JCTC Journal of Chemical Theory and Computation

Fully Automated Incremental Evaluation of MP2 and CCSD(T) Energies: Application to Water Clusters

Joachim Friedrich* and Michael Dolg

Institute for Theoretical Chemistry, University of Cologne, Greinstrasse 4, 50939 Cologne, Germany

Abstract: A fully automated implementation of the incremental scheme for CCSD energies has been extended to treat MP2 and CCSD(T) energies. It is shown in applications on water clusters that the error of the incremental expansion for the total energy is below 1 kcal/mol already at second or third order. It is demonstrated that the approach saves CPU time, RAM, and disk space. Finally it is shown that the calculations can be run in parallel on up to 50 CPUs, without significant loss of computer time.

1. Introduction

It is well-known that a quantitative description of the electronic structure in molecules is usually not possible with the Hartree−Fock (HF) method. One way to achieve higher accuracy in electronic structure calculations is to improve the wave function, which can be routinely done by many-body perturbation theory (MBPT), configuration interaction theory (CI), or coupled-cluster theory (CC). The major drawback of these approaches is their strong dependence of the computational effort on the size of the one-particle basis set. This means that these approaches depend heavily on the system size too, if atom-centered basis functions are used.

Since for large systems the canonical HF orbitals are not necessarily the best choice for a PT, CI, or CC expansion of the wave function, many groups use a local orbital basis instead to include electron correlation. This allows one to screen out insignificant contributions to the energy, and therefore the computational cost is reduced.[1−19] Conceptually different approaches divide the total system into parts and then perform a perturbation expansion to obtain the total correlation energy.[20−23] An approach designed in this way is the incremental scheme of Stoll.[24−26] It is based on the Bethe−Goldstone expansion, which was introduced to quantum chemistry by Nesbet[27−29] more than 40 years ago. The incremental scheme was successfully applied during the past 15 years to various periodic systems[30−34] and mole-

cules.[35−39] Besides the treatment of closed-shell systems, extensions to open-shell cases have been developed too.[40−42]

Recently we proposed a fully automated implementation of the incremental scheme for CCSD energies,[36] implemented an automatic distance screening,[37] extended the approach for the usage of symmetry[39] and to the RCCSD method for open-shell calculations.[42] To account for the local character of the core electrons, we introduced an efficient scheme to treat the core and core−valence correlation.[43] In this work we now extend our fully automated implementation to second-order Møller−Plesset perturbation theory (MP2) and to the coupled-cluster ansatz with singles, doubles, and perturbative triples excitations CCSD(T). Furthermore, as recently proposed, we use a second basis set to describe the environment of the orbital domains in a computationally efficient manner.[44] We note that the idea of multiple basis sets is not entirely new, since Jurgens-Lutovsky and Almlöf made use of a reduced basis for the occupied space in MP2 calculations already in 1991[45] and Klopper et al. made use of a reduced basis for the treatment of the triples in CCSD(T) calculations in 1997.[46] However, we exploit this idea in the framework of the incremental scheme and check how it performs with respect to the overall accuracy and the CPU time requirements.

2. Theory

2.1. Incremental Scheme. In an incremental calculation we divide the total system into small domains consisting of groups of localized occupied orbitals according to the procedure outlined in refs 36−38. Then we calculate the

* To whom correspondence should be addressed. Tel.: (+49) (0)221-470-6886. Fax: (+49) (0)221-470-6896. E-mail: joachim_friedrich@gmx.de.

correlation energies for these domains. To include the nonadditivity corrections, we also calculate correction energies of pairs and triples, etc., of domains, until we reach the desired accuracy. The correlation energy is then computed according to

$$E_{corr} = \sum_i \Delta\varepsilon_i + \frac{1}{2!}\sum_{ij}\Delta\varepsilon_{ij} + \frac{1}{3!}\sum_{ijk}\Delta\varepsilon_{ijk} + ... \quad (1)$$

$$\Delta\varepsilon_i = \varepsilon_i \quad \Delta\varepsilon_{ij} = \varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j$$

where ε_i is the correlation energy of the subsystem i and ε_{ij} the correlation energy of the subsystem i and j together. For the convenient treatment of higher order terms we use a notation based on simple set theory.[36] Now eq 1 reads

$$E_{corr} = \sum_{\substack{X \\ X \in P(X) \wedge |X| \le O}} \Delta\varepsilon_X \quad (2)$$

D is the set of domains, $P(D)$ stands for the power set of the set of the domains and O denotes the order of the expansion. The summation index in eq 2 runs over all increments up to the order O (see refs 36 and 37 for details). When ε_X represents the correlation energy of the unified subsystems of the general correlation energy increment, $\Delta\varepsilon_X$ is given as

$$\Delta\varepsilon_X = \varepsilon_X - \sum_{\substack{Y \\ Y \in P(X) \wedge |Y| < |X|}} \Delta\varepsilon_Y \quad (3)$$

here the summation index Y runs over the power set of X.

2.2. Introduction of a Domain-Specific Basis Set. To reduce the computational cost, it is necessary to reduce the virtual space of the domains. As an alternative to an explicit elimination of selected orbitals from the virtual space spanned by the full basis set, we introduce a small basis set to describe the environment of an n-site domain and use the original larger basis set only in the main part of the domain.[44] In the current version of the code we use the distance to the center of charge of a localized orbital (t_{main}) to determine which atoms have to be treated with the full basis set. Since a domain represents a set of occupied localized orbitals, we combine the main regions of all orbitals of the domain, to obtain the set of atoms which are treated with the full basis set.

