

# JCTC

Journal of Chemical Theory and Computation

## Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory

Michal Pitoňák<sup>\*,†,‡</sup> and Andreas Heßelmann<sup>\*,§</sup>

*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v. v. i., Flemingovo nám. 2, 166 10 Praha 6, Czech Republic, Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Mlynská Dolina, 842 15 Bratislava 4, Slovak Republic, and Lehrstuhl für Theoretische Chemie, Universität Erlangen-Nürnberg, Egerlandstrasse 3, D-91058 Erlangen*

Received November 5, 2009

**Abstract:** A new method is presented that improves the supermolecular second-order Møller–Plesset (MP2) method for dimer systems with strong dispersion interactions while preserving the generally good performance of MP2 for other types of intermolecular interactions, e.g., hydrogen-bonded systems. This is achieved by adding a correction term to the supermolecular MP2 energy that is determined using time-dependent density functional (TDDFT) response theory and that accounts for the error of the dispersion energy contained in the supermolecular MP2 method. It is shown for the S22 database set of noncovalent complexes and some potential energy curves of noncovalent bound aromatic dimers that the approach gives strong improvements over MP2 if compared to coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) reference energies. An efficient computer implementation of the method is presented that is shown to scale only with the fourth power of the system size and thus leads only to a slightly higher computational cost than that of the supermolecular MP2 itself.

### 1. Introduction

Highly accurate interaction energies of noncovalent bound molecular complexes and clusters are of high interest in the wide community of both computational chemists and experimentalists. However, to achieve the desired, so-called chemical accuracy ( $\sim 1$  kcal/mol) with present wave-function theory (WFT) computational methods is often a very demanding task. For notoriously known problematic interaction types, such as  $\pi$ – $\pi$  stacking, highly sophisticated theories such as CCSD(T) have to be applied to achieve such an accuracy. But even with ever improving computer technologies and algorithms (as implemented in various computer programs such as MOLCAS,<sup>1</sup> PSQ,<sup>2</sup> GAMESS,<sup>3</sup> ACES3,<sup>4</sup> etc.), an  $N^7$  scaling of the CCSD(T) method with

the system size allows the applicability of this method to be widened very slowly. According to a recently published series of benchmark calculations of noncovalent complexes,<sup>5–8</sup> CCSD(T) interaction energies close to the complete basis set (CBS) limit can nowadays be obtained for systems with about 30–50 second row atoms and hydrogen. Though this was just a dream a few years ago, such complexes are still too small to serve even as reliable models in biochemistry, nanoclusters, etc. To extend the investigated systems size beyond models to a “real-life” dimension is clearly impossible. Approximations of the CCSD(T) method within the strict WFT formalism for the calculation of noncovalent interactions is hardly possible. Following the series of increasing order of perturbation theory, i.e., second (MP2), third (MP3, MP4(SD), MP4(SDQ), CCSD), and fourth (MP4, CCSD(T)), none of these methods lower than fourth order is reliable enough.<sup>9</sup> For the problematic types of noncovalent interactions, this series often converges in an oscillatory way, third-order “overcorrecting” errors of the second order, etc. According to this, one could conclude that the proper

\* To whom correspondence should be addressed. E-mail: pitonak@fns.uniba.sk (M.P.); andreas.hesselmann@chemie.uni-erlangen.de (A.H.).

<sup>†</sup> Academy of Sciences of the Czech Republic.

<sup>‡</sup> Comenius University.

<sup>§</sup> Universität Erlangen-Nürnberg.

description of noncovalent interactions in WFT theory necessarily requires at least an  $\mathcal{N}^7$  scaling algorithm. This is clearly impractical for several reasons, one of which is the long-range nature of the intermolecular dispersion contribution that asymptotically decays as  $R^{-6}$  with the distance ( $R$ ) of the interacting fragments.

A possible conceptual remedy lies in the (fourth and higher order) approaches based on the idea of locality of the electron-correlation, which are, however, still not accurate, robust, or “black-box” enough. A major problem in the application of these methods on noncovalent interaction is that certain parts of the interaction energy, mainly the dispersion, is certainly not short-ranged nor can be approximated accurately enough by lower-order scaling methods, as already mentioned. Currently more practical, but computationally still quite expensive, options seem to be the empirical,  $\mathcal{N}^6$  scaling MP2.5<sup>9</sup> and SCS-CCSD methods.<sup>10</sup> Though being quite accurate, as demonstrated in several benchmark calculations,<sup>8–11</sup> chemical accuracy is certainly not guaranteed, and error propagation with increasing molecular complex size is, especially for the latter one, SCS-CCSD, not known yet. Empirical  $\mathcal{N}^5$  scaling methods, mostly based on the idea of SCS-MP2,<sup>12</sup> such as SCSN-MP2,<sup>13</sup> SCS(MI)-MP2,<sup>14</sup> or JMP2,<sup>15</sup> are surprisingly accurate and have a much lower cost than the methods mentioned above, but as they are parametrized for only certain interaction types and molecular complex dimensions, their performance outside their training sets can be quite unpredictable.

