Theor Chem Acc (1997) 97:164–176

Theoretical
Chemistry Accounts

Multiple basis sets in calculations of triples corrections in coupled-cluster theory

Wim Klopper[1], **Jozef Noga**[2], **Henrik Koch**[3], **Trygve Helgaker**[1]

[1] Department of Chemistry, University of Oslo, N-0315 Oslo, Norway
[2] Institute of Inorganic Chemistry, Slovak Academy of Sciences, SK-84236 Bratislava, Slovakia
[3] Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark

Received: 28 December 1996 / Accepted: 8 April 1997

Abstract. Multiple basis sets are used in calculations of perturbational corrections for triples replacements in the framework of single-reference coupled-cluster theory. We investigate a computational procedure, where the triples correction is calculated from a reduced space of virtual orbitals, while the full space is employed for the coupled-cluster singles-and-doubles model. The reduced space is either constructed from a prescribed unitary transformation of the virtual orbitals (for example into natural orbitals) with subsequent truncation, or from a reduced set of atomic basis functions. After the selection of a reduced space of virtual orbitals, the singles and doubles amplitudes obtained from a calculation in the full space are projected onto the reduced space, the remaining set of virtual orbitals is brought into canonical form by diagonalizing the representation of the Fock operator in the reduced space, and the triples corrections are evaluated as usual. The case studies include the determination of the spectroscopic constants of N_2, F_2, and CO, the geometry of O_3, the electric dipole moment of CO, the static dipole polarizability of F^-, and the Ne$\cdots$Ne interatomic potential.

Key words: Coupled-cluster theory – Triples corrections – Multiple basis sets – Natural orbitals – Integral-direct methods

1 Introduction

The basis-set convergence of calculated electron-correlation effects is a critical issue in computational methods of molecular electronic-structure ab initio theory. Very recently, we performed systematic calculations on the water molecule, investigating the basis-set convergence for Hartree-Fock self-consistent field theory (SCF), for perturbation theory at the second-order Møller-Plesset level (MP2), for coupled-cluster theory at the singles-and-doubles level (CCSD), and for coupled-cluster theory at the singles-and-doubles level including perturbational corrections for connected triples [CCSD(T)] [1].

The results of this study for the correlation-consistent core-valence basis sets CVXZ with $X = 2, 3, \ldots 6$ (X is sometimes referred to as the "cardinal number") are presented in Table 1. The largest basis set (CV6Z) corresponds to a $[12s11p9d7f5g3h1i]$ contraction for O and a $[6s5p4d3f2g1h]$ contraction for H, amounting to 412 atomic basis functions.

On an absolute scale, the basis-set truncation error of the CV6Z basis is relatively small both at the SCF level (0.3 mE_h) and for the CCSD(T) triples correction (0.2 mE_h), but it is substantial for the CCSD correlation energy (4.8 mE_h). For comparison, the corresponding errors are 2.6, 0.9, and 15.1 mE_h for the CVQZ basis. It seems that the CCSD correlation energy, in particular, suffers from the well-known slow basis-set convergence of correlation energies. For the triples, the convergence appears to be faster, in particular with respect to a threshold of chemical accuracy.

Thus, concerning the total CCSD(T) energy of H_2O, if the goal were to achieve a prescribed accuracy of – say – 1 mE_h, it would be sufficient to employ the CV5Z basis set for both the SCF calculation and the CCSD(T) triples correction. However, the CCSD calculation would require a much larger basis set. Based on a two-parameter fit of the form $E = A + BX^{-3}$ to the computed CCSD energies for $3 \leq X \leq 6$, we expect that the CCSD truncation error of the (nonexisting) CVXZ basis set with cardinal number $X = 10$ would be around the target accuracy of 1 mE_h [1]. This basis set would consist of as many as 1660 functions.

The difference in behavior of the doubles and triples amplitudes indicates that it may be advantageous to consider the use of different orbital spaces for these amplitudes.

Jurgens-Lutovsky and Almlöf have proposed exploiting the fact that SCF and MP2 calculations require basis sets of different size and quality, and have suggested using "dual basis sets in calculations of electron correlation" [2]. They employed a small space of orbitals for the SCF calculation (S_1) and a larger space for the MP2

Correspondence to: W. Klopper

Table 1. H_2O molecule:[a] SCF energy, CCSD correlation energy, and CCSD(T) triples correction (ΔE_T) obtained from calculations using correlation-consistent core-valence basis sets

Basis set	N_{vir} [b]	E_{SCF} (E_{h})	ΔE_{CCSD} (E_{h})	ΔE_{T} (E_{h})
CVDZ	23	−76.027 2	−0.251 8	−0.003 3
CVTZ	66	−76.057 4	−0.324 2	−0.008 2
CVQZ	139	−76.065 0	−0.346 5	−0.009 5
CV5Z	250	−76.067 1	−0.353 3	−0.010 0
CV6Z[c]	407	−76.067 3	−0.356 8	−0.010 2
Extrapolated limit [1]		−76.067 6	−0.361 6	−0.010 4

[a] Geometry: $R_{\mathrm{O-H}} = 1.80885\ a_0, \angle_{\mathrm{H-O-H}} = 104.52^\circ$

[b] Number of virtual orbitals

[c] The CV6Z basis set was developed in Ref. [1]

correlation energy (S_2), with S_1 chosen as a subset of S_2. In the present study, we explore the possibility of using yet a different orbital space for the triples correction (for obvious reasons, we will denote this space as S_3). We expect that this space can be chosen smaller than the S_2 space, and that the use of "multiple basis sets" in CCSD(T) calculations will result in efficient computational models.

The usefulness of multiple basis sets should be analyzed in view of the "integral-direct" methods [3, 4]. For example, the dual basis sets of Jurgens-Lutovsky and Almlöf are very efficient in integral-direct MP2 calculations [5–8], where the number of two-electron integrals is much reduced, but the computational scheme would not be useful in integral-direct CCSD calculations [9–11]. On the other hand, the integral-direct formulation of CCSD theory allows us to employ very large basis sets of atomic functions at this level, whereas the use of extended basis sets is much less straightforward for the computation of the triples corrections. More flexible approaches are needed.

In this paper, we present our preliminary results obtained from using a small orbital space S_3 for the CCSD(T) triples correction. We have investigated the use of different sets of atomic basis functions, for instance the atomic natural orbitals (ANO) basis sets of Almlöf and Taylor [12, 13], as well as several truncations of the space of the molecular virtual orbitals, for example by selecting orbitals with respect to the occupation numbers of (molecular) natural orbitals (NO) that were determined at the MP2 level.

The history of optimizations of the virtual space for the computation of electron-correlation effects is rich. One example is the expansion in terms of natural (spin)orbitals introduced by Löwdin [14], and approximative NOs have been applied with great success in configuration interaction (CI) and coupled-electron pair approximation (CEPA) calculations in the 1970s [15–19]. Another category of virtual orbitals is related to certain effective potentials, for example to the Hartree-Fock potential of the positively charged ion of the system under study [20–32]. The third example is the approach developed at the end of the 1980s by Adamowicz and Bartlett, who optimized the virtual space by minimizing the Hylleraas functional of MP2 theory [33–38]. The different optimizations of active virtual spaces have been reviewed quite recently by Wasilewski et al. [39, 40]. We note, however, that the goal of these past studies was to find short expansions for the whole correlated energy. In contrast, the present paper is concerned with the efficient expansion of the connected triples replacements in the framework of coupled-cluster theory, and it is based on the observation that the connected triples corrections appear to converge faster to within chemical accuracy than the doubles amplitudes. In a recent and similar study, Sherrill and Schaefer have investigated variational wave functions that incorporate limited triples and quadruples substitutions [41].

The present study concentrates on the CCSD(T)-type triples correction, which is the perturbational correction that is most commonly used today [42]. Of course, the reduced virtual space approach can be applied to other types of perturbational corrections as well, for example to the CCSD + T(CCSD) or CCSD-T corrections [43–45]. Furthermore, methods where the triples amplitudes enter the computation in an iterative manner, as in the full singles-and-doubles-and-triples model (CCSDT), could also take great advantage of the reduced virtual space approach, not least in view of the storage and input/output (I/O) bottlenecks that occur when the triples amplitudes are stored on external disk.

We refer to Ref. [46] for a recent overview of coupled-cluster theory, and we only summarize the working formulae for the CCSD(T) triples correction in the next section. Section 3.1 describes a few possible ways of constructing reduced spaces of virtual orbitals, for example in terms of canonical virtual orbitals, improved virtual orbitals, MP2 NOs, or optimized virtual orbitals. The multiple basis sets of atomic functions are described in Sect. 3.2. Section 4 is concerned with the computational aspects, for example with the projection of the CCSD amplitudes obtained in S_2 onto the reduced space S_3. Our preliminary results are reported in Sect. 5 and some general conclusions are presented in Sect. 6.