2.3. Localized Orbitals and Perturbation Theory. The computationally cheap canonical representation of the MP2 energy as well as the canonical representation of the perturbative triples correction in CCSD(T) are not invariant with respect to unitary transformations within the occupied space. To account for this within the canonical representation of the theory, we construct the transformed Fock matrix $\tilde{F}(C_L)$ of the local basis

$$\tilde{F}(C_L) = (S^{-1/2})^\dagger F(C_L)S^{-1/2} \quad (4)$$

where S is the overlap matrix, $S^{-1/2}$ is constructed to orthogonalize the basis symmetrically, and C_L is the MO coefficient matrix in the local basis. Next we build the matrix ϵ

$$\epsilon = C_L'^\dagger \tilde{F}(C_L)C_L' \quad (5)$$

with $C_L' = S^{1/2}C_L$. In the canonical basis, ϵ is diagonal and it contains the orbital energies, but in the local basis it is not diagonal. Now we classify the occupied orbitals into four classes, the frozen core orbitals, which are not correlated in all calculations, the environment orbitals, which are frozen in a specific calculation, the domain orbitals, which have to be correlated in this specific domain and the virtual orbitals, which are unoccupied in the HF reference. Furthermore we do not distinguish between core and environment orbitals, since they are treated equally in the subsequent steps. According to these criteria, we classify the blocks of the ϵ matrix into nine blocks:

$$\begin{array}{c} \\ \text{core} \\ \text{domain} \\ \text{virtual} \end{array} \begin{array}{ccc} \text{core} & \text{domain} & \text{virtual} \\ \left(\begin{array}{ccc} \epsilon_{cc} & \epsilon_{cd} & \epsilon_{cv} \\ \epsilon_{dc} & \epsilon_{dd} & \epsilon_{dv} \\ \epsilon_{vc} & \epsilon_{vd} & \epsilon_{vv} \end{array}\right) \end{array} = \epsilon = C'^\dagger\tilde{F}(C_L)C' \quad (6)$$

with (frozen core + environment) = core. In our approach we diagonalize ϵ in the subspace of the domain. In other words, we diagonalize the ϵ_{dd} block of ϵ. The unitary matrix U which is necessary to transform the MOs has the form

$$\begin{array}{c} \\ \text{core} \\ \text{domain} \\ \text{virtual} \end{array} \begin{array}{ccc} \text{core} & \text{domain} & \text{virtual} \\ \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \tilde{U} & 0 \\ 0 & 0 & 1 \end{array}\right) \end{array} = U \quad (7)$$

where $\tilde{U}$ is the matrix which diagonalizes ϵ_{dd}. The corresponding MOs are then obtained by

$$\tilde{C}_L = S^{-1/2}C'U \quad (8)$$

This scheme includes more and more correction terms to the orbitals at higher order, and therefore it converges to the canonical treatment. In the limit of a full incremental expansion, the exact result is obtained, since the original diagonal Fock matrix is recovered.

2.4. Distance Screening. The second-order energy increments decay usually very rapidly with increasing distance of the domains. This can be used to introduce a distance criterion, in order to screen out the small contributions.[37] Since the incremental contributions decay with increasing order, we use an order-dependent distance threshold of

$$t_{dist} = \frac{f_{method}}{(O\text{-}1)^2}$$

where O is the order and f_{method} is an adjustable parameter for every method. A value of $f_{method} = \infty$ means that no distance truncation is performed.

2.5. Obtaining the Correlation Energies. The AO basis for a domain is determined by the t_{main} parameter (vide supra). Since we change the AO basis for every domain, we have to construct MOs for every domain. In the current work this is done by HF calculations in the AO basis of the domain with a subsequent Boys localization. The orbitals of the domain K are then identified by the centers of charge. This requires a unique mapping of the charge centers in the basis B_1 to the charge center in the basis B_2 (see ref 44 for details).

MP2 and CCSD(T) Energies Incremental Evaluation

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **289**

In the present work we apply the MP2, CCSD, and the CCSD(T) approach as implemented in MOLPRO[47−49] to evaluate the correlation energies. For a domain K we correlate all electrons in the orbitals of K.

3. Computational Details

3.1. Incremental Calculations. First we perform a HF calculation for the total system in a minimal basis set with a subsequent Boys localization.[50,51] In the next step we extract the molecular orbital coefficient matrix, the overlap matrix in AO basis, and the dipole integrals in AO basis from the MOLPRO calculation, in order to construct the n-site domains,[36] where the parameter t_{con} determines the connectivity in the edge-weighted graph of the correlated occupied orbitals and the parameter dsp controls the size of the domains (for details see refs 36−39). Then we determine the basis set for a domain by the distance t_{main}. Now a HF calculation with the dual basis set is performed, and the orbitals are localized with the Boys procedure of MOL-PRO.[47] Next we classify the new MOs into the domains according to the centers of charge and construct the pseudocanonical orbitals for the domain using the procedure above. Finally we calculate the correlation energy for the domain using the MOLPRO MP2, CCSD, and CCSD(T) codes. To avoid numerical problems due to the error propagation in the incremental series, we determine the HF energy to 10^{-11} hartree and use a dynamical energy threshold e_{thres} to determine the accuracy of the correlation calculations, as described in ref 43.

