU with access to 64 GB of memory. The number of virtual orbitals included in each computation is also presented.

deviations occur for FNO CCSD(T) with the aggresive 10^{-4} threshold than were observed in the aug-cc-pVDZ basis. The MAE and maximum errors for both methods are given in Table 2. The maximum errors for both schemes with the most aggressive threshold are on the order of tenths of a kilocalorie per mole and too large to be acceptable for benchmark-quality computations. The MAEs for the 10^{-5} threshold are 0.009 and 0.020 for CCSD + FNO(T) and FNO CCSD(T), respectively, which should be acceptable as benchmark-quality solutions, particularly if the CCSD(T) computations would not be possible in the full MO basis.

Finally, we discuss the computational advantages of the FNO CCSD(T) scheme. Tables 3 and 4 provide the time in minutes required for each module involved in counterpoise-corrected FNO CCSD(T) computations for adenine−thymine (WC) in the aug-cc-pVDZ basis set and benzene−methane in the aug-cc-pVTZ basis set.

In the double-ζ basis set, Table 3 shows that the number of virtual orbitals in a dimer computation can be reduced from 468 to 406 for an occupation threshold of 10^{-5}. With this threshold, the cost of the CCSD and (T) portions of the algorithm are reduced by factors of 1.56 and 1.79, respectively. The time for the post-Hartree−Fock computation is reduced by a factor of 1.61 relative to the full basis computation. Because we use semicanonical orbitals, the FNO computations typically require about the same number of iterations as conventional CCSD. In fact, for the test cases presented in Tables 3 and 4, all dimer computations required the same number of iterations, regardless of FNO cutoff. The performance advantages of FNO techniques are more evident for monomer computations in the supermolecular dimer basis, where a threshold of 10^{-5} can reduce the size of the virtual space by more than a factor of 2. For the adenine monomer, we observe speedups in the CCSD and (T) routines of 27.7 and 21.6, respectively. Similar speedups are observed for the thymine monomer (35.9 and 26.1 for CCSD and (T), respectively), and we can safely conclude that the coupled cluster portions of the monomer calculations are essentially free (compared to the dimer calculations) in the FNO scheme.

The computational savings of the FNO scheme increase with the size of the basis set. According to Table 4, the virtual space for the benzene−methane dimer is reduced from 526 to 360 orbitals using a threshold of 10^{-5}. The post-Hartree−Fock portion of the full dimer computation is performed about 2.13 times faster using this threshold than when using the full MO basis. The cost of the CCSD and (T) portions of the algorithm were reduced by factors of 3.43 and 4.84, respectively. The coupled cluster portions of the monomer computations are again essentially free; the cost of the post-Hartree−Fock procedure is dominated by the integral transformation.

Recall that the accuracy of the CCSD + FNO(T) and FNO CCSD(T) methods was comparable across the entire S22 set. The relative utility of the methods is therefore dependent on the computational advantages afforded by each. The savings offered by the truncation of virtual space prior to the solution of the CCSD equations is evident from these timings; both the CC integral sort and solution of CCSD equations greatly benefit from the FNO truncation. The integral sort is performed out-of-core and is thus inherently serial, while our CCSD and (T) algorithms use OpenMP and parallel BLAS libraries. In some cases, this serial sort requires more wall time than the evaluation of the (T) contribution to the energy; any reduction in the cost of the integral sort is important.

The total cost of the AO/NO integral transformation is sometimes more expensive in the FNO CCSD(T) relative to canonical CCSD(T). The reason for this peculiarity is that FNO CCSD(T) requires two separate integral transformations. First, the (ov|ov) integrals must be transformed to the MO basis to determine the MP2 natural orbitals. Second, the full AO/NO transformation must be performed. The cost of the second transformation does reduce with increasingly aggressive NO tolerances, but the cost of the initial (ov|ov) transformation remains constant. Finally, we observe massive savings in the amount of memory and disk space required for our CCSD algorithm when using FNOs. The largest block of integrals in a CCSD computation scales as v^4, so by reducing the size of the virtual space from 547 to 83 for the methane monomer in the dimer basis, the FNO algorithm requires 1800 times less disk space for the storage of this class of integrals. The analogous reduction for the benzene−methane dimer computation leads to a factor of 4.6 savings in the required disk storage. FNOs also allow the CCSD(T) equations to be solved more efficiently with respect to CPU time, but, much more importantly, they facilitate computations with very large basis sets that may otherwise have memory and disk requirements that exceed practical hardware limitations.