2 Perturbational triples corrections

The CCSD(T) perturbational triples correction [42] to the CCSD energy is the most widely used approximate triples model and is found to be very robust and stable in determining molecular structures [47]. However, with a computational scaling of O^3V^4 (where O is the number of occupied orbitals and V is the number of virtual orbitals), the method becomes rather demanding for large systems. In the following, we investigate the possible savings obtainable using a smaller set of virtual orbitals in the conventional [48, 49] and integral-direct [50, 51] CCSD(T) models.

The CCSD wave function for a closed-shell system is given by the *ansatz*

$$|\mathrm{CC}\rangle = \exp(T)|\mathrm{HF}\rangle \ , \tag{1}$$

where the cluster operator T is

$$T = T_1 + T_2 \ . \tag{2}$$

The connected singles $\{t_i^a\}$ and doubles $\{t_{ij}^{ab}\}$ cluster amplitudes enter T as

$$T_1 = \sum_{ai} t_i^a E_{ai} \ , \tag{3}$$

$$T_2 = \tfrac{1}{2}\sum_{aibj} t_{ij}^{ab} E_{ai} E_{bj} \ , \tag{4}$$

with $t_{ij}^{ab} = t_{ji}^{ba}$. The operators $\{E_{ai}\}$ are the unitary group generators

$$E_{ai} = a_{a\uparrow}^{\dagger} a_{i\uparrow} + a_{a\downarrow}^{\dagger} a_{i\downarrow} \ . \tag{5}$$

Here and in the following, the labels $i, j, k, \ldots$ and $a, b, c, \ldots$ are used for occupied and virtual canonical orbitals, respectively. The Mulliken notation is used for the two-electron integrals, and $\varepsilon_i, \varepsilon_j, \varepsilon_k, \ldots, \varepsilon_a, \varepsilon_b, \varepsilon_c, \ldots$ denote Hartree-Fock orbital energies.

The expression for the triples correction is given by Lee et al. [48, 49] as

$$\Delta E_T = \tfrac{1}{3}\sum_{ijk}\sum_{abc}\left(4W_{ijk}^{abc} + W_{ijk}^{bca} + W_{ijk}^{cab}\right)\left(V_{ijk}^{abc} - V_{ijk}^{cba}\right)/D_{ijk}^{abc} , \tag{6}$$

where

$$W_{ijk}^{abc} = P_{ijk}^{abc}\left[\sum_d (bd|ai) t_{kj}^{cd} - \sum_l (ck|jl) t_{il}^{ab}\right] \ , \tag{7}$$

$$V_{ijk}^{abc} = W_{ijk}^{abc} + (bj|ck) t_i^a + (ai|ck) t_j^b + (ai|bj) t_k^c \ , \tag{8}$$

$$D_{ijk}^{abc} = \varepsilon_i + \varepsilon_j + \varepsilon_k - \varepsilon_a - \varepsilon_b - \varepsilon_c \ . \tag{9}$$

P_{ijk}^{abc} is a permutation operator defined as

$$\begin{aligned} P_{ijk}^{abc}\begin{pmatrix} abc \\ ijk \end{pmatrix} &= \begin{pmatrix} abc \\ ijk \end{pmatrix} + \begin{pmatrix} bac \\ jik \end{pmatrix} + \begin{pmatrix} cba \\ kji \end{pmatrix} \\ &\quad + \begin{pmatrix} acb \\ ikj \end{pmatrix} + \begin{pmatrix} bca \\ jki \end{pmatrix} + \begin{pmatrix} cab \\ kij \end{pmatrix} \ . \end{aligned} \tag{10}$$

The O^3V^4 and O^4V^3 computational scalings appear in the construction of the intermediate W_{ijk}^{abc} in Eq. (7), whereas the other contractions scale as O^3V^3. Thus, if a smaller set of virtual orbitals is used for the triples correction, we obtain a large saving due to the fourth power dependence on the number of virtual orbitals. If the ratio between the full and reduced virtual spaces is about 3, the computational work will be reduced by a factor of 81. Judging from the results presented in the present work, a reduction by a factor of 3 is not unrealistic for large basis sets, and may be seen as a consequence of the faster convergence of the triples correction in comparison with the energy contribution from doubles replacements. In the conventional approach, the amount of I/O operations is also reduced, and because these scale as O^3V^3, the reduction in I/O may be viewed as equally important as the reduction of the computational effort.

In the integral-direct approach, the same reduction as for the conventional method is obtained concerning the computational effort, since the calculation is organized in terms of multiple passes over the integrals. The total amount of I/O scales as OV^2N (where N is the number of atomic basis functions), and this is negligible compared to the computational effort. To summarize, we expect that the use of a reduced virtual space in our integral-direct CCSD(T) implementation will considerably increase the application range of the method.

3 Multiple basis sets

3.1 Reduced virtual orbital spaces

The virtual orbital space can be reduced straightforwardly by deleting a selected number of canonical SCF orbitals, for example those with orbital energies larger than a predefined threshold. However, this approach is probably not the most efficient one, and other reduced spaces of virtual orbitals should be investigated.

We proceed by allowing the virtual orbitals to rotate among themselves to form some kind of optimal set of correlating orbitals. Once this is achieved, the least important orbitals are removed and the Fock operator is transformed into the basis of the remaining orbitals and then diagonalized. In this manner, the final active virtual orbital space consists yet again of canonical orbitals. The associated orbital energies are also obtained. Thus, a formalism based on canonical orbitals and their orbital energies can be utilized to compute the CCSD(T) triples correction, even though the reduced virtual orbital space is defined in terms of a set of noncanonical orbitals.

In the present work, we study three types of virtual orbitals:

1. Canonical virtual orbitals (CVO): These are the usual virtual SCF orbitals obtained by diagonalizing the all-electron Fock operator **F**.
2. Improved virtual orbitals (IVO): These orbitals are obtained by diagonalizing the virtual-virtual block of a Fock matrix $\mathbf{F}^{\mathrm{core}}$ that is built from a prescribed number of core orbitals [28],

$$F_{ab}^{\mathrm{core}} = h_{ab} + \sum_i^{\mathrm{core}} \left[2(ii|ab) - (ia|ib)\right] \ , \tag{11}$$

   where h is the one-electron Hamiltonian.

3. MP2 natural orbitals (MP2 NO): These orbitals are obtained by diagonalizing the virtual-virtual block of the MP2 one-electron density matrix [52],

$$D_{ab} = \sum_i \sum_j \sum_c \frac{\sqrt{2}(ib|jc)\left[2(ia|jc) - (ic|ja)\right]}{(\varepsilon_i + \varepsilon_j - \varepsilon_c - \varepsilon_b)(\varepsilon_i + \varepsilon_j - \varepsilon_c - \varepsilon_a)} \ . \tag{12}$$

   We can sum either over all occupied orbitals, including the core, or over the valence orbitals only, without further modifications. We have not considered an alternative approach, where the whole density matrix (not only the virtual-virtual block) is diagonalized. In that case, the weakly occupied natural orbitals would have to be orthogonalized to the occupied SCF orbitals (orthogonalized natural orbitals, ONO). We also note that, for our exploratory study, we have chosen the MP2 NOs for the sake of simplicity. At a later stage of the study, we would like to investigate the CCSD NOs.

A comparison of the three types of orbitals is presented in Fig. 1 for a calculation of the CCSD(T) triples correction of the H_2O molecule in the VQZ basis. The $1a_1$ core orbital was not correlated in the CCSD(T) calculation, and the $\mathbf{F}^{\text{core}}$ operator was built from this orbital. Hence, $\mathbf{F}^{\text{core}}$ is the Fock operator of the H_2O^{8+} cation constructed from the lowest occupied orbital of the neutral molecule. The CVOs and IVOs were ordered with respect to increasing eigenvalues of the $\mathbf{F}$ and $\mathbf{F}^{\text{core}}$ operators, respectively, and a predefined number of orbitals with the highest eigenvalues were removed. The MP2 NOs, in contrast, were ordered with respect to decreasing eigenvalues of $\mathbf{D}$ – that is, with respect to the occupation numbers – and a predefined number of orbitals with the lowest eigenvalues were removed. The MP2 density matrix was obtained by summing over all occupied orbitals in Eq. (12).

It is demonstrated in Fig. 1 that the MP2 NOs provide the most effective expansion of the virtual space. With only 50% of the original number of virtual orbitals, about 95% of the triples correction to the energy is recovered by the MP2 NOs. We further observe that a very smooth, S-shaped curve is obtained with the MP2 NOs, whereas the CVOs display a considerable structure, converging less smoothly. The IVO curve is not as smooth as the MP2 NO curve, but nevertheless represents a noticeable improvement on the CVOs.