3.2. Geometries. If nothing else is stated, we optimized the geometries with the RI-BP86/SVP gradient-corrected density functional theory method[52,53] in the TURBOMOLE 5.6[54] quantum chemistry package. Stationary points were characterized by analyzing the Hessian matrix. Note that the goal of the current work is not to derive very accurate structural data for the compounds investigated here but rather to obtain reasonable geometrical parameters for the incremental MP2, CCSD, and CCSD(T) calculations.

3.3. Hardware. The calculations were performed on a cluster of Intel Core2Quad Q6600 PCs with 2.4 GHz, 4 GB random access memory (RAM), and 160 GB disk space per node. The PCs are connected with 1 Gbit ethernet.

4. Applications

The calculation of water clusters is an active field in quantum chemistry.[22,23,55−65] Since molecular clusters have natural domains, these objects were studied with a perturbation series in terms of single molecules already in the 1970 of the past century at the HF level.[66] Later Xantheas did a many-body analysis for these clusters in terms of water molecules at the HF and at the correlated level.[56] Since chemical reactions in solution are very important, we decided to study the performance of our scheme for water clusters.

4.1. $(H_2O)_6(S_6)$. For intermolecular clusters such as $(H_2O)_6$ (Figure 1) we can use the full S_6 symmetry of the system to reduce the number of calculations significantly. In this work we use the symmetry analysis as introduced in ref 39. The accuracy of the new scheme is tested for the correlation
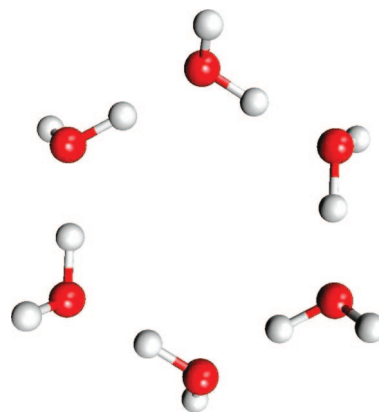
Figure 1. RI-BP86/SVP optimized structure of $(H_2O)_6$ (S_6).

Table 1. Comparison of the Incremental cc-pVDZ CCSD(T), CCSD, and MP2 Correlation Energies with the Canonical Energies for $(H_2O)_6$ (t_{main} = 3 bohr; dsp = 4, t_{con} = 3 bohr; e_{thres} = 1 × 10^{-6} au; core = 6; RAM = 800 MB; fit basis, H, O = STO-3G; $f_{CCSD(T)}$ = f_{CCSD} = f_{MP2} = ∞; S_6)

method	order	ith order correction (au)	$E_{corr}(i)$ (au)	error (kcal/mol)
CCSD(T)	1	−1.278313	−1.278313	35.06
	2	−0.056613	−1.334926	−0.47
canonical			−1.334179	
CCSD	1	−1.260129	−1.260129	31.03
	2	−0.050369	−1.310497	−0.57
canonical			−1.309585	
MP2	1	−1.210674	−1.210674	34.15
	2	−0.055719	−1.266393	−0.81
canonical			−1.265103	

consistent double- and triple-ζ basis sets of Dunning and co-workers as well as the corresponding augmented basis sets[67,68] (cc-pVXZ; aug-cc-pVXZ; X = D,T). Furthermore we explore the accuracy and the timing with respect to the fit basis for the environment and the energy threshold e_{thres}. The energies of the CCSD(T), CCSD, and MP2 calculations are given in Tables 1−4, the timings and the disk space requirements of the CCSD(T) calculations are given in Table 5. The dsp parameter has been set to 4 in all calculations on water, since we freeze the 1s orbitals of the oxygen atom, which means that we have to correlate 4 occupied orbitals per water molecule. Therefore a dsp = 4 in combination with a t_{con} = 3 forces the one-site domains to be a single water molecule.

In Table 1 we use a minimal basis set (STO-3G) as fit basis for the cc-pVDZ calculations. In this case we get chemical accuracy of 1 kcal/mol already at the second order for all applied correlation methods. If a minimal basis is used for the environment in combination with the aug-cc-pVDZ basis with diffuse functions, which are clearly necessary for the correct description of intermolecular interactions, we need a third-order expansion to obtain chemical accuracy (Table 2). The reason for this is most likely the basis set superposition error (BSSE) in the incremental energies due to the minimal basis set. If we extend the basis set of the environment and use a 6-31G basis for the oxygen atoms, the errors of the applied correlation methods are reduced significantly. For the second-order expansion we already have chemical accuracy, and at the third-order level the errors

