The Journal of Chemical Physics

Benchmark studies of electron correlation in six-electron systems

Cite as: J. Chem. Phys. **100**, 8132 (1994); https://doi.org/10.1063/1.466807
Submitted: 01 November 1993 . Accepted: 24 February 1994 . Published Online: 31 August 1998

Justin T. Fermann, C. David Sherrill, T. Daniel Crawford, and Henry F. Schaefer

AIP Publishing

J. Chem. Phys. **100**, 8132 (1994); https://doi.org/10.1063/1.466807 **100**, 8132

Benchmark studies of electron correlation in six-electron systems

Justin T. Fermann,[a),b)] C. David Sherrill,[c)] T. Daniel Crawford,[d)] and Henry F. Schaefer III
Center for Computational Quantum Chemistry, University of Georgia, Athens, Georgia 30602

(Received 1 November 1993; accepted 24 February 1994)

Benchmark studies of H_6, H_7^+, and $H_5^+\cdot He$ were initiated in order to develop more efficient theoretical methods for describing the electron correlation energy, due to the relative ease with which the full configuration interaction (full CI) results could be obtained for these six-electron systems. Single-point energies which approach the quality of the full CI results are reported for a variety of coupled-cluster (CC) and configuration interaction (CI) methods using optimized basis sets and full CI optimized geometries. Emphasis is placed on multireference CI (MRCI) methods. By carefully limiting the configurations included in the CI and by using CI natural orbitals, we find it possible to reduce the number of configuration state functions (CSFs) by two orders of magnitude or more with little loss in the correlation energy recovered for the six-electron systems studied here. To judge the applicability of the MRCI methods to the study of potential energy surfaces, the energies of H_2O at three geometries are compared to previously published full CI and complete active space self-consistent-field (CASSCF) second-order CI (SOCI) results. Finally, we propose a compact MRCI wave function incorporating limited triple and quadruple excitations. Indirect tests suggest that this new approach should be highly effective.

INTRODUCTION

Due to continuing increases in computing power available to theoreticians and the development of robust and efficient implementations of theoretical methods, it is now possible to use the full CI method to exactly solve the Schrödinger equation using limited basis sets for systems with up to approximately eight electrons.[1,2] Obviously, this does not yet preclude the need to develop methods for the evaluation of the correlation energy that remain good approximations to a full CI calculation yet require less effort. It has long been known that truncating the CI procedure by leaving out higher order excitations results in drastically reduced configuration lists while still producing useful wave functions.[3] Similarly, deleting higher-lying virtual molecular orbitals (MOs) and removing low-lying core MOs from the active space is an effective means of reducing the size of the CI space while still describing correlation adequately.[4]

A routine CI wave function typically includes all configurations corresponding to single and double excitations from a Hartree–Fock reference, denoted CISD. This method garners approximately 95% of the correlation energy for systems with ten electrons;[3] most of the remaining five percent is attributed to the inclusion of triple and quadruple excitations from the reference function.[3] However, including all triple and quadruple excitations (CISDTQ) increases the number of configuration state functions (CSFs) in the CI expansion by an order of magnitude for a system with six electrons and 30 basis functions. Thus the length of a full CISDTQ configuration list often makes such a wave function prohibitively expensive. When a well-ordered set of MOs is chosen, it is observed that the triple and quadruple excitations into the higher lying virtual space make negligible contributions compared to the same order excitations into the lower virtual space.[3,5] By constructing a wave function that specifically eliminates those unimportant CSFs, it should be possible to obtain a very efficient method for estimating the contribution of higher-order excitations. The first-order CI (FOCI) and second-order CI (SOCI) methods were developed for this purpose.[6] The FOCI and SOCI methods are employed here by taking all single or all single and double excitations, respectively, from a reference space consisting of all the CSFs resulting from a full CI in the valence space. FOCI is far less effective than simple CISD for systems qualitatively described by the single configuration Hartree–Fock wave function; therefore, apart from the inclusion of FOCI results in the tables, this method will not be discussed further. SOCI, on the other hand, is much more complete than CISD, and has been successfully employed in such notoriously difficult correlation problems as N_2.[7] Unfortunately, SOCI remains intractable for systems with more than a few heavy atoms. Two newer, more compact methods, here referred to as CISD[T] and CISD[TQ], are multireference CISD [MRCI(SD)] techniques that use a set of reference functions generated *a priori* by taking all single or all single and double excitations from the Hartree–Fock reference configuration within a valence space determined by choosing MOs based upon the number of symmetry orbitals that can be formed from the valence atomic orbitals.[8,9] One may think of CISD[T] as CISDT in which no more than two electrons are allowed to simultaneously reside outside the valence space, and CISD[TQ] as CISDTQ in which no more than two electrons may reside outside the valence space. It is hoped that through the use of the CISD[TQ] method, one may account for the dominant contributions from triple and quadruple excitations without the expense of a CISDTQ expansion. Grev and Schaefer[8] have shown that the efficiency of such a MRCI method is further increased by using CI natural orbitals (CINOs), which are ordered according to their occupancy through diagonalization of the one-particle

---

[a)] Author to whom correspondence should be addressed.
[b)] Friedrich Hund Graduate Fellow.
[c)] National Science Foundation Graduate Fellow.
[d)] Department of Defense Graduate Fellow.