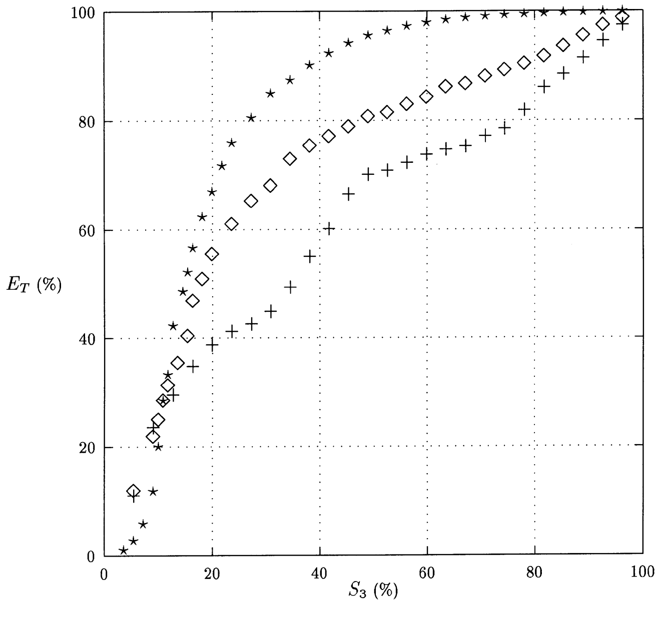

Fig. 1. Valence shell CCSD(T) triples correction (E_T in %) obtained from calculations employing canonical virtual orbitals (+), improved virtual orbitals (◇), and MP2 natural orbitals (⋆), as a function of the number of orbitals in S_3 (in %). Results for H_2O using the VQZ basis set

Finally, the space of virtual orbitals can also be optimized by the methods developed by Adamowicz and Bartlett about a decade ago [33–38]. These authors investigated the optimized virtual orbitals (OVO). These orbitals are obtained by minimizing – by Newton's method, for example – the Hylleraas second-order energy functional of Møller-Plesset perturbation theory.

We have not considered the OVOs in the present work since these orbitals appear to be somewhat biased towards an optimal description of doubles replacements. Adamowicz and Bartlett reported that the poor description of the triples replacements is a negative aspect of the OVOs [33]. Still, the OVOs have been used for the computation of triples corrections in previous work. For example, an accurate CCSD + T(CCSD) correlation energy for the HF molecule was obtained by computing the T(CCSD) triples correction using 56 OVOs ($-0.0089\ E_{\text{h}}$) and adding this correction to the CCSD energy computed in the full space of 75 virtual orbitals ($-0.3673\ E_{\text{h}}$) [34]. Furthermore, we note that Adamowicz has studied the minimization of that part of the fourth-order energy functional that involves the singles replacements (the usual $E_{\text{S}}^{(4)}$ energy contribution of fourth-order Møller-Plesset theory) in order to obtain OVOs for cases where singles replacements are important [38]. Obviously, the most promising procedure for generating OVOs with respect to the triples corrections would consist of the minimization of the triples part of the fourth-order energy functional – that is, $E_{\text{T}}^{(4)}$. Of course, such a procedure would not lead to any computational savings if we are interested only in perturbational triples corrections. It could perhaps become attractive in conjunction with the full CCSDT model. At this level, the perturbational optimization (at the MP4 level) of the virtual orbitals for an efficient description of triples replacements would be fully analogous to the MP2-level optimization of the OVOs for use in CCSD calculations [33]. Preferably, one would consider a "fourth-order functional" based on the converged $\{t_{ij}^{ab}\}$ amplitudes of the CCSD calculation, in this manner optimizing the perturbational triples correction itself at the fourth-order level of perturbation theory.

3.2 Multiple AO basis sets

In the preceding section, we discussed a number of possible ways of reducing and optimizing the space of virtual orbitals, based on rotations among the virtual orbitals followed by a truncation of the space. Alternatively, we can start with a separate basis set of atomic orbitals (AO) for the computation of the triples corrections, orthogonalized to the occupied SCF orbitals. In principle, the AO basis set for the triples correction (S_3) can be chosen completely independent from the AO basis set used for the CCSD calculation (S_2), but it is advantageous to choose S_3 as a subset of S_2 since the orthogonalization procedure will yield a final set of virtual orbitals that is expanded in the union of the two AO basis sets. Hence, we have restricted our study to the use of true subsets of larger AO basis sets for the computation of the triples corrections.

A seeming advantage of the multiple AO basis sets approach is that no thresholds are required for the selection of the reduced space. Furthermore, the method provides the flexibility of selecting the type of basis functions that is expected to be important for the triples correction – diffuse or tight functions, low or high angular momentum functions, and so on. However, one must be careful not to choose a subset that is linearly dependent on the occupied space (numerically, linear dependence occurs with respect to a predefined threshold and the multiple AO basis sets method is consequently not completely free of thresholds). For example, it is anticipated that in some cases, it may be difficult or even impossible to orthogonalize the innermost *s*-orbitals of atomic basis sets of the ANO or correlation-consistent type to the occupied SCF orbitals of the molecule under study.

The following example illustrates the use of multiple AO sets. Let us consider the correlation-consistent VTZ basis set for the H_2O molecule, which constitutes the S_2 space. It consists of $4s3p2d1f$ and $3s2p1d$ contractions for O and H, respectively, and contains 58 AO basis functions. We choose the S_3 space by removing the f-type functions from the VTZ basis set and denote the remaining basis as VTZ′. This basis contains 51 functions. However, we cannot construct 51 virtual orbitals that are orthogonal to the occupied space. Of the 7 f-functions that are removed, two belong to the a_1 irreducible representation in C_{2v} symmetry, one belongs to a_2, two belong to b_1, and two belong to b_2. The VTZ′ basis will therefore be linearly dependent on the space spanned by the occupied orbitals, since three a_1 orbitals are occupied but only two are removed from the S_2 space. Therefore, one extra a_1 orbital is deleted by the orthogonalization procedure and the final (orthogonal) reduced virtual space consists of 50 orbitals.

3.3 Thresholds

It might be argued that the use of thresholds for the selection of virtual orbitals prohibits the computation of smooth potential energy surfaces, or that it would introduce complications concerning the computation of analytic first and second derivatives of the energy and properties in general. This is not so. The way we proceed is that we first determine the reduced virtual space for the molecule at a prescribed geometry (e.g., at its equilibrium). In all subsequent computations of the potential energy curve or surface, the number of virtual orbitals within each irreducible representation is kept fixed at the number determined in the exploratory calculation. Thus, the use of thresholds, for example for the selection of MP2 NOs or the orthogonalization procedure, constitutes no problems whatsoever.

4 Computational details

In this section we briefly discuss the implementation of the multiple basis sets. We write the cluster operator in terms of orthonormal singlet-excited configurations,

$$T_1|\mathrm{HF}\rangle = \sum_{ai} c_i^a |\Psi_i^a\rangle \ , \tag{13}$$

$$\begin{aligned} T_2|\mathrm{HF}\rangle &= \sum_{i\le j}\sum_{a\le b} c_{ij}^{ab}(+)|\Psi_{ij}^{ab}(+)\rangle \\ &\quad + \sum_{i<j}\sum_{a<b} c_{ij}^{ab}(-)|\Psi_{ij}^{ab}(-)\rangle \ , \end{aligned} \tag{14}$$

where

$$|\Psi_i^a\rangle = \tfrac{1}{\sqrt{2}} E_i^a |\mathrm{HF}\rangle \ , \tag{15}$$

$$|\Psi_{ij}^{ab}(+)\rangle = \tfrac{1}{2}(1+\delta_{ij})^{-\frac{1}{2}}(1+\delta_{ab})^{-\frac{1}{2}}(E_{ai}E_{bj}+E_{bi}E_{aj})|\mathrm{HF}\rangle \ , \tag{16}$$

$$|\Psi_{ij}^{ab}(-)\rangle = \tfrac{1}{2\sqrt{3}}(E_{ai}E_{bj}-E_{bi}E_{aj})|\mathrm{HF}\rangle \ . \tag{17}$$

The singles and doubles amplitudes that were computed using the full space $\{a\}$ can be projected onto the reduced virtual space $\{\bar{a}\}$ by evaluating the overlap between the orthogonal spin-adapted configurations. For the singles amplitudes we get

$$t_i^{\bar{a}} = \tfrac{1}{\sqrt{2}} c_i^{\bar{a}} = \tfrac{1}{\sqrt{2}} \sum_a c_i^a \langle \Psi_i^{\bar{a}} | \Psi_i^a \rangle = \tfrac{1}{\sqrt{2}} \sum_a c_i^a S_{\bar{a}a} = \sum_a t_i^a S_{\bar{a}a} \ . \tag{18}$$