Table 2. Comparison of the Incremental aug-cc-pVDZ CCSD(T), CCSD, and MP2 Correlation Energies with the Canonical Energies for $(H_2O)_6$ (t_{main} = 3 bohr; dsp = 4; t_{con} = 3 bohr; e_{thres} = 1 × 10⁻⁶ au; core = 6; RAM = 1520 MB; $f_{CCSD(T)} = f_{CCSD} = f_{MP2} = \infty$; S_6)

| method | order | fit basis: H, O = STO-3G | | | fit basis: H = STO-3G; O = 6-31G | | | fit basis: aug-cc-pVDZ | | |
		correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)	correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)	correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)
CCSD(T)	1	−1.384842	−1.384842	38.37	−1.387153	−1.387153	36.92	−1.391714	−1.391714	34.06
	2	−0.063037	−1.447878	−1.18	−0.059193	−1.446346	−0.22	−0.053907	−1.445621	0.23
	3	0.002724	−1.445154	0.53	0.000465	−1.445881	0.07	−0.000329	−1.445950	0.03
exact			−1.445993			−1.445993			−1.445993	
CCSD	1	−1.353862	−1.353862	33.23	−1.356129	−1.356129	31.81	−1.360119	−1.360119	29.30
	2	−0.054940	−1.408802	−1.25	−0.051162	−1.407292	−0.30	−0.046544	−1.406663	0.10
	3	0.002728	−1.406074	0.46	0.000571	−1.406720	0.06	−0.000124	−1.406787	0.02
exact			−1.406814			−1.406814			−1.406814	
MP2	1	−1.313124	−1.313124	36.26	−1.314587	−1.314587	35.34	−1.318964	−1.318964	32.59
	2	−0.059355	−1.372479	−0.99	−0.056873	−1.371460	−0.35	−0.051761	−1.370726	0.11
	3	0.002330	−1.370149	0.48	0.000646	−1.370814	0.06	−0.000181	−1.370906	0.00
exact			−1.370907			−1.370907			−1.370907	

Table 3. Comparison of the Incremental cc-pVTZ CCSD(T), CCSD, and MP2 Correlation Energies with the Canonical Energies for $(H_2O)_6$ (t_{main} = 3 bohr; dsp = 4; t_{con} = 3 bohr; e_{thres} = 1 × 10⁻⁶ au: core = 6; RAM = 1520 MB; $f_{CCSD(T)} = f_{CCSD} = f_{MP2} = \infty$; S_6)

| method | order | fit basis: H, O = STO-3G | | | fit basis: H = STO-3G; O = 6-31G | | |
		correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)	correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)
CCSD(T)	1	−1.639658	−1.639658	39.16	−1.643736	−1.643736	36.60
	2	−0.063213	−1.702870	−0.51	−0.058272	−1.702008	0.03
	3	0.001173	−1.701698	0.23			
canonical			−1.702060			−1.702060	
CCSD	1	−1.593649	−1.593649	33.77	−1.597473	−1.597473	31.37
	2	−0.054810	−1.648459	−0.63	−0.050153	−1.647626	−0.10
	3	0.001242	−1.647217	0.15			
canonical			−1.647516			−1.647516	
MP2	1	−1.564739	−1.564739	38.24	−1.568374	−1.568374	35.96
	2	−0.062153	−1.626892	−0.76	−0.057661	−1.626035	−0.23
	3	0.001530	−1.625362	0.20			
canonical			−1.625751			−1.625751	

Table 4. Comparison of the Incremental aug-cc-pVTZ CCSD(T), CCSD, and MP2 Correlation Energies with the Canonical Energies for $(H_2O)_6$ (t_{main} = 3; dsp = 4; t_{con} = 3; e_{thres} = 1 × 10⁻⁶ au; core = 6; RAM = 1520 MB; fit basis, H = STO-3G and O = 6-31G; $f_{CCSD(T)} = f_{CCSD} = f_{MP2} = \infty$; S_6)

method	order	correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)
CCSD(T)	1	−1.678737	−1.678737	−
	2	−0.064205	−1.742942	−
CCSD	1	−1.627791	−1.627791	34.25
	2	−0.055066	−1.682857	−0.31
canonical			−1.682365	
MP2	1	−1.603516	−1.603516	38.74
	2	−0.062281	−1.665797	−0.34
canonical			−1.665256	

are negligible. To compare the error of the reduced basis set, we included a calculation with the aug-cc-pVDZ for the environment. From this we see that the usage of a smaller basis set does not change the convergence behavior for CCSD, and CCSD(T) and the errors are of the same order of magnitude compared to the 6-31G/STO-3G basis in the environment. Furthermore the calculation with the full basis set in the environment shows that the incremental expansion of the diagonalization corrections leads to the correct canonical energy in this case.

Considering the cc-pVTZ basis set, we have similar findings (Table 3). The second-order errors are larger, if a minimal basis is used to model the environment, and they become much smaller, if we use the unpolarized 6-31G basis of double-ζ quality for oxygen. For the aug-cc-pVTZ basis we find chemical accuracy at the second order of the incremental expansion with the STO-3G/6-31G fit basis for the CCSD and the MP2 energies. In this case the canonical CCSD(T) calculation was infeasible with the 32-bit executable of MOLPRO and 2.4 GB RAM. Since the errors of CCSD and MP2 are only −0.31 and −0.34 kcal/mol, respectively, we expect a similar error for the CCSD(T) energy, too.
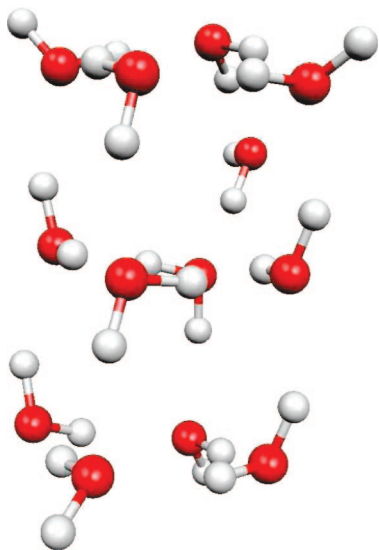
The timings of the incremental calculations are calculated by adding the real time of the client processes and the master process. Therefore they include the time for all processes required for the calculation except the time for the HF calculation in the full basis set. Since the time of the HF calculation is negligible compared to the time spent in the CCSD(T) calculation, the timings of the incremental calculations are directly comparable to the real time of the canonical calculation. Comparing the timings of the incremental calculations with the timings of the canonical calculations (Table 5), we find large improvements in the CPU time, without significant loss of accuracy. For the CCSD(T)/aug-cc-pVDZ calculations we find a reduction to less than 3.3% of the time used in the canonical case (last column), if a