0021-9606/94/100(11)/8132/8/$6.00

density matrix, as the references generated from such orbitals are then more appropriate for describing nondynamical correlation.[3,10]

In addition, it is possible to make a slight departure from traditional methods of constructing a MRCI wave function intended to be applicable to several nuclear arrangements on a potential energy surface (PES). For many such applications, the references used in generating a MRCI(SD) wave function are chosen to be those with the largest CI coefficients in a multiconfiguration self-consistent-field (MCSCF) wave function. Typically, values of 0.02 to 0.05 are used to determine what constitutes an "important" CSF. At each geometry, a MCSCF wave function produces a set of MOs and a list of references for use in further MRCI expansions. The lists of references from each point are combined and applied throughout the surface, producing a compact wave function that contains CSFs important in the description of the electronic structure at each geometry. One can expect such a wave function to produce energies which continuously parallel the full CI PES. If the MOs are instead obtained from a CISD natural orbital procedure and the references selected from a list of the important CSFs in a CISD or, better still, a CISD[TQ] or SOCI expansion, the resulting wave function is also compact and of high quality.

In this work, the geometrical structures of H_6, H_7^+, and $H_5^+ \cdot He$ have been optimized using full CI with optimized double zeta plus polarization (DZP) basis sets. A variety of CI and coupled-cluster (CC) wave functions are analyzed in terms of their relative sizes and their ability to reproduce full CI energies. All energies are reported in tabular form with the expectation that this will provide a thorough comparative set of several of the more widely applied methods. In particular, MRCI(SD) methods with references chosen based upon their CI coefficients in some previous CI expansion are compared to CISD[TQ] which selects the references *a priori* as all single and double excitations in the valence space.

In addition, single-point energies are determined for H_2O at three geometries and compared to the full CI and MRCI(SD) results of Bauschlicher and Taylor,[2] gauging not only the extent to which each method recovers the correlation energy but also how well each method parallels the full CI PES. Finally, a proposal for a variation on the CISD[TQ] method is examined in anticipation of the need for a highly compact MRCI wave function that includes dominant triple and quadruple excitations.

METHODS

We have applied theoretical methods ranging in complexity from self-consistent-field (SCF) to full CI to all the systems studied here. Initially, the geometries at which single-point energies were calculated for the six-electron systems were obtained by optimization of the structures using analytic gradient methods at the full CI level of theory.[11] The DZP basis sets for hydrogen used in the study of H_6, H_7^+, and $H_5^+ \cdot He$ are based on the standard Huzinaga[12]–Dunning[13] $(4s/2s)$ contraction plus polarization. The DZP basis set for He used in the study of $H_5^+ \cdot He$ revises van Duijneveldt's $(6s/1s)$ contraction[14] to a $(6s/2s)$ contraction augmented by a set of p polarization functions. For each molecule, the s-orbital scaling factor, ζ, and the p-orbital exponent, α_p, were optimized separately.

For H_6, the basis set and geometry were optimized iteratively at the full CI level of theory yielding $\zeta=1.10$ and $\alpha_p=0.742$. The full CI calculations for H_7^+ and $H_5^+ \cdot He$ were substantially larger due to their lower symmetry, so the basis sets were optimized iteratively with the geometry at the CISDTQ level of theory in these cases. For H_7^+, the basis set optimization yielded $\alpha_p=0.681$ and $\zeta=1.187$. The results for $H_5^+ \cdot He$ were $\alpha_p=0.726$ and $\zeta=1.136$ for hydrogen and $\alpha_p=1.233$ and $\zeta=1.048$ for helium. Using the optimized basis sets, the geometries of H_6, H_7^+, and $H_5^+ \cdot He$ were reoptimized at the full CI level of theory. For the H_2O calculations we used the DZP basis set and geometries reported previously[15] in Ref. 2.

The following single-reference theoretical methods were employed for all systems using SCF canonical orbitals: SCF, CISD, CISDT, CISDTQ, full CI, coupled cluster including single and double excitations (CCSD),[16] CCSD with perturbatively included triple excitations [CCSD(T)],[17] and CCSD with full inclusion of triple excitations (CCSDT).[18] The complete active space-SCF (CAS SCF) wave functions from the benchmark H_2O calculations of Ref. 2 and MRCI wave functions based on them were reproduced as per the methods of Bauschlicher *et al.* to illustrate the compatibility of implementations. Minor discrepancies (<20 μH) were discovered between our MRCI results and those of Bauschlicher *et al.* We expect these are due to different convergence criteria in the SCF wave functions. As the full CI method is rigorously invariant with respect to rotations among the active orbitals,[3] we accept their full CI energy benchmarks as published.

In addition to these standard methods, first-order CI (FOCI) and second-order CI (SOCI), as well as CISD[T] and CISD[TQ] were applied. With valence space chosen as described, these methods were employed using both the canonical SCF orbitals and the CI natural orbitals obtained from CISD calculations. To explore the rapidity with which successively larger CI spaces converged to the full CI limit, NO-MRCI(SD) calculations were performed in which the references were chosen according to their CI coefficients in the NO-CISD wave function. At times, the addition of another CSF to the reference space of the NO-MRCI(SD) wave function did not increase the size of the CI space because two of the reference CSFs described the same orbital occupations. Because the CSFs are not restricted to the first-order interacting space, the sets of excitations will be identical. All wave functions were constructed using the highest possible symmetry except for $D_{6h}H_6$, which was formed in the nondegenerate point group D_{2h}. To avoid confusion, the numbers of CSFs in the entire CI space for each wave function are included in the tables.