The overlap between doubly excited configurations is

$$\begin{aligned} \langle \Psi_{ij}^{\bar{a}\bar{b}}(+)|\Psi_{ij}^{ab}(+)\rangle &= (1+\delta_{\bar{a}\bar{b}})^{-\frac{1}{2}}(1+\delta_{ab})^{-\frac{1}{2}} \\ &\quad \times (S_{\bar{a}a}S_{\bar{b}b} + S_{\bar{a}b}S_{\bar{b}a}) \ , \end{aligned} \tag{19}$$

$$\langle \Psi_{ij}^{\bar{a}\bar{b}}(-)|\Psi_{ij}^{ab}(-)\rangle = S_{\bar{a}a}S_{\bar{b}b} - S_{\bar{a}b}S_{\bar{b}a} \ , \tag{20}$$

and thus, noting that

$$c_{ij}^{ab}(+) = c_{ij}^{ba}(+), \qquad c_{ij}^{ab}(-) = -c_{ij}^{ba}(-) \ , \tag{21}$$

the projection of the doubles amplitudes becomes

$$\begin{aligned} c_{ij}^{\bar{a}\bar{b}}(+) &= \sum_{a\le b} c_{ij}^{ab}(+)\langle \Psi_{ij}^{\bar{a}\bar{b}}(+)|\Psi_{ij}^{ab}(+)\rangle \\ &= \sum_{ab} c_{ij}^{ab}(+)\left(\frac{1+\delta_{ab}}{1+\delta_{\bar{a}\bar{b}}}\right)^{\frac{1}{2}} S_{\bar{a}a}S_{\bar{b}b} \ , \end{aligned} \tag{22}$$

$$c_{ij}^{\bar{a}\bar{b}}(-) = \sum_{a<b} c_{ij}^{ab}(-)\langle \Psi_{ij}^{\bar{a}\bar{b}}(-)|\Psi_{ij}^{ab}(-)\rangle = \sum_{ab} c_{ij}^{ab}(-)S_{\bar{a}a}S_{\bar{b}b}. \tag{23}$$

We now obtain the following simple expression for the projected doubles amplitudes,

$$\begin{aligned} t_{ij}^{\bar{a}\bar{b}} &= \tfrac{1}{2}(1+\delta_{ij})^{\frac{1}{2}}(1+\delta_{\bar{a}\bar{b}})^{\frac{1}{2}} c_{ij}^{\bar{a}\bar{b}}(+) + \tfrac{1}{2\sqrt{3}} c_{ij}^{\bar{a}\bar{b}}(-) \\ &= \sum_{ab} t_{ij}^{ab} S_{\bar{a}a}S_{\bar{b}b} \ . \end{aligned} \tag{24}$$

It thus appears that the $\{t_i^a\}$ and $\{t_{ij}^{ab}\}$ amplitudes that occur in the coupled-cluster approach where the Schrödinger equation is projected onto a bi-orthogonal basis can be computed in a straightforward manner from a two-index transformation. The transformation of the $\{c_i^a\}$ and $\{c_{ij}^{ab}\}$ amplitudes, however, requires some care with respect to the normalization of the "diagonal" terms where $a = b$.

We collect the orbital coefficients of the full set of canonical virtual orbitals in a matrix $\mathbf{V}$, the coefficients of the reduced set of virtual orbitals in a matrix $\bar{\mathbf{V}}$, and the singles and doubles amplitudes in the vectors $\mathbf{T}_1^i$ and matrices $\mathbf{T}_2^{ij}$, respectively,

$$(\mathbf{T}_1^i)_a = t_i^a, \qquad (\mathbf{T}_2^{ij})_{ab} = t_{ij}^{ab} \ . \tag{25}$$

The projection of the CCSD amplitudes onto S_3 is then given by

$$\bar{\mathbf{T}}_1^i = \bar{\mathbf{V}}^\dagger \mathbf{S} \mathbf{V} \mathbf{T}_1^i \ , \tag{26}$$

$$\bar{\mathbf{T}}_2^{ij} = \bar{\mathbf{V}}^\dagger \mathbf{S} \mathbf{V} \mathbf{T}_2^{ij} \mathbf{V}^\dagger \mathbf{S} \bar{\mathbf{V}} \ , \tag{27}$$

where $\mathbf{S}$ is the overlap matrix. Hence, the computational procedure consists of transforming the singles and doubles amplitudes, followed by computing the triples correction as usual from the transformed amplitudes.

5 Results and discussion

The calculations were performed on IBM RS/6000 590 workstations with the program DIRCCR12-95 [53] using correlation-consistent [54–59] and ANO basis sets [60, 61]. Note that, in order to save space, we use the abbreviations VXZ, CVXZ, and AVXZ for the cc-pVXZ, cc-pCVXZ, and aug-cc-pVXZ basis sets, respectively, where X = D, T, Q, 5, 6. For the computation of the electric properties we have employed the doubly and triply augmented basis sets, which are denoted as d-AVXZ and t-AVXZ, respectively [57].

The following atomic masses were used: $M(\mathrm{C}^{12}) = 12.000\,000\,000$ amu, $M(\mathrm{N}^{14}) = 14.003\,074\,008$ amu, $M(\mathrm{O}^{16}) = 15.994\,914\,640$ amu, $M(\mathrm{F}^{19}) = 18.998\,403\,250$ amu, and $M(\mathrm{Ne}^{20}) = 19.992\,439\,100$ amu [62].

5.1 Calculations in a reduced AO *basis:* N_2 *and* Ne_2

We first consider the use of reduced AO sets for the calculation of the triples correction. In Table 2, we have listed for N_2 the bond distances, harmonic vibrational frequencies, and electronic energies obtained with a variety of S_2 and S_3 spaces. There are two sets of calculations in this table: one set of calculations based on the [*6s5p4d3f*] ANO basis of Widmark et al. and another set based on the AVQZ basis of Dunning and coworkers.

For the ANO calculations, we have employed three basis sets denoted ANO_1, ANO_2, and ANO_3. The ANO_3 basis corresponds to the full [*6s5p4d3f*] ANO basis. The [*5s4p3d2f*] ANO_2 basis has been generated from ANO_3 by the removal of one set of functions (with lowest occupation) for each angular momentum. Finally, by removing one more set of functions, we arrive at the smallest [*4s3p2d1f*] basis denoted ANO_1. We note that these basis sets constitute a hierarchy of sets where the smaller ones are true subsets of the larger ones.

Two sets of calculations have been carried out with these basis sets. First, we calculated the CCSD energies for all three basis sets. Next, we calculated the CCSD(T) energies, employing only the largest ANO_3 set for the CCSD wave function and the basis sets ANO_1, ANO_2, and ANO_3 for the triples correction.

We first note that for the largest basis set ANO_3, the bond distance is 109.36 pm at the CCSD level and 110.08 pm at the CCSD(T) level. The corresponding numbers for the harmonic frequency are 2429.7 and 2349.8 cm^{-1}. Clearly, the triples corrections are quite significant: 0.72 pm and -79.9 cm^{-1}. Next, we observe that triples corrections are only moderately affected by the reductions in the AO space. In the ANO_1 S_3 basis, the triples corrections are 0.69 pm and -77.5 cm^{-1} – that is, only 0.03 pm and 2.4 cm^{-1} smaller in magnitude than the full corrections. The corresponding errors in the ANO_2 basis are 0.01 pm and 0.7 cm^{-1}. These differences are much smaller than those observed at the CCSD level, where for example the ANO_1 basis gives a bond length that is 0.68 pm too long and and a frequency that is 11.6 cm^{-1} too low relative to the ANO_3 basis. We also note that the calculation of the triples correction with

Table 2. Results for the N_2 molecule: equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) obtained from coupled-cluster calculations using multiple AO basis sets. The 1*s* core orbitals were not correlated

	S_2	S_3	r_e (pm)	ω_e (cm^{-1})	E_{min} (E_h)	N_{vir}[a]
CCSD	*4s3p2d1f*		110.04	2418.1	−109.351 873	
CCSD	*5s4p3d2f*		109.45	2425.1	−109.373 911	
CCSD	*6s5p4d3f*		109.36	2429.7	−109.380 042	
CCSD(T)	*6s5p4d3f*	*4s3p2d1f*	110.05	2352.2	−109.397 161	60
CCSD(T)	*6s5p4d3f*	*5s4p3d2f*	110.07	2350.5	−109.399 808	92
CCSD(T)	*6s5p4d3f*	*6s5p4d3f*	110.08	2349.8	−109.400 134	117
CCSD	VQZ′		109.53	2415.0	−109.344 486	
CCSD	AVQZ′		109.53	2415.2	−109.345 613	
CCSD	VQZ		109.31	2435.6	−109.384 266	
CCSD	AVQZ		109.32	2434.3	−109.386 848	
CCSD(T)	AVQZ	VQZ′	109.99	2359.4	−109.404 715	64
CCSD(T)	AVQZ	AVQZ′	110.00	2358.1	−109.404 791	82
CCSD(T)	AVQZ	VQZ	110.04	2355.1	−109.407 039	110
CCSD(T)	AVQZ	AVQZ	110.05	2354.5	−109.407 243	153
Experiment [63]			109.77	2358.6		

[a] Number of virtual orbitals in S_3

the ANO_1 basis is about 15 times faster than the evaluation of the triples correction with the full ANO_3 basis, and that the use of the ANO_2 basis gives a speedup of about 2.6 relative to the ANO_3 basis. We conclude that the use of a reduced AO space for the calculation of triples corrections is a promising approach, which may give significant computational savings with an acceptable loss of accuracy.