MP2 and CCSD(T) Energies Incremental Evaluation

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **291**

Table 5. Timings and Disk Space Requirements of the Incremental and Canonical CCSD(T) Calculations for $(H_2O)_6$, Using the MOLPRO Quantum Chemistry Package

method	order	no. of slaves	wall time (s)	wall time (%)	total time (s)	total time (%)	disk space (GB)	disk space (%)
cc-pVDZ								
canonical[a]		1	5842.6	100.0	5842.6	100.0	1.0	100.0
STO-3G fit[a]	2	3	107.6	1.8	260.6	4.5	0.1	8.6
aug-cc-pVDZ								
canonical[d]		1	60860.8	100.0	60860.8	100.0	6.0	100.0
STO-3G fit[b]	3	7	2279.2	3.7	10493.3	17.2	1.6	26.8
6-31G/STO-3G fit[b]	3	7	3041.5	5.0	14292.3	23.5	2.1	34.4
6-31G/STO-3G fit[b]	2	4	613.0	1.0	2022.2	3.3	0.8	13.3
aug-cc-pVDZ fit[b,c]	3	8	6692.1	11.0	41259.1	67.8		
aug-cc-pVDZ fit[b,c]	2	4	2955.2	4.9	12748.4	20.9		
cc-pVTZ								
canonical[d]		1	257026.5	100.0	257026.5	100.0	11.1	100.0
STO-3G fit[b]	3	7	8800.5	3.4	39230.6	15.3	3.5	31.5
6-31G/STO-3G fit[b]	2	4	1586.9	0.6	5975.1	2.3	1.7	14.8
aug-cc-pVTZ								
canonical (CCSD)[d,e]		1	406096.4		406096.4		115.4	
6-31G/STO-3G fit (CCSD(T))[b]	2	4	8769.7		31130.3		9.7	

[a] 800 MB RAM. [b] 1520 MB RAM. [c] 64 bit executable of MOLPRO of the 2006.1 version. [d] 2400 MB RAM. [e] Not enough memory for the triples calculation.

Figure 2. Structure of the $(H_2O)_{13}$ cluster reported by Bulusu et al.[63]

STO-3G/6-31G fit basis is used. Since the calculation can be run in parallel, we end up with a wall time of 1% for this calculation using four CPUs. If the incremental calculations are run in the full basis set of the environment, the required computer times are much larger compared to the reduced basis set calculations, but still stay below the time for the canonical calculation for the second and third order. For all timings of this work we included the wall time of the master process into the sum of the CPU times, even though it does not require this time completely. For the 6-31G/STO-3G fit in the cc-pVTZ basis the wall time is only 0.6% of the time, needed for the canonical calculation.

A comparison of the disk space requirements for the canonical and the incremental calculations in different basis sets is given in Table 5. As one can see from this, the disk space requirements are significantly reduced, for all incremental calculations. The disk space for the incremental calculations depend on the order of the expansion, on the applied fit basis and on the threshold t_{main}. If the fit basis is increased, the disk space requirements increase, too. Since it was sufficient to do a second-order expansion on the water cluster to obtain chemical accuracy with the larger fit basis, the disk space requirement is lower compared to the total STO-3G fitting and a third-order expansion. Finally we conclude that the incremental scheme can be used to reduce the disk space requirements significantly.

4.2. $(H_2O)_{13}$. To check the performance of our approach for larger water clusters, we studied the $(H_2O)_{13}$ cluster as reported by Bulusu et al.[63] (Figure 2). In the small 6-31G** basis of Pople and co-workers,[69,70] we find chemical accuracy at third order for CCSD(T), CCSD, and MP2 using a STO-3G basis to fit the environment (Table 6). If we use the unpolarized 6-31G basis for oxygen and the STO-3G on hydrogen in the environment, we get chemical accuracy already at the second order of the incremental expansion. Therefore we conclude that a mixed STO-3G/6-31G basis set is a good choice to model water clusters with high accuracy at reduced cost. Since the reference calculation in the small 6-31G** basis is already very expensive, we were not able to perform the CCSD(T)/CCSD calculations using the aug-cc-pVDZ/cc-pVTZ basis sets, whereas the incremental calculations were still feasible (Table 8). From the previous results we expect that the third-order calculations provide chemical accuracy for these systems. Therefore we conclude that the proposed variant of the incremental method is a useful tool to calculate high-level CCSD(T) energies for large systems which are not accessible with the standard approaches.

Figure 3 shows the error of the MP2, CCSD, and CCSD(T) third-order energies for $(H_2O)_{13}$ with respect to the truncation parameter f. For $f = 60$ all terms where included in the incremental series which means that larger values of f will not change the energies for third-order calculations. Since $f = 20$ leads to small errors in this case, we use this value to perform the distance screening.