In the study of H_2O, additional MRCI wave functions were employed to treat the PES evenly. The MOs used in each case were the CINOs from a standard CISD procedure, and three different sets of reference functions were used. Lists of important CSFs were generated separately at each geometry from CISD, CISD[TQ], and SOCI wave functions using a minimum coefficient threshold of 0.020. The result-

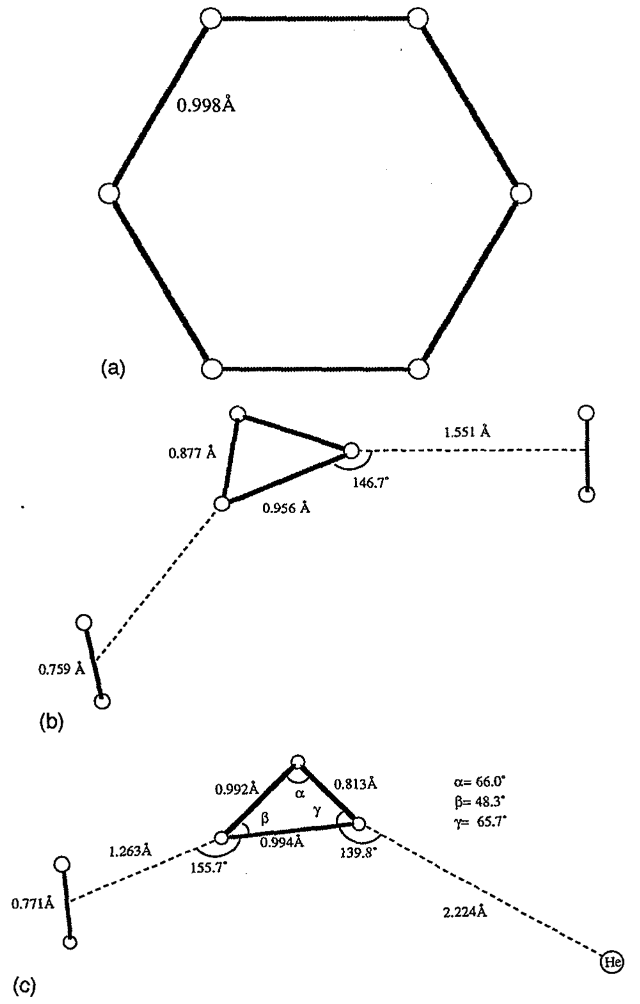

FIG. 1. (a) Full CI D_{6h} stationary point of H_6, with a full CI optimized DZP basis. (b) Full CI C_{2v} equilibrium geometry of H_7^+, with a CISDTQ optimized DZP basis. (c) Full CI C_s equilibrium geometry of $H_5^+ \cdot He$, with a CISDTQ optimized DZP basis.

ing CSF lists for each method were then merged to form the three overall reference CSF lists for use in MRCI(SD) expansions at all geometries. The overall reference list from the CISD wave functions contained 42 CSFs, the list from the CISD[TQ] wave functions contained 52 CSFs, and the list from the SOCI wave functions contained 50 CSFs. These produced MRCI(SD) wave functions consisting of 46 136, 53 611, and 52 396 CSFs, respectively. Each of these MRCI(SD) wave functions is about twice as compact as the SOCI wave function, yet they show markedly better performance in reproducing the full CI PES than almost any other truncated CI method examined, including CISDTQ.

TABLE I. Energies for (a) H_6 at the full configuration interaction D_{6h} stationary point[a] seen in Fig. 1(a), (b) H_7^+ at the full configuration interaction C_{2v} equilibrium geometry[a] seen in Fig. 1(b), and (c) $H_5^+ \cdot He$ at the full configuration interaction C_s equilibrium geometry[a] seen in Fig. 1(c).

Method	Number of CSFs	Energy[b]	$\%E_{corr}$
(a)			
SCF	1	−3.251 9928	0.00
CISD	596	−3.382 9989	94.54
CISDT	8680	−3.386 0295	96.73
CISDTQ	70 263	−3.390 4620	99.93
CISD[T]	3232	−3.385 3224	96.22
CISD[TQ]	7709	−3.387 8204	98.02
FOCI	1000	−3.322 5308	50.90
SOCI	10 572	−3.387 8475	98.04
CINO-CISD[T][c]	3232	−3.385 8190	96.58
CINO-CISD[TQ][c]	7709	−3.389 4550	99.20
CINO-FOCI[c]	1000	−3.344 8811	67.03
CINO-SOCI[c]	10 572	−3.389 5145	99.24
CCSD[d]		−3.387 2932	97.64
CCSD(T)[d]		−3.390 0035	99.60
CCSDT[d]		−3.390 4590	99.92
FCI	594 712	−3.390 5630	100.00
(b)			
SCF	1	−3.572 1951	0.00
CISD	1271	−3.683 8431	96.36
CISDT	24 468	−3.684 4365	96.87
CISDTQ	248 149	−3.688 0146	99.96
CISD[T]	8898	−3.684 2895	96.75
CISD[TQ]	26 255	−3.686 2517	98.44
FOCI	3587	−3.632 1482	51.74
SOCI	43 573	−3.686 2646	98.45
CINO-CISD[T][c]	8898	−3.684 3544	96.80
CINO-CISD[TQ][c]	26 255	−3.687 3529	99.39
CINO-FOCI[c]	3587	−3.649 5000	66.72
CINO-SOCI[c]	43 573	−3.687 3822	99.41
CCSD[d]		−3.687 4136	99.44
CCSD(T)[d]		−3.687 9373	99.89
CCSDT[d]		−3.688 0460	99.99
FCI	2 923 933	−3.688 0604	100.00
(c)			
SCF	1	−5.295 6324	0.00
CISD	1918	−5.402 8461	96.97
CISDT	30 550	−5.403 3691	97.44
CISDTQ	257 755	−5.406 1691	99.98
CISD[T]	10 470	−5.403 2290	97.32
CISD[TQ]	24 513	−5.404 3776	98.36
FOCI	2700	−5.340 0261	40.15
SOCI	33 540	−5.404 3801	98.36
CINO-CISD[T][c]	10 470	−5.403 3054	97.39
CINO-CISD[TQ][c]	24 513	−5.405 3686	99.25
CINO-FOCI[c]	2700	−5.353 3679	52.22
CINO-SOCI[c]	33 540	−5.405 3779	99.26
CCSD[d]		−5.405 6282	99.49
CCSD(T)[d]		−5.406 0860	99.90
CCSDT[d]		−5.406 1826	99.99
FCI	2 289 280	−5.406 1961	100.00

[a] Unless otherwise noted, SCF canonical orbitals were used in the correlation procedure.