We now consider the calculations carried out using the correlation-consistent basis sets. Four different basis sets have been used: the $[5s4p3d2f1g]$ VQZ basis, the $[5s4p3d]$ VQZ$'$ basis, the $[6s5p4d3f2g]$ AVQZ basis, and the $[6s5p4d]$ AVQZ$'$ basis. Thus, the VQZ$'$ and AVQZ$'$ basis sets have been obtained from the full VQZ and AVQZ sets by removing the f and g correlating orbitals. The augmented basis sets AVQZ and AVQZ$'$ are generated from the VQZ and VQZ$'$ sets by the addition of diffuse functions, one set for each angular momentum present in the original basis.

From Table 2, we see that the omission of the f and g correlating functions changes the triples corrections by about 0.05 pm for the bond distance and 4 cm^{-1} for the frequency. These corrections should be compared with the full triples corrections of 0.73 pm and -79.8 cm^{-1} for the AVQZ basis. The error introduced into the triples correction by the removal of the f and g correlating functions is therefore not negligible (ca. 7%) and perhaps too large to be acceptable for general use, although it should be kept in mind that the removal of the f and g functions gives a 10-fold speed-up in the calculation of the triples. What is needed, however, is a more flexible way of reducing the size of the virtual orbital space. Such an approach will be studied in the next subsection.

Before we go to Sect. 5.2, let us first consider another example: the weak van der Waals interaction between two Ne atoms. The results are collected in Table 3. We have chosen this example for two reasons. First, the triples correction is quite large (ca. 20% of the interaction energy), and second, this case study demonstrates that it is possible to apply the counterpoise correction in a straightforward manner when multiple AO basis sets are used. On the other hand, the counterpoise procedure is perhaps not so well defined for calculations in a reduced virtual basis of the type reported in the next subsection. For example, if the selection of the reduced virtual basis is based on NO occupation numbers, it is likely that all of the functions on the "ghost" atom will give rise to NOs with such low occupations that they will be omitted from the calculation right away. We return to this point in Sect. 5.3.

The results in Table 3 show that the S_3 basis set $[4s4p3d2f]$ is sufficient for the computation of the triples correction. The basis set truncation error of the $[6s5p4d3f]$ basis at the CCSD level is much larger than the error introduced by using the $[4s4p3d2f]$ set for the triples correction instead of the full $[6s5p4d3f]$ basis. Finally, we note that for the calculations on the Ne $\cdots$ Ne interaction we had to remove the innermost s- and p-type ANOs (rather than the outermost s and p as in the N_2 calculations) because these atomic functions could not be orthogonalized to the occupied orbital space. This was possible for the N_2 molecule in the previous example, but due to the very weak interaction, the orbitals of the Ne atoms change so little that the atomic funcions are linearly dependent on the occupied orbitals of the van der Waals complex.

5.2 Calculations in a reduced virtual basis: N_2, F_2, *and* CO

We now discuss the calculation of the triples correction in an active virtual space. In Table 4, we present the results from calculations on the N_2 system using the correlation-consistent basis sets VTZ, VQZ, and V5Z and a variety of active triples virtual spaces. In Table 5, the errors for these calculations are given relative to the full V5Z/CCSD(T) results. Finally, in Table 6, we have listed the triples corrections relative to the CCSD result for the different active virtual spaces. Tables 7–9 contain the corresponding numbers for the F_2 system, and in Tables 10–12 the results for the CO molecule are presented. In the following, we denote the active virtual space spanned by MP2 NOs of occupation numbers larger than 10^{-n} by S_3^n.

The first thing to note about these calculations is the very orderly improvement in the properties (i.e., the bond distance, the harmonic frequency, and the energy) with extension of the primary doubles space S_2 and with the triples virtual space S_3. The energy decreases monotonically with the extension of either space. The bond distance, in contrast, decreases with each extension of the primary space but increases as the triples virtual space is extended. In accordance with the behavior of the distance, the harmonic frequencies increase with

Table 3. Ne$\cdots$Ne van der Waals interaction: equilibrium interatomic separation (r_e), harmonic wavenumber (ω_e), and well depth (ε) obtained from coupled-cluster calculations using multiple AO basis sets. The 1s core orbitals were not correlated

	S_2	S_3	r_e (a_0)	ω_e (cm^{-1})	ε(μE_h)	N_vir[a]
Uncorrected						
CCSD	$6s5p4d3f$		5.988	25.2	−103.0	
CCSD(T)	$6s5p4d3f$	$4s3p2d1f$	5.943	26.6	−114.4	60
CCSD(T)	$6s5p4d3f$	$4s4p3d2f$	5.914	27.5	−122.0	90
CCSD(T)	$6s5p4d3f$	$6s5p4d3f$	5.913	27.6	−123.0	114
Counterpoise corrected						
CCSD	$6s5p4d3f$		6.089	21.9	−80.3	
CCSD(T)	$6s5p4d3f$	$4s3p2d1f$	6.044	23.2	−89.7	60
CCSD(T)	$6s5p4d3f$	$4s4p3d2f$	6.011	24.1	−96.8	90
CCSD(T)	$6s5p4d3f$	$6s5p4d3f$	6.002	24.4	−99.0	114

[a] Cf. Table 2

Table 4. Using MP2 NOs for the CCSD(T) triples correction: equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) for N_2. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	r_e (pm)	ω_e (cm^{-1})	E_{min} (E_{h})	N_{vir}[b]
CCSD	VTZ		109.67	2423.8	−109.355 361	
CCSD(T)	VTZ	10^{-3}	110.14	2371.8	−109.366 598	17
CCSD(T)	VTZ	10^{-4}	110.35	2348.4	−109.373 448	44
CCSD(T)	VTZ	VTZ	110.38	2346.0	−109.373 937	53
CCSD	VQZ		109.31	2435.6	−109.384 266	
CCSD(T)	VQZ	10^{-3}	109.77	2384.9	−109.395 367	17
CCSD(T)	VQZ	10^{-4}	109.99	2360.7	−109.402 632	48
CCSD(T)	VQZ	VQZ	110.03	2356.2	−109.404 391	103
CCSD	V5Z		109.22	2439.9	−109.393 470	
CCSD(T)	V5Z	10^{-3}	109.68	2389.8	−109.404 500	17
CCSD(T)	V5Z	10^{-4}	109.89	2365.7	−109.411 619	48
CCSD(T)	V5Z	10^{-5}	109.93	2361.1	−109.413 741	103
CCSD(T)	V5Z	V5Z	109.94	2359.8	−109.414 195	175
Experiment [63]			109.77	2358.6		

[a] NOs with occupation number below the given threshold are deleted. The MP2 one-electron density matrix was formed using all occupied orbitals

[b] Number of virtual orbitals in S_3

Table 5. Using MP2 NOs for the CCSD(T) triples correction: errors in the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) of N_2

	S_2	CCSD	S_3^3	S_3^4	S_3^5
δr_e (pm)	VTZ	−0.27	0.20	0.41	0.44
	VQZ	−0.63	−0.17	0.05	0.09
	V5Z	−0.72	−0.26	−0.05	−0.01
$\delta\omega_e$ (cm^{-1})	VTZ	64.0	12.0	−11.4	−13.8
	VQZ	75.8	25.1	0.9	−3.6
	V5Z	80.1	30.0	5.9	1.3
δE_{min} (mE_{h})	VTZ	58.8	47.6	40.7	40.3
	VQZ	29.9	18.8	11.6	9.8
	V5Z	20.7	9.7	2.6	0.5

Table 6. CCSD(T) triples correction to the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) of N_2

	S_2	S_3^3	S_3^4	S_3^5
Δr_e (pm)	VTZ	0.47	0.68	0.71
	VQZ	0.46	0.68	0.72
	V5Z	0.46	0.67	0.71
$\Delta\omega_e$ (cm^{-1})	VTZ	−52.0	−75.4	−77.8
	VQZ	−50.7	−74.9	−79.4
	V5Z	−50.1	−74.2	−78.8
ΔE_{min} (mE_{h})	VTZ	−11.2	−18.1	−18.6
	VQZ	−11.1	−18.4	−20.1
	V5Z	−11.0	−18.1	−20.3

extensions of the primary space but decrease with extensions of the triples virtual space.

The different behavior of the bond distances and vibrational frequencies with respect to extensions of the doubles and triples orbitals spaces is clearly displayed in Tables 5, 8, and 11 – for the bond distances, the negative errors are located in the lower left corner of the table; for the frequencies, the negative errors are located in the upper right corner. Accordingly, we find that the smallest errors are located on the diagonal – that is, at the VTZ level with an S_3^3 triples space, at the VQZ/S_3^4 level, and at the V5Z/S_3^5 level. These levels thus appear to represent balanced treatments of the doubles and triples spaces, where the errors in the doubles and triples spaces systematically cancel each other out. Although we do not unconditionally advocate the use of the hierarchy VTZ/S_3^3, VQZ/S_3^4, and V5Z/S_3^5, it nevertheless appears to offer an attractive sequence of models for accurate calculations of correlation effects. A considerably more extensive statistical investigation of the performance of these models would be required before these models can be adopted for routine calculations.