Table 6. Comparison of the Incremental 6-31G** CCSD(T), CCSD, and MP2 Correlation Energies with the Canonical Energies for $(H_2O)_{13}$ (t_{main} = 2 bohr; dsp = 4, t_{con} = 3 bohr; e_{thres} = 1 × 10^{-6} au; core = 13; RAM = 800 MB; fit basis, H, O = STO-3G; $f_{CCSD(T)}$ = 20 bohr; f_{CCSD} = 50 bohr; f_{MP2} = 50 bohr; C_1)

| method | order | fit basis: H, O = STO-3G | | | fit basis: H = STO-3G; O = 6-31G | | |
		correction(i) (au])	$E_{corr}(i)$ (au)	error (kcal/mol)	correction(i) (au)	$E_{corr}(i)$ (au)	error (kcal/mol)
CCSD(T)	1	−2.646002	−2.646002	67.32	−2.669007	−2.669007	52.88
	2	−0.105492	−2.751494	1.12	−0.083867	−2.752874	0.26
	3	−0.002987	−2.754480	−0.75	−0.000353	−2.753226	0.04
canonical			−2.753284			−2.753284	
CCSD	1	−2.614089	−2.614089	59.36	−2.635218	−2.635218	46.10
	2	−0.094038	−2.708126	0.35	−0.073568	−2.708786	−0.06
	3	−0.001472	−2.709598	−0.57	−0.000050	−2.708836	−0.09
canonical			−2.708687			−2.708687	
MP2	1	−2.506036	−2.506036	65.48	−2.527991	−2.527991	51.71
	2	−0.105408	−2.611444	−0.66	−0.083279	−2.611270	−0.55
	3	0.000399	−2.611045	−0.41	0.000858	−2.610412	−0.01
canonical			−2.610393			−2.610393	

Table 7. Timings of the Incremental CCSD(T) Calculations with Respect to the Canonical Ones for $(H_2O)_{13}$, Using the MOLPRO Quantum Chemistry Package

| method | order | no. of slaves | wall time | | total time | |
			(s)	(%)	(s)	(%)
canonical[a]		1	1003263.6	100.0	1003263.6	100.0
STO-3G fit[b]	3	50	1614.0	0.2	54173.6	5.4
6-31G/STO-3G fit[c]	2	20	4382.8	0.4	79996.5	8.0
6-31G/STO-3G fit[c]	3	50	5083.2	0.5	193714.6	19.3

[a] 2650 MB RAM. [b] 800 MB RAM. [c] 1520 MB RAM.

Table 8. Incremental CCSD(T), CCSD, and MP2 Correlation Energies for $(H_2O)_{13}$ (t_{main} = 2 bohr; dsp = 4; t_{con} = 3 bohr; e_{thres} = 1 × 10^{-7} au; core = 13; RAM = 1040 MB; fit basis, H = STO-3G and O = 6-31G; $f_{CCSD(T)}$ = 30 bohr; f_{CCSD} = 40 bohr; f_{MP2} = 40 bohr; C_1)

| method | order | aug-cc-pVDZ[a] | | cc-pVTZ[b] | |
		correction(i) (au)	$E_{corr}(i)$ (au)	correction(i) (au)	$E_{corr}(i)$ (au)
CCSD(T)	1	−2.974964	−2.974964	−3.529315	−3.529315
	2	−0.138585	−3.113549	−0.138913	−3.668228
	3	0.005169	−3.108381	−0.000497	−3.668725
CCSD	1	−2.911434	−2.911434	−3.433496	−3.433496
	2	−0.119177	−3.030611	−0.119799	−3.553295
	3	0.004951	−3.025661	0.000601	−3.552694
MP2	1	−2.815292	−2.815292	−3.364553	−3.364553
	2	−0.128732	−2.944024	−0.136900	−3.501453
	3	0.004889	−2.939135	0.003631	−3.497821

[a] The canonical reference calculations with 533 contracted Gaussian basis functions, 52 correlated occupied orbitals, and 468 virtual orbitals were not feasible. [b] The canonical reference calculations with 754 contracted Gaussian basis functions, 52 correlated occupied orbitals, and 689 virtual orbitals were not feasible.
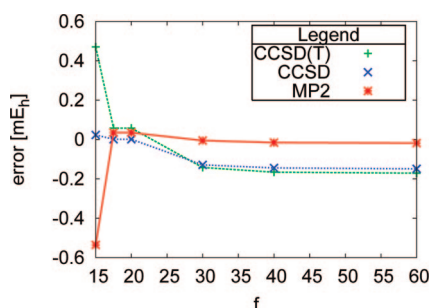
Figure 3. Error of the MP2, CCSD, and CCSD(T) energy with respect to the truncation parameter f.