[b] Energies in hartrees obtained using a DZP basis optimized at the FCI level in part (a) and at the CISDTQ level in parts (b) and (c).

[c] Using CISD natural orbitals.

[d] Nonvariational method. %Error relative to E_{corr}.

RESULTS AND DISCUSSION

Using the optimized DZP basis sets and full CI optimized geometries shown in Fig. 1, energies for H_6, H_7^+, and $H_5^+ \cdot He$ were obtained at numerous levels of theory. The total energies are summarized in Table I, along with the percent of the electron correlation energy (ECE) that each wave function recovers. At a stationary point that is fairly well described by single reference methods, the common CISD wave function recovers 94% to 97% of the ECE. For the six-electron systems under study, the effect of triple and quadruple excitations accounts for nearly all of the remaining 3% to 6%, with quintuple and sextuple excitations contributing a mere 0.07% (or 0.10 mH absolute) of the ECE. Since the reliable treatment of triple and quadruple excitations is a major goal of quantum chemistry, the systems chosen are in this sense illustrative.

Although not optimal for use in this context, the SCF canonical orbitals were employed with the CISD[TQ] and SOCI methods. The two methods yield almost identical results, recovering more than 98% of the correlation energy. An obvious improvement is made by constructing the CISD[TQ] and SOCI wave functions using CINOs. This leads to recovery of greater than 99% of the ECE by both methods. The fact that the energy obtained by the two methods is so similar implies that the extra CSFs included in the SOCI are not important in evaluating the correlation energy and that the CISD[TQ] method provides a substantial savings in wave function size at a negligible cost in energy.

To examine the effectiveness of the CISD[TQ] method, we construct a number of different MRCI(SD) wave functions by taking the important CSFs from a NO-CISD wave function and using them as references. Table II has been generated by adding one new reference configuration at a time to the reference space. For each system, the MRCI energy approaches the full CI energy surprisingly quickly, approaching to within 0.5 mH when a maximum of 25 CSFs form the reference space. Attaining agreement with the full CI result that is within the realm of "chemical accuracy" using a wave function that is orders of magnitude shorter than a full CI expansion is a noteworthy accomplishment. The efficiency is unquestionable—in all cases, the SOCI method is bested with a more compact wave function, as much as 63% smaller in the case of $H_5^+\cdot$He. Upon the examination of the important CSFs in the CISD wave functions for H_7^+ and $H_5^+\cdot$He, we note that many involve excitations into MOs not included in the valence space. Because the CISD[TQ] and SOCI methods use only references with electrons restricted to the valence space, CSFs with excitations to these higher-lying MOs will not be listed as a reference in either case and including these necessarily produces a better wave function.

The three coupled-cluster methods perform very well in these cases and are computationally efficient. For $H_5^+\cdot$He, a typical CCSD(T) calculation takes only 67 seconds of cpu time while the SOCI uses 124 s on an IBM RS/6000 model 550. In all three cases, the CC energies achieve a smaller relative error with respect to the full CI than the best MRCI results. Although this suggests remarkable capabilities for these methods, their shortcomings due to their single reference nature become evident.

TABLE II. Multireference configuration interaction energies[a] for (a) H_6 at the full configuration interaction D_{6h} stationary point using CISD natural orbitals,[b] (b) H_7^+ at the full configuration interaction C_{2v} equilibrium geometry using CISD natural orbitals,[b] and (c) $H_5^+\cdot$He at the full configuration interaction C_s equilibrium geometry using CISD natural orbitals.[b]

Method	Number of CSFs	Energy	$\%E_{\text{corr}}$
(a)			
CISD	596	−3.383 0031	94.54
2R-CISD	2502	−3.386 8722	97.34
3R-CISD	2502	−3.386 8722	97.34
4R-CISD	3186	−3.388 1535	98.26
5R-CISD	3186	−3.388 1535	98.26
6R-CISD	5656	−3.388 8785	98.78
7R-CISD	5656	−3.388 8785	98.78
8R-CISD	6172	−3.389 2697	99.07
9R-CISD	6172	−3.389 2697	99.07
10R-CISD	6172	−3.389 2697	99.07
11R-CISD	10 216	−3.389 6595	99.35
12R-CISD	10 216	−3.389 6595	99.35
13R-CISD	10 520	−3.389 7057	99.38
14R-CISD	10 520	−3.389 7057	99.38
15R-CISD	10 520	−3.389 7057	99.38
16R-CISD	12 313	−3.389 8316	99.47
17R-CISD	12 313	−3.389 8316	99.47
18R-CISD	12 920	−3.389 9471	99.56
19R-CISD	12 920	−3.389 9471	99.56
20R-CISD	15 328	−3.390 0609	99.64
21R-CISD	16 230	−3.390 1015	99.67
22R-CISD	16 230	−3.390 1015	99.67
23R-CISD	16 999	−3.390 1323	99.69
24R-CISD	16 999	−3.390 1323	99.69
25R-CISD	18 152	−3.390 1599	99.71
FCI	594 712	−3.390 5630	100.00
(b)			
CISD	1271	−3.683 8515	96.37
2R-CISD	5487	−3.684 8184	97.20
3R-CISD	6613	−3.684 9818	97.34
4R-CISD	7658	−3.685 5264	97.81
5R-CISD	8663	−3.686 0244	98.24
6R-CISD	9175	−3.686 3175	98.50
7R-CISD	9687	−3.686 5756	98.72
8R-CISD	10 995	−3.686 7930	98.91
9R-CISD	11 843	−3.686 9791	99.07
10R-CISD	12 590	−3.687 0468	99.13
11R-CISD	15 510	−3.687 2168	99.27
12R-CISD	16 503	−3.687 2929	99.34
13R-CISD	16 843	−3.687 3221	99.36
14R-CISD	17 828	−3.687 3853	99.42
15R-CISD	20 634	−3.687 4622	99.48
16R-CISD	22 960	−3.687 5313	99.54
17R-CISD	26 595	−3.687 6153	99.62
18R-CISD	29 999	−3.687 6789	99.67
19R-CISD	30 897	−3.687 6935	99.68
20R-CISD	31 717	−3.687 7093	99.70
21R-CISD	33 631	−3.687 7386	99.72
22R-CISD	33 631	−3.687 7386	99.72
23R-CISD	33 631	−3.687 7386	99.72
24R-CISD	35 123	−3.687 7905	99.77
25R-CISD	35 123	−3.687 7905	99.77
FCI	2923 933	−3.688 0604	100.00
(c)			
CISD	1918	−5.402 8511	96.97
2R-CISD	3526	−5.403 6695	97.71
3R-CISD	6317	−5.404 3531	98.33
4R-CISD	7919	−5.404 9229	98.85
5R-CISD	9239	−5.405 1367	99.04
6R-CISD	12 175	−5.405 4341	99.31
7R-CISD	12 652	−5.405 5112	99.38