The next thing to note about the calculations is that the triples corrections are almost independent of the size of the primary orbital space. Thus, we find that, for the N_2 molecule, the S_3^3 triples correction to the bond length is 0.46–0.47 pm for all three primary basis sets. The S_3^4 corrections are 0.67 to 0.68 pm and the S_3^5 corrections are 0.71–0.72 pm. The corresponding intervals for the F_2 system are 1.48–1.53 pm at the S_3^3 level, 2.00–2.07 pm at the S_3^4 level, and 2.12–2.22 pm at the S_3^5 level. For CO, we find the intervals 0.44–0.47 pm, 0.66–0.68 pm, and 0.70–0.71 pm, respectively.

Concerning the S_3^5 level, we note that for the VTZ basis, the full virtual space S_3^{∞} is significantly smaller than the corresponding S_3^5 space at the V5Z level. In contrast, the S_3^{∞} space for the VQZ basis is only slightly smaller than the V5Z S_3^5 space. We also note in passing that the size of the active virtual space is more or less independent of the primary space as long as the primary space contains a sufficient number of orbitals.

The same behavior is observed for the frequencies and the energies. At the S_3^4 level, for example, the triples corrections to the frequency are −75.4 to −74.2 cm^{-1} for the N_2 molecule, −88.2 to −88.0 cm^{-1} for the F_2 system, and −67.3 to −66.7 cm^{-1} for CO. For the energies, at the S_3^4 level, the intervals are −18.4 to −18.1 mE_{h} (N_2), −18.3 to −17.3 mE_{h} (F_2), and −16.8 to −16.6 mE_{h} (CO).

From these numbers, we conclude that the triples corrections are almost independent of the quality of the primary basis set, at least for sets of correlation-consistent triple-zeta quality or better. This observation implies that we may converge the CCSD energy and the triples corrections independently and that, for a given accuracy, the cost of the triples correction is independent of the size of the primary basis set. For example, the calculation of

Table 7. Using MP2 NOs for the CCSD(T) triples correction: equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) for F_2. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	r_e (pm)	ω_e (cm^{-1})	E_{min} (E_{h})	N_{vir}[a]
CCSD	VTZ		139.46	1012.5	−199.278 305	
CCSD(T)	VTZ	10^{-3}	140.99	945.8	−199.287 073	18
CCSD(T)	VTZ	10^{-4}	141.48	924.3	−199.295 646	48
CCSD(T)	VTZ	VTZ	141.58	919.9	−199.296 112	51
CCSD	VQZ		139.07	1015.6	−199.338 314	
CCSD(T)	VQZ	10^{-3}	140.52	952.9	−199.347 041	18
CCSD(T)	VQZ	10^{-4}	141.14	927.6	−199.356 311	50
CCSD(T)	VQZ	VQZ	141.29	921.1	−199.358 906	101
CCSD	V5Z		138.80	1021.0	−199.359 280	
CCSD(T)	V5Z	10^{-3}	140.28	960.5	−199.367 934	18
CCSD(T)	V5Z	10^{-4}	140.80	935.0	−199.377 603	52
CCSD(T)	V5Z	10^{-5}	141.00	928.0	−199.380 309	107
CCSD(T)	V5Z	V5Z	141.10	926.0	−199.380 987	173
Experiment [63]			141.19	916.6		

[a] Cf. Table 4

Table 8. Using MP2 NOs for the CCSD(T) triples correction: errors in the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) of F_2

	S_2	CCSD	S_3^3	S_3^4	S_3^5
δr_e (pm)	VTZ	−1.64	−0.11	0.38	0.48
	VQZ	−2.03	−0.58	0.04	0.19
	V5Z	−2.30	−0.82	−0.30	−0.10
$\delta\omega_e$ (cm^{-1})	VTZ	86.5	19.8	−1.7	−6.1
	VQZ	89.6	26.9	1.6	−4.9
	V5Z	95.0	34.5	9.0	2.0
δE_{min} (mE_{h})	VTZ	102.7	93.9	85.3	84.9
	VQZ	42.7	33.9	24.7	22.1
	V5Z	21.7	13.1	3.4	0.7

Table 9. CCSD(T) triples correction to the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy (E_{min}) of F_2

	S_2	S_3^3	S_3^4	S_3^5
Δr_e (pm)	VTZ	1.53	2.02	2.12
	VQZ	1.45	2.07	2.22
	V5Z	1.48	2.00	2.20
$\Delta\omega_e$ (cm^{-1})	VTZ	−66.7	−88.2	−92.6
	VQZ	−62.7	−88.0	−94.5
	V5Z	−60.5	−86.0	−93.0
ΔE_{min} (mE_{h})	VTZ	−8.8	−17.3	−17.8
	VQZ	−8.7	−18.0	−20.6
	V5Z	−8.7	−18.3	−21.0

the triples correction at the S_3^5 level costs exactly the same for the VQZ and the V5Z primary basis sets.

Another important observation to be made about the calculations for the N_2, F_2, and CO molecules is that the convergence of the CCSD energies and properties with respect to the primary basis is slower than the convergence of the triples corrections with respect to the active virtual space. Thus, at the S_3^5 level, the triples correction to the bond distance appears to have been converged to within 0.1 pm, the frequency to within a few wave numbers, and the energy to within 1–2 mE_{h}. In contrast, at the VQZ level (which contains the full S_3^5 space), the CCSD(T) energy is still in error by as much as 10–20 mE_{h}, the bond distance by 0.1–0.2 pm, and the vibrational frequency by 1–5 cm^{-1}.

The different convergence characteristics of the CCSD energy and properties on the one hand and the triples correction on the other make the use of multiple basis sets even more attractive, since it now becomes possible to concentrate the efforts where they are needed – that is, we may converge the CCSD energy well without having to pay a penalty in terms of computation time in the calculation of the triples correction.

5.3 Ne_2 *in a reduced virtual basis*

In Sect. 5.1, we described the calculation of the Ne_2 interaction energy using multiple AO basis sets. We now consider the same system in a reduced virtual basis of MP2 NOs, (see Table 13). It is particularly interesting to investigate the counterpoise correction when a reduced virtual basis set is used, because for a given occupation number threshold, the AOs of the ghost atom do not increase the number of selected MP2 NOs, and therefore not the size of the virtual space. One might perhaps worry that the counterpoise correction cannot be applied within the framework of a reduced virtual space of MP2 NOs.

Comparing the present results with those in Table 3 (obtained using multiple AO sets), we find that the two sets of results are very similar. It thus does not appear that noticeable errors are introduced when the counterpoise method is applied in a reduced virtual basis. We note, however, that the triples contribution to the total counterpoise correction concerning the interaction energy is only 1.3 μE_{h}. Using the full 6*s*5*p*4*d*3*f* basis set, the counterpoise correction is 22.7 μE_{h} at the CCSD level and 24.0 μE_{h} at the CCSD(T) level. Thus, any small errors related to the triples contribution to the counterpoise correction must be smaller than 1.3 μE_{h} and negligible in comparison with the fraction of the total triples contribution recovered by the reduced space of virtual orbitals. Certainly, by far the most important contribution to the counterpoise correction occurs at the CCSD level, neither at the Hartree-Fock nor at the triples level.

5.4 *Calculations on ozone*

We have determined the molecular structure of ozone at the CCSD(T) level using the correlation-consistent

Table 10. Using MP2 NOs for the CCSD(T) triples correction: equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy ($E_{\min}$) for CO. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	r_e (pm)	ω_e (cm^{-1})	$E_{\min}$ (E_h)	N_{vir}[a]
CCSD	VTZ		112.87	2222.9	−113.138 549	
CCSD(T)	VTZ	10^{-3}	113.34	2173.7	−113.148 299	17
CCSD(T)	VTZ	10^{-4}	113.55	2155.6	−113.155 109	44
CCSD(T)	VTZ	VTZ	113.57	2153.7	−113.155 579	53
CCSD	VQZ		112.43	2234.8	−113.169 411	
CCSD(T)	VQZ	10^{-3}	112.88	2187.4	−113.179 013	17
CCSD(T)	VQZ	10^{-4}	113.10	2167.7	−113.186 250	48
CCSD(T)	VQZ	VQZ	113.14	2164.4	−113.187 906	103
CCSD	V5Z		112.36	2236.3	−113.179 113	
CCSD(T)	V5Z	10^{-3}	112.80	2189.5	−113.188 657	17
CCSD(T)	V5Z	10^{-4}	113.02	2169.6	−113.195 847	48
CCSD(T)	V5Z	10^{-5}	113.06	2166.0	−113.197 764	103
CCSD(T)	V5Z	V5Z	113.07	2165.1	−113.198 188	175
Experiment [63]			112.83	2169.8		