Considering the timing for the $(H_2O)_{13}$ cluster in Table 7, we find a reduction of the total CPU time of 94.6% compared to the canonical calculation for the STO-3G fit. The wall time is 0.2% on 50 slave PCs. The parallel efficiency as

defined by Pulay and co-workers[71] (total CPU time divided by the product of the wall time and the number of CPUs) is in this case 65.8%, which is similar to the efficiency reported by other groups.[64,71,72] If we calculate the efficiency with respect to the canonical calculation:

$$\text{total efficiency} = \frac{\text{time of the canonical calculation}}{\text{product of wall time and no. of CPUs}} \quad (9)$$

we get 1218.9%, due to the local approximations. Note that the master process was included in the total CPU time as well in the denominator of the efficiency. We want to point out that the accuracy is still within 1 kcal/mol in this case. Considering the results of the $(H_2O)_6$ cluster, we expect that we will not meet the high accuracy for calculations with larger basis sets. If we use the more accurate STO-3G/6-31G fit, we end up with

MP2 and CCSD(T) Energies Incremental Evaluation

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **293**

a total CPU time of 8.0% at second order. The wall time is 0.4% on 20 slave PCs and the parallel efficiency 86.9%. The reason for the larger parallel efficiency is the better ratio between the number of calculations (91) and the number of slaves (20). In general the efficiency of our approach will decrease, if the number of CPUs gets comparable to the number of calculations. The reason for this is that the time for the calculations of domains with different size may vary largely. Due to the nature of the incremental scheme the domain sizes vary largely for different orders of the incremental expansion; e.g., for one-site domains of the same size, the size is doubled at second order and tripled at third order. However, since we have no idle processes in our approach, it is not very important to optimize the efficiency defined in this way. The reason why the efficiency as defined by Pulay may become small in our scheme is just because the denominator of eq 9 counts contributions of CPUs which may already compute another job.

Going to the third-order level, we have a total CPU time of 19.3% and a wall time of 0.5% using 50 slave PCs. This corresponds to a parallel efficiency of 74.7% and a total efficiency of 387.0%.

The timing of the STO-3G/6-31G fit is somewhat worse compared to the timing of the STO-3G fit. However, due to the higher accuracy and the robustness of the STO-3G/6-31G fit for larger basis sets we conclude that it is superior to the STO-3G fit. As judged from the timings and the high accuracy, we conclude that the proposed approach is a useful tool to calculate the correlation energies of large systems at reduced cost.

Since the calculations in the domains can be run independently, we conclude that the incremental scheme is inherently parallel. Therefore it is not necessary to have all slaves available at the beginning of the calculation, and no slave has to wait for the other ones. We do not waste CPU time by idle CPUs; therefore the total CPU efficiency as defined by Pulay[71] is not that important for our incremental approach. The only process which collects the total data is the master process. Since our master and slave structure is build upon the socket++ library, the master process does not consume much CPU time. On the basis of the improved timings compared to canonical calculations and the high accuracy (within 1 kcal/mol) achieved in all cases, we conclude that our modified variant of the incremental scheme is probably competitive with the currently proposed parallel CCSD(T) approaches of Olson et al.[64] and Auer et al.[72] Although the current pilot implementation is not generally applicable, we demonstrated for an important class of chemical systems that the approach considerably improves the speed of the calculations. In our future research we plan to use a projection technique to remove the mapping step for the identification of the domains, in order to make the approach generally applicable.

If we compare the proposed incremental method with the approach of Werner and Schütz[73] or Head-Gordon and Subotnik,[74] we obtain more accurate correlation energies if the incremental series is truncated in a proper way. On the other hand the approach of Werner and Schütz as well as the approach of Subotnik and Head-Gordon scale linearly with the system size. Therefore we conclude that the proposed incremental scheme is somewhat in the middle between the efficiency of the local coupled cluster methods and the accuracy of the parallel CCSD(T) implementations.

5. Conclusion

We introduced a new efficient variant of the incremental scheme to calculate CCSD(T), CCSD, and MP2 correlation energies of large systems. We demonstrated for CCSD(T), CCSD, and MP2 energies that chemical accuracy can be reached at reduced computational cost, if a dual basis set approach is followed and the incremental expansion is truncated in a proper way. Furthermore we have shown that the disk space and memory requirements are reduced significantly. We have shown that our scheme is systematically improvable by including higher orders of the expansion and by applying better basis sets for the description of the environment orbitals. In addition the approach is inherently parallel with essentially no loss in CPU time due to dependencies of the individual processes. Therefore we conclude that our modified incremental scheme is an alternative way to calculate high-level correlation energies for large systems. Although in the current work we only applied the approach to water clusters, it can also be applied to hydrocarbons and the glycine tetramer as demonstrated elsewhere.[44]

References

(1) Pulay, P.; Saebø, S. *Theor. Chim. Acta* **1986**, *69*, 357.

(2) Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86* (2), 914.

(3) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213.

(4) Boughton, J. W.; Pulay, P. *J. Comput. Chem.* **1993**, *14*, 736.

(5) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286.

(6) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111* (13), 5691.

(7) Maslen, P. E.; Head-Gordon, M. *Chem. Phys. Lett.* **1998**, *283*, 102.

(8) Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 7093.

(9) Maslen, P. E.; Lee, M. S.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *319*, 205.

(10) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592.

(11) Schütz, M. *J. Chem. Phys.* **2000**, *113* (22), 9986.

(12) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661.

(13) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.

(14) Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *123*, 64108.

(15) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 74116.

(16) Auer, A.; Nooijen, M. *J. Chem. Phys.* **2006**, *125*, 24104.

(17) Weijo, V.; Manninen, P.; Jørgenson, P.; Christiansen, O.; Olsen, J. *J. Chem. Phys.* **2007**, *127*, 074106.