Tables III and IV report energies for H_2O at three geometries in order to indicate the ability of each method to map a PES parallel to the full CI surface generated by Bauschlicher *et al.* At $1.0{*}r_e$, the three coupled-cluster methods show a strong ability to reproduce the full CI energy. The

TABLE II. (*Continued.*)

Method	Number of CSFs	Energy	$\%E_{\text{corr}}$
8R-CISD	13 166	−5.405 5719	99.44
9R-CISD	13 643	−5.405 6110	99.47
10R-CISD	14 157	−5.405 6409	99.50
11R-CISD	19 308	−5.405 7243	99.57
12R-CISD	20 737	−5.405 8189	99.66
13R-CISD	22 207	−5.405 9122	99.74
14R-CISD	26 151	−5.405 9776	99.80
15R-CISD	26 151	−5.405 9776	99.80
16R-CISD	27 546	−5.406 0396	99.86
17R-CISD	29 965	−5.406 0736	99.89
18R-CISD	32 236	−5.406 0986	99.91
19R-CISD	35 990	−5.406 1123	99.92
20R-CISD	36 924	−5.406 1221	99.93
21R-CISD	37 900	−5.406 1320	99.94
22R-CISD	38 596	−5.406 1337	99.94
23R-CISD	39 003	−5.406 1354	99.95
24R-CISD	39 774	−5.406 1371	99.95
25R-CISD	42 231	−5.406 1398	99.95
FCI	2289 280	−5.406 1961	100.00

[a] Energies in hartrees obtained using a DZP basis optimized at the full CI level in parts (a) and at the CISDTQ level in parts (b) and (c).

[b] See text for a discussion of the choice of reference functions.

performance of these methods is markedly poorer at stretched geometries, where the system attains multireference character. At $2.0*r_e$, the CCSDT ran for 220 iterations after which the energy step size was less than 10^{-6} and the T_1 diagnostic was 0.054, too large to imply the expected accuracy in the calculation.[19] Even so, these nonvariational methods parallel the full CI PES fairly well and still require less cpu time in most cases than even the CISD[TQ].

Of the variational methods, conclusions similar to those based on the results for H_6, H_7^+, and $H_5^+\cdot$He can be drawn from the energies reported for H_2O at $1.0*r_e$. The agreement with the full CI results for the various methods is somewhat worse, due in part to the increased number of electrons and the possibility of important higher-order excitations. Of more interest are the results from distorted geometries. The CINO-SOCI method is known to be biased slightly in favor of dissociated geometries while it maintains a rather constant error from the full CI,[8] and the results in Table III support this observation. Using CINO-CISD[TQ], although not as accurate as SOCI, provides a relatively inexpensive and straightforward approximation which is far superior to any single-reference CI calculation of comparable size.

Table IV presents a series of (n)R-CISD wave functions in which the references were taken as the (n) most important CSFs of a NO-CISD calculation at each geometry. This procedure is *not* valid for mapping the PES, as it will generally produce one which is discontinuous. However, this method does illustrate the maximum effectiveness of obtaining reference CSFs from a CISD wave function. By the inclusion of 25 CSFs in the reference space, the MRCI(SD) space is approximately one third the size of the SOCI space. Near the

TABLE III. Total energies[a] for H_2O from a variety of correlated methods.

Method	Number of CSF's	$1.0*r_e$ energy	$1.5*r_e$ energy	$2.0*r_e$ energy
SCF	1	−76.040 5419	−75.800 4944	−75.582 2860
CISD	926	−76.243 7722	−76.040 9834	−75.876 6251
CISDT	15 520	−76.247 0005	−76.047 8969	−75.891 9354
CISDTQ	151 248	−76.256 2267	−76.069 8583	−75.945 9889
CISD[T]	8080	−76.246 2572	−76.047 5127	−75.891 6067
CISD[TQ]	32 361	−76.252 9029	−76.067 8152	−75.944 6330
FOCI	10 040	−76.168 7867	−76.014 8666	−75.906 7708
SOCI	96 072	−76.253 1258	−76.069 0314	−75.950 2897
CINO-CISD[T][b]	8080	−76.246 7693	−76.047 6535	−75.891 6067
CINO-CISD[TQ][b]	32 361	−76.254 9941	−76.068 8680	−75.945 4024
CINO-FOCI[b]	10 040	−76.197 6439	−76.028 9173	−75.916 5864
CINO-SOCI[b]	96 072	−76.255 3477	−76.070 3470	−75.951 2486
CCSD		−76.252 5024	−76.061 2468	−75.930 8648
CCSD(T)		−76.255 9072	−76.069 4071	−75.956 9029
CCSDT		−76.256 0924	−76.069 3726	−75.954 7414
MRCI(SD)[b] with reference configurations selected[c] from				
CINO-CISD[b]	46 136	−76.255 7422	−76.069 4358	−75.945 9049
CINO-CISD[TQ][b]	53 611	−76.255 9625	−76.070 6291	−75.951 1979
CINO-SOCI[b]	52 396	−76.255 9381	−76.070 6630	−75.951 2898
CAS[d]	12	−76.094 713	−75.924 781	−75.823 721
MRCI[d]	8805	−76.251 643	−76.066 885	−75.948 557
CAS(BIG)[d]	55	−76.129 876	−75.953 141	−75.839 916
MRCI(BIG)[d]	31 096	−76.254 108	−76.069 363	−75.950 517
FCI[d]	6740 280	−76.256 624	−76.071 405	−75.952 269