[a] Cf. Table 4

Table 11. Using MP2 NOs for the CCSD(T) triples correction: errors in the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy ($E_{\min}$) of CO

	S_2	CCSD	S_3^3	S_3^4	S_3^5
δr_e (pm)	VTZ	−0.20	0.27	0.48	0.50
	VQZ	−0.64	−0.19	0.03	0.07
	V5Z	−0.71	−0.27	−0.05	−0.01
$\delta\omega_e$ (cm^{-1})	VTZ	57.8	8.6	−9.5	−11.4
	VQZ	69.7	22.3	2.6	−0.7
	V5Z	71.2	24.4	4.5	0.9
$\delta E_{\min}$ (mE_h)	VTZ	59.6	49.9	43.1	42.6
	VQZ	28.8	19.2	11.9	10.3
	V5Z	19.1	9.5	2.3	0.4

Table 12. CCSD(T) triples correction to the equilibrium bond length (r_e), harmonic wavenumber (ω_e), and minimum total energy ($E_{\min}$) of CO

	S_2	S_3^3	S_3^4	S_3^5
Δr_e (pm)	VTZ	0.47	0.68	0.70
	VQZ	0.45	0.67	0.71
	V5Z	0.44	0.66	0.70
$\Delta\omega_e$ (cm^{-1})	VTZ	−49.2	−67.3	−69.2
	VQZ	−47.4	−67.1	−70.4
	V5Z	−46.8	−66.7	−70.3
$\Delta E_{\min}$ (mE_h)	VTZ	−9.7	−16.6	−17.0
	VQZ	−9.6	−16.8	−18.5
	V5Z	−9.5	−16.7	−18.7

valence triple-zeta (VTZ) basis set (see Table 14). Ozone was selected since the triples corrections are particularly large for this molecule. In the VTZ basis, it increases the equilibrium bond length by as much as 2.5 pm and it decreases the bond angle by −0.7°. In the S_3^4 space, the respective corrections are 2.5 pm and −0.6°, but the S_3^4 space is almost as large as the full VTZ orbital space. The S_3^3 space, on the other hand, is significantly smaller, and for this reduced virtual space the triples corrections are 1.8 pm and −0.5°, respectively. Thus, using about 35% of the virtual orbitals space yields about 70% of the full VTZ/CCSD(T) triples correction.

We would like to stress that, for ozone and the calculations of the electric properties that are reported in the next subsection, we have used valence-shell MP2 NOs – that is, the MP2 one-electron density matrix was formed by summing over the valence orbitals only. This contrasts with the previously reported calculations on N_2, F_2, and CO, where the MP2 one-electron density matrix also included the contributions from the core orbitals. Both possibilities seem to work equally well.

We conclude that the S_3^3 space of the VTZ basis offers attractive computational opportunities. Today, CCSD(T) calculations at this level are possible for extended molecular systems, for example for systems containing 10–20 first-row atoms. CCSD calculations that correlate about 100 electrons using 300–600 atomic basis functions are tractable now, but the full CCSD(T) triples correction is still beyond reach for calculations of this size. Using the S_3^3 space, however, the computation of the triples correction would become feasible.

5.5 Calculations of electric properties

Tables 15 and 16 show the results from CCSD(T) calculations of the electric dipole moment of CO and the dipole polarizability of the F^- ion, respectively, using reduced virtual spaces for the triples correction. We have used doubly and triply augmented correlation-consistent basis sets, as it is crucial for accurate computations of electric properties to provide diffuse atomic basis functions, in particular for the calculation of the dipole polarizability of a negatively charged ion.

In order to avoid a corruption of the true reduced space effects by unwanted effects due to changes in the geometry, we have computed the dipole moment of CO at a fixed bond length of $r_{\text{C–O}} = 112.83$ pm, which is the experimental equilibrium bond length.

First, we observe that the MP2 NOs in the augmented basis sets cover a broad range of occupation numbers. The d-AVTZ basis for CO, for example, provides 18 orbitals with occupation numbers below 10^{-6}, and the t-AVTZ basis for F^- provides 16 NOs below this threshold.

Second, we note that the S_3^3 space is not adequate for the computation of the triples correction to the polarizability of F^-. This correction is ca. 2 a.u. in the t-AVTZ basis, and the S_3^3 space recovers only about 4% of it. Note that this space consists of 9 orbitals only. However, it appears that the S_3^4 space already contains those vir-

Table 13. Ne···Ne van der Waals interaction: equilibrium interatomic separation (r_e), harmonic wavenumber (ω_e), and well depth (ε) obtained from coupled-cluster calculations using valence-shell MP2 NOs. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	r_e (a_0)	ω_e (cm^{-1})	ε (μE_h)	N_{vir}[b]
Uncorrected						
CCSD	$6s5p4d3f$		5.988	25.2	−103.0	
CCSD(T)	$6s5p4d3f$	10^{-3}	5.955	25.8	−107.2	18
CCSD(T)	$6s5p4d3f$	10^{-4}	5.946	26.3	−111.5	50
CCSD(T)	$6s5p4d3f$	10^{-5}	5.930	26.9	−116.7	82
CCSD(T)	$6s5p4d3f$	$6s5p4d3f$	5.913	27.6	−123.0	114
Counterpoise corrected						
CCSD	$6s5p4d3f$		6.089	21.9	−80.3	
CCSD(T)	$6s5p4d3f$	10^{-3}	6.049	22.7	−84.8	18
CCSD(T)	$6s5p4d3f$	10^{-4}	6.032	23.4	−90.2	50
CCSD(T)	$6s5p4d3f$	10^{-5}	6.017	23.9	−94.3	82
CCSD(T)	$6s5p4d3f$	$6s5p4d3f$	6.002	24.4	−99.0	114

[a] NOs with occupation number below the given threshold are deleted. The MP2 one-electron density matrix was formed using the occupied valence orbitals

[b] Number of virtual orbitals in S_3

Table 14. Ozone molecule: equilibrium geometry (r_e and $\angle_{\mathrm{O-O-O}}$) and minimum total energy (E_{min}) obtained from coupled-cluster calculations using valence-shell MP2 NOs. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	r_e (pm)	$\angle_{\mathrm{O-O-O}}$ (deg)	E_{min} (E_h)	N_{vir}[a]
CCSD	VTZ		125.0	117.6	−225.086 66	
CCSD(T)	VTZ	10^{-3}	126.8	117.1	−225.116 99	27
CCSD(T)	VTZ	10^{-4}	127.5	117.0	−225.131 74	68
CCSD(T)	VTZ	VTZ	127.5	116.9	−225.132 78	78
Experiment [64]			127.2	116.8		

[a] Cf. Table 13

Table 15. CO molecule (internuclear distance fixed at $r_{\mathrm{C-O}} = 112.83$ pm): electric dipole moment (μ) and total energy (E_{tot}) obtained from coupled-cluster calculations using valence-shell MP2 NOs. The 1*s* core orbitals were not correlated

	S_2	S_3[a]	μ (D)	E_{tot} (E_h)	N_{vir}[a]
CCSD	AVDZ		0.0963	−113.060 978	
CCSD(T)	AVDZ	10^{-3}	0.1257	−113.067 238	13
CCSD(T)	AVDZ	10^{-4}	0.1375	−113.073 013	28
CCSD(T)	AVDZ	d-AVDZ	0.1407	−113.073 280	39
CCSD	d-AVTZ		0.0713	−113.145 401	
CCSD(T)	d-AVTZ	10^{-3}	0.1225	−113.155 022	17
CCSD(T)	d-AVTZ	10^{-4}	0.1284	−113.162 072	44
CCSD(T)	d-AVTZ	10^{-5}	0.1282	−113.162 963	72
CCSD(T)	d-AVTZ	10^{-6}	0.1283	−113.163 003	99
CCSD(T)	d-AVTZ	d-AVTZ	0.1283	−113.163 003	117

[a] Cf. Table 13

Table 16. F^{-} ion: static dipole polarizability (α) and total energy (E_{tot}) obtained from coupled-cluster calculations using valence-shell MP2 NOs. The 1*s* core orbital was not correlated

	S_2	S_3[a]	α (a.u.)	E_{tot} (E_h)	N_{vir}[a]
CCSD	d-AVDZ		13.58	−99.663 598	
CCSD(T)	d-AVDZ	10^{-3}	13.60	−99.664 636	9
CCSD(T)	d-AVDZ	10^{-4}	14.82	−99.669 833	18
CCSD(T)	d-AVDZ	d-AVDZ	14.84	−99.670 054	27
CCSD	t-AVTZ		15.09	−99.739 951	
CCSD(T)	t-AVTZ	10^{-3}	15.17	−99.740 715	9
CCSD(T)	t-AVTZ	10^{-4}	17.13	−99.749 234	28
CCSD(T)	t-AVTZ	10^{-5}	17.10	−99.750 318	41
CCSD(T)	t-AVTZ	10^{-6}	17.07	−99.750 416	57
CCSD(T)	t-AVTZ	t-AVTZ	17.07	−99.750 419	73

[a] Cf. Table 13

tual orbitals that are important for a quantitative computation of the triples correction of the electric properties under study. In this reduced space, the results are very close to the respective results from the full basis sets.