294 *J. Chem. Theory Comput., Vol. 5, No. 2, 2009*

Friedrich and Dolg

(18) Doser, B.; Lambrecht, D. S.; Ochsenfeld, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3335.

(19) Hughes, T. F.; Flocke, N.; Bartlett, R. J. *J. Phys. Chem. A* **2008**, *112*, 5994.

(20) Li, W.; Li, S. *J. Chem. Phys.* **2004**, *121*, 6649.

(21) Federov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.

(22) Kamiya, M.; Hirata, S.; Valiev, M. *J. Chem. Phys.* **2008**, *128*, 074103.

(23) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.

(24) Stoll, H. *Chem. Phys. Lett.* **1992**, *191*, 548.

(25) Stoll, H. *Phys. Rev. B* **1992**, *46*, 6700.

(26) Stoll, H. *J. Chem. Phys.* **1992**, *97*, 8449.

(27) Nesbet, R. K. *Phys. Rev.* **1967**, *155*, 51.

(28) Nesbet, R. K. *Phys. Rev.* **1968**, *175*, 2.

(29) Nesbet, R. K. *Adv. Chem. Phys.* **1969**, *14*, 1.

(30) Doll, K.; Dolg, M.; Fulde, P.; Stoll, H. *Phys. Rev. B* **1997**, *55*, 10282.

(31) Rosciszewski, K.; Paulus, B.; Fulde, P.; Stoll, H. *Phys. Rev. B* **1999**, *60*, 7905.

(32) Shukla, A.; Dolg, M.; Fulde, P.; Stoll, H. *Phys. Rev. B* **1999**, *60*, 5211.

(33) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, *123*, 144108.

(34) Voloshina, E.; Paulus, B. *J. Phys. Chem.* **2006**, *124*, 234711.

(35) Paulus, B. *Int. J. Quantum Chem.* **2004**, *100*, 1026.

(36) Friedrich, J.; Hanrath, M.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 154110.

(37) Friedrich, J.; Hanrath, M.; Dolg, M. *J. Phys. Chem. A* **2007**, *111*, 9830.

(38) Friedrich, J.; Hanrath, M.; Dolg, M. *Chem. Phys.* **2007**, *338*, 33.

(39) Friedrich, J.; Hanrath, M.; Dolg, M. *Chem. Phys.* **2008**, *346*, 266.

(40) Mödl, M.; Dolg, M.; Fulde, P.; Stoll, H. *J. Chem. Phys.* **1997**, *106*, 1836.

(41) Bezugly, V.; Birkenheuer, U. *Chem. Phys. Lett.* **2004**, *399*, 57.

(42) Friedrich, J.; Hanrath, M.; Dolg, M. *J. Phys. Chem. A* **2008**, *112*, 8762.

(43) Friedrich, J.; Walczak, K.; Dolg, M. *Chem. Phys.* **2008**, doi: 10.1016/j.chemphys.2008.10.030. .

(44) Friedrich, J.; Dolg, M. *J. Chem. Phys.* **2009**, *129*, 244105.

(45) Jurgens-Lutovsky, R.; Almlöf, J. *Chem. Phys. Lett.* **1991**, *178*, 452.

(46) Klopper, W.; Noga, J.; Koch, H.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 164.

(47) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D. Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO, a package of ab initio programs designed* version 2002; Werner H.-J., Knowles, P. J., Eds.; Technical Report; University of Birmingham: Birmingham, England, 2002.

(48) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219.

(49) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **2000**, *112*, 3106.

(50) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.

(51) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

(52) Becke, A. D. *Phys. Rev. A* **1988**, *38* (6), 3098.

(53) Perdew, J. P. *Phys. Rev. B* **1986**, *33* (12), 8822.

(54) Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; Huber, C.; Huniar, U.; Kölmel, C.; Kollwitz, M.; Ochsenfeld, C.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H. *Turbomole 5*; Institut für Physikalische Chemie, Universität Karlsruhe: Karlsruhe, Germany, 2002.

(55) Laasonen, K.; Parrinello, M.; Car, R.; Lee, C.; Vanderbilt, D. *Chem. Phys. Lett.* **1993**, *207*, 208.

(56) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.

(57) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163.

(58) Kim, J.; Majumdar, D.; Lee, H. M.; Kim, K. S. *J. Chem. Phys.* **1999**, *110*, 9128.

(59) Kozmutza, C.; Kryachko, E. S.; Tfirst, E. *THEOCHEM* **2000**, *501*, 435.

(60) Day, P.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063.

(61) Upadhyay, D. M.; Shukla, M. K.; Mishra, P. C. *Int. J. Quantum Chem.* 2001.

(62) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, *427*, 185.

(63) Bulusu, S.; Yoo, S.; Apra, E.; Xantheas, S.; Zeng, X. C. *J. Phys. Chem. A* **2006**, *110*, 11781.

(64) Olson, R. M.; Bentz, J. L. R. A. K.; Schmidt, M. W.; Gordon, M. S. *J. Chem. Theory Comput.* **2007**, *3*, 1312.

(65) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.

(66) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. *J. Chem. Phys.* **1970**, *53*, 4544.

(67) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(68) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(69) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(70) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(71) Janowski, T.; Ford, A. R.; Pulay, P. *J. Chem. Theory Comput.* **2007**, *3*, 1368.

(72) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. *J. Chem. Theory Comput.* **2008**, *4*, 64.

(73) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370.

(74) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 034103.