[a] Energies in hartrees obtained using the DZP basis and geometry of Ref. 2. Unless otherwise noted, SCF canonical orbitals were used in the correlation procedure.

[b] Using CISD natural orbitals.

[c] See text for a discussion of the choice of reference functions.

[d] Results taken from Ref. 2.

TABLE IV. Multireference configuration interaction total energies[a] for H_2O using CISD natural orbitals.[b]

Method	Number of CSFs	$1.0*r_e$ energy	Number of CSFs	$1.5*r_e$ energy	Number of CSFs	$2.0*r_e$ energy
CISD	926	−76.243 5507	926	−76.040 6383	926	−75.876 9099
2R-CISD	4390	−76.248 6330	4390	−76.058 6444	4390	−75.926 6792
3R-CISD	5011	−76.249 8465	5011	−76.061 8014	5011	−75.935 2382
4R-CISD	5632	−76.250 8312	5632	−76.064 6523	5632	−75.942 2994
5R-CISD	6423	−76.251 9327	6199	−76.065 2690	6199	−75.942 7795
6R-CISD	9348	−76.252 7825	6750	−76.065 7370	6750	−75.943 2265
7R-CISD	9348	−76.252 7825	7539	−76.066 6149	7539	−75.943 8286
8R-CISD	11 977	−76.253 4314	10 424	−76.067 2870	7539	−75.943 8286
9R-CISD	11 977	−76.253 4314	13 025	−76.067 8047	10 424	−75.944 2689
10R-CISD	14 424	−76.253 7509	13 025	−76.067 8047	13 035	−75.944 6574
11R-CISD	16 612	−76.254 0307	13 025	−76.067 8047	15 478	−75.944 8481
12R-CISD	16 612	−76.254 0307	15 478	−76.068 0506	15 478	−75.944 8481
13R-CISD	19 836	−76.254 3631	17 628	−76.068 2355	15 478	−75.944 8481
14R-CISD	21 765	−76.254 5468	19 551	−76.068 3704	17 626	−75.945 0055
15R-CISD	22 261	−76.254 6412	22 203	−76.068 6099	19 551	−75.945 0983
16R-CISD	25 420	−76.254 9581	24 077	−76.068 6834	22 303	−75.945 2808
17R-CISD	25 896	−76.255 0319	26 594	−76.068 8690	24 855	−75.945 4495
18R-CISD	25 896	−76.255 0319	26 594	−76.068 8690	26 822	−75.945 5095
19R-CISD	26 685	−76.255 0925	26 890	−76.068 9088	28 847	−75.945 5575
20R-CISD	27 308	−76.255 1514	28 748	−76.068 9896	30 571	−75.945 6089
21R-CISD	29 821	−76.255 3086	31 477	−76.069 0891	30 571	−75.945 6089
22R-CISD	29 821	−76.255 3086	31 477	−76.069 0891	32 391	−75.945 6559
23R-CISD	29 993	−76.255 3443	31 477	−76.069 0891	32 391	−75.945 6559
24R-CISD	29 993	−76.255 3443	31 868	−76.069 1187	34 218	−75.945 6910
25R-CISD	30 340	−76.255 3719	34 470	−76.069 2051	35 727	−75.945 7145
FCI[c]	6740 280	−76.256 624	6740 280	−76.071 405	6740 280	−75.952 269

[a] Energies in hartrees obtained using the DZP basis and geometry of Ref. 2.
[b] See text for a discussion of the choice of reference functions.
[c] Results taken from Ref. 2.

equilibrium geometry ($1.0*r_e$), this wave function attains a total energy lower than the SOCI wave function. At larger distortions, it becomes evident that the single reference CISD wave function ceases to be an optimum source of references, as the deviations from the full CI result become substantially larger. This is primarily due to the increasing importance of CSFs corresponding to greater than double excitations from the Hartree–Fock reference in describing nondynamical correlation.

Such is the justification for using the MCSCF method to generate references for a MRCI(SD) as well as for the alternate methods explored here. Using a coefficient threshold of 0.020 for elimination of CSFs from the reference space, the important CSFs in the CISD wave functions at all three geometries were merged and used to generate a MRCI(SD) wave function at each geometry. The same was done using the important CSFs from CISD[TQ] wave functions and from SOCI wave functions. These results are presented in Table III, and the absolute energy differences from the full CI for these and several other methods are compared with the MRCI results of Bauschlicher *et al.* in Table V. Because the SOCI is intractable for most chemical systems, a SOCI-based MRCI(SD) is taken as an ordinarily unattainable limit for this type of wave function, very close to the full CI-based MRCI(SD). All three of these wave functions perform satisfactorily; the CISD-based MRCI(SD) is just slightly better than CISDTQ at reproducing the full CI surface. The CISD[TQ]- and SOCI-based MRCI(SD) methods each produce a PES more accurate than all but the full SOCI, and outstrip any other truncated CI method examined here in terms of recovery of ECE. CISD[TQ]-based MRCI(SD), the most practical in terms of both performance and effort, generated a PES parallel to the full CI result to within 0.4 mH,