Third, for the first time in the present study, we observe that the magnitude of the triples correction is overestimated by some of the reduced spaces of virtual orbitals. The CCSD(T) triples correction to the dipole moment of CO in the S_3^4/d-AVTZ space is 100.1% of the full space correction, and the corrections to the polarizability of F$^-$ amount to 100.4% and 100.2% in the S_3^4/t-AVTZ and S_3^5/t-AVTZ spaces, respectively. The reason for this is that, in comparison with calculations without electric field, these spaces recover slightly more of the energy correction if the electric field is applied.

We believe that the CO and F$^-$ examples represent rather difficult cases, and we are very pleased with the performance of the reduced virtual spaces for the calculation of the electric properties.

6 Conclusion

We have reduced the size of the basis set of virtual orbitals for the computation of perturbational triples corrections in coupled-cluster theory. It has been shown in a number of case studies that the virtual space can be reduced considerably (e.g., by 50%), without a significant loss in accuracy. In order to achieve the best possible accuracy using limited computer resources, the best strategy is to employ the largest basis set that is feasible for the CCSD calculation and a smaller one for the triples correction – just small enough to make possible its computation.

The important observation was made that the convergence of the CCSD energy with the size of the basis set and the convergence of the triples correction are largely mutually independent. Thus, it would be advantageous to investigate the convergence behavior of the S_2 and S_3 spaces independently, thereby yielding additional insight into the overall accuracy of the CCSD(T) calculation.

For the bond distances and vibrational frequencies, the extensions of the S_2 and S_3 spaces have opposite effects that may be exploited in studies on large systems (although careful calibration is then needed).

Most of the calculations reported in this paper were based on MP2 NOs. We do not expect these orbitals to be the optimal orbitals for the triples corrections, and it seems worthwhile to develop more sophisticated approaches, for instance methods that minimize the (always negative) CCSD+T(CCSD)-type triples correction. Such a method would be similar to the OVO approach concerning the MP2 energy functional developed by Adamowicz and Bartlett.

In conjunction with the integral-direct CCSD method, the use of active spaces of virtual orbitals for the triples replacements will result in a vast increase in the size of the molecular systems accessible by coupled-cluster methods that include perturbational triples corrections. This will also apply to the full CCSDT method.

Acknowledgements. The authors would like to thank Poul Jørgensen and Jeppe Olsen for discussions. This work was supported by the Slovak grant agency VEGA (project no. 2-501), the Danish Natural Science Research Council (project no. 11-0924), and the Research Council of Norway (project no. 111182/410). A generous allocation of computing time by the MetaCenter for Supercomputing in Norway (grant no. NN2694K) is gratefully acknowledged.

References

1. Helgaker T, Klopper W, Koch H, Noga J (1997) J Chem Phys 106:9639
2. Jurgens-Lutovsky R, Almlöf JE (1991) Chem Phys Lett 178:451
3. Almlöf JE (1995) In: Yarkony DR (ed) Modern electronic structure theory, part I. World Scientific, Singapore, p 110
4. Almlöf JE, Fægri K Jr, Korsell K (1982) J Comput Chem 3:385
5. Head-Gordon M, Pople JA, Frisch MJ (1988) J Chem Phys 153:503
6. Sæbø S, Almlöf JE (1989) Chem Phys Lett 154:83
7. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165
8. Klopper W, Almlöf JE (1993) J Chem Phys 99:5167
9. Koch H, Christiansen O, Kobayashi R, Jørgensen P, Helgaker T (1994) Chem Phys Lett 228:233
10. Klopper W, Noga J (1995) J Chem Phys 103:6127
11. Koch H, Sánchez de Merás A, Helgaker T, Christiansen O (1996) J Chem Phys 104:4157
12. Almlöf JE, Taylor PR (1987) J Chem Phys 86:4070
13. Almlöf JE, Taylor PR (1992) Adv Quantum Chem 22:301
14. Löwdin PO (1955) Phys Rev 97:1474
15. Kutzelnigg W (1963) Theor Chim Acta 1:327
16. Ahlrichs R, Kutzelnigg W (1968) J Chem Phys 48:1819
17. Meyer W (1973) J Chem Phys 58:1017
18. Ahlrichs R, Lischka H, Staemmler V, Kutzelnigg W (1975) J Chem Phys 62:1225
19. Karlström G, Jönsson B, Roos BO, Siegbahn PEM (1978) Theor Chim Acta 48:59
20. Kelly HP (1964) Phys Rev 136:B896
21. Silverstone HJ, Yin ML (1968) J Chem Phys 49:2026
22. Hunt WJ, Goddard WA III (1969) Chem Phys Lett 3:414
23. Luken WL (1970) Chem Phys 40:301
24. Huzinaga S, Arnau C (1970) Phys Rev A 1:1285
25. Huzinaga S, McWilliams D, Cantu AA (1973) Adv Quantum Chem 7:187
26. Silver DM, Bartlett RJ (1976) Phys Rev A 13:1
27. Cooper IL, Pounder CNM (1979) J Chem Phys 71:957
28. Bauschlicher CW Jr (1980) J Chem Phys 72:880
29. Feller D, Davidson ER (1981) J Chem Phys 74:3977
30. Cooper IL, Pounder CNM (1982) J Chem Phys 77:5045
31. Luken WL, Seiders BAB (1985) Chem Phys 92:235
32. Fantucci P, Bonačić-Koutecky V, Koutecky J (1985) J Comput Chem 6:462
33. Adamowicz L, Bartlett RJ (1987) J Chem Phys 86:6314
34. Adamowicz L, Bartlett RJ (1988) Phys Rev A 37:1
35. Adamowicz L, Bartlett RJ, Sadlej AJ (1988) J Chem Phys 88:5749
36. Adamowicz L (1989) J Comput Chem 10:928
37. Adamowicz L (1990) J Phys Chem 93:1780
38. Adamowicz L (1991) Int J Quantum Chem Symp 25:71
39. Wasilewski J (1991) Int J Quantum Chem 39:649
40. Wasilewski J, Zelek S, Wierbowska M (1996) Int J Quantum Chem 60:1027
41. Sherrill CD, Schaefer HF III (1996) J Phys Chem 100:6069
42. Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) Chem Phys Lett 157:479
43. Urban M, Noga J, Cole SJ, Bartlett RJ (1985) J Chem Phys 83:4041
44. Bartlett RJ, Watts JD, Kucharski SA, Noga J (1990) Chem Phys Lett 165:513
45. Deegan MJO, Knowles PJ (1994) Chem Phys Lett 227:321
46. Bartlett RJ (1995) In: Yarkony DR (ed) Modern electronic structure theory, part II. World Scientific, Singapore, p 1047
47. Helgaker T, Gauss J, Jørgensen P, Olsen J (1997) J Chem Phys 106:6430
48. Lee TJ, Rendell AP, Taylor PR (1990) J Phys Chem 94:5463
49. Rendell AP, Lee TJ, Komornicki A (1991) Chem Phys Lett 178:462
50. Koch H, Jørgensen P, Helgaker T (1996) J Chem Phys 104:9528

51. Koch H, Christiansen O, Jørgensen P, Sánchez de Merás A, Helgaker T (1997) J Chem Phys 106:1808
52. Jensen HJ Aa, Jørgensen P, Ågren H, Olsen J (1988) J Chem Phys 88:3834
53. Noga J, Klopper W, DIRCCR12-95, a coupled-cluster program
54. Dunning TH Jr (1989) J Chem Phys 90:1007
55. Kendall RA, Dunning TH Jr, Harrison RJ (1992) J Chem Phys 96:6796
56. Woon DE, Dunning TH Jr (1993) J Chem Phys 98:1358
57. Woon DE, Dunning TH Jr (1994) J Chem Phys 100:2975
58. Woon DE, Dunning TH Jr (1995) J Chem Phys 103:4572
59. The correlation-consistent basis sets AVXZ (with X = D, T, Q, 5, and 6) were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 1.0, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory, which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multi-program laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830
60. Widmark PO, Malmqvist PÅ, Roos BO (1990) Theor Chim Acta 77:291
61. The ANO basis sets were obtained from the MOLCAS Version 3 basis set library: Andersson K, Fülscher MP, Karlström G, Lindh R, Malmqvist PÅ, Olsen J, Roos BO, Sadlej AJ, Blomberg RA, Siegbahn PEM, Kellö V, Noga J, Urban M, Widmark PO, University of Lund, Sweden
62. Wapstra AH, Bos K (1977) At Data Nucl Data Tables 19:175
63. Huber KP, Herzberg G (1979) Molecular spectra and molecular structure. IV. Constants of diatomic molecules. Van Nostrand Reinhold, New York
64. Tanaka T, Morino Y (1978) J Mol Spectrosc 33:538