4. CONCLUSIONS

Frozen natural orbital coupled cluster methods are applied to the evaluation of interaction energies of noncovalently bound systems. Two FNO schemes are found to yield solutions of comparable quality across the entire S22 test set. For a modest occupation threshold of 10^{-5}, both FNO CCSD(T) and CCSD + FNO(T) yield results that can be considered of benchmark

quality, with FNO CCSD(T) methods yielding an MAE of 0.012 kcal mol^{-1} for the S22 set in an aug-cc-pVDZ basis set versus CCSD(T) using the full MO basis. For the subset of 11 smallest molecules in the aug-cc-pVTZ basis set, the FNO CCSD(T) method yields interaction energies with an MAE of 0.020 kcal mol^{-1}, versus canonical CCSD(T). CCSD + FNO(T) results are slightly more accurate than FNO CCSD(T) results in the larger basis, but the computational benefits of the FNO CCSD(T) scheme are far greater. Our results indicate that, for modest occupation thresholds on the order of 10^{-5}, FNO CCSD(T) methods consistently yield intermolecular interaction energies of benchmark quality and are a viable alternative to CCSD(T) computations using the full molecular orbital basis.

■ ASSOCIATED CONTENT

Supporting Information

The values for $\delta_{MP2}^{CCSD(T)}$ computed for several FNO occupation thresholds for the S22 data set described by the aug-cc-pVDZ basis set and the S11 data set described by the aug-cc-pVTZ basis set. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sherrill@gatech.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.D.S. acknowledges support from the U. S. National Science Foundation (Grant No. CHE-1011360). The computer resources of the Center for Computational Molecular Science and Technology are funded through a National Science Foundation CRIF Award (CHE-0946869). A.E.D. acknowledges support from the National Science Foundation American Competitiveness in Chemistry Postdoctoral Fellowship (CHE-1137288).

■ REFERENCES

(1) Bartlett, R. J.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291−352.
(2) Lee, T. J.; Scuseria, G. E. In *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Langhoff, S. R., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp 47−108.
(3) East, A. L. L.; Allen, W. D. *J. Chem. Phys.* **1993**, *99*, 4638−4650.
(4) Császár, A. G.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **1998**, *108*, 9751−9764.
(5) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146−10159.
(6) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *135*, 194102.
(7) Marshall, M. S.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7*, 3978−3982.
(8) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27−32.
(9) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287−5293.
(10) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282−289.
(11) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(12) Pitonak, M.; Hesselmann, A. *J. Chem. Theory Comput.* **2010**, *6*, 168−178.
(13) Burns, L. A.; Vázquez-Mayagoitia, Á.; Sumpter, B. G.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *134*, 084107.
(14) Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106.
(15) Knizia, G.; Adler, T.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.
(16) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(17) Sosa, C.; Geertsen, J.; Trucks, G. W.; Bartlett, R. J.; Franz, J. A. *Chem. Phys. Lett.* **1989**, *159*, 148−154.
(18) Klopper, W.; Noga, J.; Koch, H.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 164−176.
(19) Taube, A. G.; Bartlett, R. J. *Collect. Czech. Chem. Commun.* **2005**, *70*, 837−850.
(20) Taube, A. G.; Bartlett, R. J. *J. Chem. Phys.* **2008**, *128*, 164101.
(21) Rolik, Z.; Kállay, M. *J. Chem. Phys.* **2011**, *134*, 124111.
(22) Landau, A.; Khistyaev, K.; Dolgikh, S.; Krylov, A. I. *J. Chem. Phys.* **2010**, *132*, 014109.
(23) Sherrill, C. D.; Schaefer, H. F. *J. Phys. Chem.* **1996**, *100*, 6069−6075.
(24) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *121*, 10905−10918.
(25) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *121*, 10852−10862.
(26) Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130*, 11410.8.
(27) Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131*, 064103.
(28) Hylleraas, E. A. *Z. Phys* **1929**, *54*, 347.
(29) Adamowicz, L.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 6314−6324.
(30) Neogrády, P.; Pitoňák, M.; Urban, M. *Mol. Phys.* **2005**, *103*, 2141−2157.
(31) Adamowicz, L. *Mol. Phys.* **2010**, *108*, 3105.
(32) Pitoňák, M.; Neogrády, P.; Řezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829−1834.
(33) Dedíková, P.; Pitoňák, M.; Neogrády, P.; Černušák, I.; Urban, M. *J. Phys. Chem. A* **2008**, *112*, 7115−7123.
(34) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *133*, 104107.
(35) Kraus, M.; Pitoňák, M.; Hobza, P.; Urban, M.; Neogrády, P. *Int. J. Quantum Chem.* **2012**, *112*, 948−959.
(36) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(37) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. *WIREs Comput. Mol. Sci.* **2012**, *2*, 556−565.
(38) Hohenstein, E. G.; Sherrill, C. D. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304−326.