TABLE V. Absolute energy differences in hartrees from FCI for various methods.[a]

Method	$1.0*r_e$	$1.5*r_e$	$2.0*r_e$
SCF	0.216 082	0.270 911	0.369 983
CISD	0.012 851	0.030 421	0.075 644
CISDTQ	0.000 397	0.001 547	0.006 280
CISD[TQ][b]	0.001 630	0.002 537	0.006 867
FOCI[b]	0.058 980	0.042 488	0.035 683
SOCI[b]	0.001 276	0.001 058	0.001 020
CCSD	0.004 122	0.010 158	0.021 404
CCSD(T)	0.000 717	0.001 998	−0.004 634
CCSDT	0.000 532	0.002 032	−0.002 472
[MRCI(SD)][b] reference configurations selected[c] from			
CINO-CISD[b]	0.000 882	0.001 969	0.006 364
CINO-CISD[TQ][b]	0.000 662	0.000 776	0.001 071
CINO-SOCI[b]	0.000 686	0.000 742	0.000 979
CAS[d]	0.161 911	0.146 624	0.128 548
MRCI[d]	0.004 891	0.004 520	0.003 712
CAS(BIG)[d]	0.126 748	0.118 264	0.112 353
MRCI(BIG)[d]	0.002 516	0.002 042	0.001 752

[a] Energies obtained using the DZP basis and geometry of Ref. 2.
[b] Using CISD natural orbitals.
[c] See text for a discussion of the choice of reference functions.
[d] Results taken from Ref. 2.

but required roughly *half* the CSFs of the SOCI in the final calculation. The total time spent generating the references and then applying them in the final calculation places this method on a more equal footing with the SOCI in this particular case. In successively larger molecules, however, we expect that the SOCI will become prohibitively large far sooner than either of the steps needed to obtain the CISD[TQ]-based MRCI(SD) wave function.

PROPOSAL FOR A MORE COMPACT CISD[TQ] WAVE FUNCTION

We have observed that the CISD[TQ] wave function proves to be a good approximation to SOCI at a substantially reduced computational expense. Nevertheless, CISD[TQ] is still too large to be applied beyond relatively small systems. We believe that the effectiveness of the CISD[TQ] method can be radically increased by a conceptually simple but practically challenging modification of our GUGA CI method.[9] It has been known since the 1960 work of Watson[20] on the Be atom that the most important quadruple excitations have much smaller CI coefficients than the most important double excitations. This finding was qualitatively explained by Sinanoglu,[21] who showed that the importance of an unlinked quadruple excitation is related to the product of the two double excitations from which it may be derived. This suggests that a much smaller NO basis can satisfactorily treat the triple and quadruple excitation contributions than is required to recover the same fraction of the correlation energy due to single and double excitations.

Preliminary evidence that this may be the case was set forth by Grev and Schaefer in 1992, in the analysis of N_2.[8] There the authors found that 92.8% of the energy lost by reducing the number of active orbitals from 108 to 50 by deleting the 58 most weakly occupied virtual NOs used in a CISD[TQ] expansion could be recovered by the CISD method. Table VI demonstrates that this is again possible in the case of H_7^+. There, both the CISD and CISDTQ methods were applied using the NOs generated by a CISD. The highest-lying NOs were deleted one at a time, and the CISD and CISDTQ energies for these smaller NO spaces are given in relation to the corresponding energies with the full 35 orbitals ($E_n - E_{35}$). We observe that for NO spaces nearly as large as the full NO space, the energy difference $E_n - E_{35}$ is virtually identical for the CISD and CISDTQ methods, indicating that triples and quadruples which occupy the deleted NOs contribute negligibly to the energy. For instance, removing the first NO results in a loss of 48 μH for CISD and 50 μH for CISDTQ. The difference between these two, labeled "TQ loss" in the fourth column of Table IV, is a mere 2 μH. This difference is also given as a percentage of the total energy due to triples and quadruples in the CISDTQ with 35 orbitals, which is 0.004 164 hartree. It is remarkable that we may eliminate more than *half* the high-lying virtual NOs before we lose even 10% of the energy due to triples and quadruples.

Thus it is proposed to split the virtual NO space into two distinct parts. The CISD part of the configuration space will include all of the NOs (or, in cases where only valence electron correlation is being considered, corelike NOs may be deleted). The triple and quadruple excitations, however, will be constrained to the lower portion of the virtual NO space, which will consist of less than half the virtual NOs. Since the number of triple and quadruple excitations is always very large compared to the number of singles and doubles, this will result in a drastic decrease in the number of configurations included in the CISD[TQ] wave function.

We suspect that such a split virtual NO approach to the CISD[TQ] method will yield most of the benefits of the CISD[TQ] method at a cost not tremendously higher than the simple CISD method. We hope to implement this method within the GUGA formalism and to compare the new method with such high level levels of correlation as CCSDT, CISDTQ, and SOCI.

TABLE VI. Energy loss (hartrees) due to deletion of high-lying NOs for H_7^+ with the CISD and the CISDTQ methods.[a]

Orbitals n	CISD $E_n - E_{35}$	CISDTQ $E_n - E_{35}$	TQ loss[b]
35	0.000 000	0.000 000	0.000 000(0.0)
34	0.000 048	0.000 050	0.000 002(0.0)
33	0.000 080	0.000 084	0.000 004(0.1)
32	0.000 134	0.000 145	0.000 011(0.3)
31	0.000 222	0.000 244	0.000 022(0.5)
30	0.000 421	0.000 459	0.000 037(0.9)
29	0.000 740	0.000 806	0.000 067(1.6)
28	0.001 124	0.001 209	0.000 086(2.1)
27	0.001 537	0.001 644	0.000 107(2.6)
26	0.002 009	0.002 150	0.000 140(3.4)
25	0.002 462	0.002 613	0.000 151(3.6)
24	0.002 590	0.002 747	0.000 156(3.8)
23	0.003 083	0.003 270	0.000 187(4.5)
22	0.003 275	0.003 478	0.000 203(4.9)
21	0.003 839	0.004 061	0.000 222(5.3)
20	0.004 259	0.004 503	0.000 244(5.9)
19	0.004 772	0.005 044	0.000 273(6.6)
18	0.005 368	0.005 685	0.000 317(7.6)
17	0.005 893	0.006 249	0.000 356(8.6)
16	0.007 060	0.007 521	0.000 462(11.1)
15	0.008 225	0.008 784	0.000 559(13.4)
14	0.012 462	0.013 261	0.000 799(19.2)
13	0.016 262	0.017 249	0.000 987(23.7)
12	0.022 782	0.024 082	0.001 300(31.2)
11	0.024 603	0.025 986	0.001 383(33.2)
10	0.031 758	0.033 506	0.001 748(42.0)
9	0.035 454	0.037 391	0.001 937(46.5)
8	0.045 381	0.047 813	0.002 432(58.4)
7	0.052 259	0.054 972	0.002 713(65.2)

[a] Energies calculated at the FCI optimized geometry using CISD natural orbitals.

[b] Energy loss relative to E_{35}(CISDTQ) due to triples and quadruples involving the deleted orbitals. The percentage relative to E_{35}(CISD)$-E_{35}$(CISDTQ)$=0.004$ 164 hartree is found in parentheses.

CONCLUSIONS

In this work, we present further evidence that treatment of electron correlation in molecular systems to extents which approach the quality of the full CI method does not necessarily require computations which approach the full CI in size. When computational resources become a factor in the extent to which a particular problem can be addressed, proper construction of such a wave function is particularly important. It is seen that the *a priori* selection of references by excitation class in a MRCI(SD) expansion leading to a CISD[TQ] wave function is an accurate approximation applicable to a broad range of systems, especially if the suggested limitations on the virtual space are implemented. For even greater accuracy, one can construct a different MRCI(SD) wave function based on the CSFs found to be important in the CISD[TQ] expansion.

ACKNOWLEDGMENTS

This research was financially supported by the National Science Foundation, Grant No. CHE-8718469. The authors would like to thank Y. Yamaguchi for helpful discussions and for providing the CAS SCF code used, C. Bauschlicher for assisting with the problems reproducing his energy calculations on H_2O, and B. J. DeLeeuw for thorough proofing of the manuscript. Special recognition is due R. S. Grev for innumerable helpful discussions. This material is based upon work supported under a National Science Foundation Graduate Fellowship as well as a Department of Defense Graduate Fellowship. The Quantum Trio thank K., A., and Em. for hearing about it a lot.

[1] P. Saxe, H. F. Schaefer, and N. C. Handy, Chem. Phys. Lett. **79**, 202 (1981); C. W. Bauschlicher, S. R. Langhoff, P. R. Taylor, and H. Partridge, *ibid.* **126**, 436 (1986); C. W. Bauschlicher, S. R. Langhoff, N. C. Handy, and P. J. Knowles, J. Chem. Phys. **85**, 1469 (1986); C. W. Bauschlicher and S. R. Langhoff, *ibid.* **86**, 5595 (1987).

[2] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **85**, 2779 (1986).

[3] I. Shavitt, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer (Plenum, New York, 1977), Vol. 3, pp. 189–275.

[4] C. W. Bauschlicher, S. R. Langhoff, and P. R. Taylor, J. Chem. Phys. **88**, 2540 (1988).

[5] E. R. Davidson, Rev. Mod. Phys. **44**, 451 (1972).

[6] H. F. Schaefer, Ph.D. thesis, Department of Chemistry, Stanford University, April, 1969.

[7] J. Almlöf, B. J. DeLeeuw, P. R. Taylor, C. W. Bauschlicher, and P. Siegbahn, Int. J. Quantum Chem. Symp. **23**, 345 (1989).

[8] R. S. Grev and H. F. Schaefer, J. Chem. Phys. **96**, 6850 (1992).

[9] P. Saxe, D. J. Fox, H. F. Schaefer, and N. C. Handy, J. Chem. Phys. **77**, 5584 (1982); B. R. Brooks and H. F. Schaefer, J. Chem. Phys. **70**, 5092 (1979).

[10] P. O. Löwdin, Phys. Rev. **97**, 1474 (1955); E. R. Davidson, *Reduced Density Matrices in Quantum Chemistry* (Academic, New York, 1976).

[11] P. Pulay, Mol. Phys. **17**, 197 (1969); in *Modern Theoretical Chemistry*, edited by H. F. Schaefer (Plenum, New York, 1977), Vol. 4, pp. 153–185.

[12] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).

[13] T. H. Dunning, J. Chem. Phys. **53**, 2823 (1970).

[14] F. B. van Duijneveldt, IBM Research Report RJ 945, 1971.

[15] We correct the contraction coefficient of the third primitive Gaussian function in the $1s$-like orbital on the oxygen atom to 0.073 771.

[16] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982); G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, *ibid.* **89**, 7382 (1988).

[17] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[18] M. R. Hoffmann and H. F. Schaefer, Adv. Quant. Chem. **18**, 207, (1986); J. Noga and R. J. Bartlett, J. Chem. Phys. **86**, 7041 (1987); **89**, 3401(E) (1988); G. E. Scuseria and H. F. Schaefer, Chem. Phys. Lett. **152**, 382 (1988).

[19] T. J. Lee and P. R. Taylor, Int. J. Quantum Chem. Symp. **23**, 199 (1989).

[20] R. E. Watson, Phys. Rev. **119**, 170 (1960).

[21] O. Sinanoğlu, Proc. Natl. Acad. Sci. **47**, 492 (1961).