A different approach would be to leave the WFT concept and follow the, in general computationally much more economic, route of density functional theory (DFT). Conventional DFT based on the local density approximation (LDA) or the generalized gradient approximation (GGA) does not account, however, for the important long-range correlation or dispersion contribution to the intermolecular interaction energy. The reason for this is that these density functionals are only locally dependent on the density and its gradients and therefore are not able to describe electron correlations between remote parts of the molecular complex. As this failure of standard DFT is widely known, a number of possible extensions have been developed in the past few years.<sup>16–21</sup>

A deeper investigation of the reasons of failure of the otherwise often quite accurate MP2 method for  $\pi$ - $\pi$  stacked complexes was done by Cybulski and Lytle<sup>22</sup> and Hesselmann.<sup>23</sup> The source of error was identified to be the 10–20% overestimated uncoupled Hartree–Fock (UCHF) dispersion energy component of the supermolecular MP2 interaction energy. In both works the same idea of substituting the inaccurate UCHF dispersion energy by the more accurate dispersion energy obtained from either scaled time-dependent Hartree–Fock (TDHF)<sup>22</sup> or time-dependent DFT (TDDFT)<sup>23</sup> has been proposed. Note that this approach is related to recent developments of new types of exchange-correlation functionals that are based on the random phase approximation (RPA),<sup>24–26</sup> because the RPA method itself is known to account for long-range correlation energies on a coupled Hartree–Fock level.<sup>27</sup> Both the approach from ref 23 and

from ref 22 seems to alleviate the problem of the overestimation of the dispersion contribution in the problematic complexes. It was found that the TDDFT-based approach is most accurate if an exchange-only potential from the localized Hartree–Fock method<sup>28</sup> combined with an exchange-only adiabatic local density approximation (ALDA) kernel is employed. In both works, results for several test cases, such as rare-gas dimers, hydrogen-bonded (H-bonded) complexes of small diatomic molecules, and DNA base pairs, were presented, strongly validating this approach. Similar to this approach, Tkatchenko et al.<sup>29,30</sup> recently proposed the so-called MP2+ $\Delta$ vdW method. The main idea is to improve the long-range interaction MP2 potential by using the series of  $\Delta C_n R^{-n}$  ( $\Delta C_6$ ,  $\Delta C_8$ , ..., being differences between the MP2 and the “accurate” dispersion coefficient for  $n = 6, 8, \dots$ ;  $R$  being the distance between the interacting molecules) in combination with a proper damping function for short distances. This approach was shown also to be quite accurate and can, in contrast to the method presented in this work, also be applied to study intramolecular dispersion effects if the scheme presented in refs 29 and 30 is extended by deriving intermolecular dispersion coefficients from atomic contributions. However, as with corresponding DFT+dispersion methods, it relies on an empirically determined damping function which has to reduce the double counting of correlation effects for short intermonomer distances. Moreover, in the MP2+ $\Delta$ vdW method, the multipolar expansion of the long-range dispersion energy is restricted to the  $C_6$  and  $C_8$  terms and will therefore not be accurate for short intermonomer distances.

The main goal of this work is to extend the tests of the new MP2 “coupled” (MP2C) approach from ref 23 to the systematic S22 database of noncovalent complexes of Hobza et al.,<sup>31</sup> as well as to several challenging noncovalent complexes, such as the benzene dimer in several conformations, the H-bonded and stacked uracil dimer, and the methyl-adenine...methyl-thymine dimer (mAmT), for which highly accurate CCSD(T) benchmark interaction energies were published.<sup>5–8</sup> Because the balanced performance, not only of the equilibrium geometries but over a wide area of the potential energy surface (PES), is important, we selected a few cuts through the PES, i.e., potential energy curves (PEC), of the benzene dimer and nitrogen-substituted heterocyclic derivatives of the benzene dimer as well. Finally, we will also present an efficient implementation of the correction scheme, due to which the overall MP2C method is computationally by 1 order of magnitude less demanding than the supermolecular MP2 itself.

## 2. Method

The supermolecular MP2 interaction energy can be obtained from the energy difference:

$$E_{\text{int}}^{\text{MP2}} = E_{\text{AB}}^{\text{MP2}} - E_{\text{A}}^{\text{MP2}} - E_{\text{B}}^{\text{MP2}} \quad (1)$$

where  $E_{\text{AB}}^{\text{MP2}}$  is the total energy of the dimer and  $E_{\text{A,B}}^{\text{MP2}}$  are the monomer energies of the two systems A and B. Note that the use of eq 1 introduces the so-called basis set superposition error (BSSE) if finite basis sets are used. An effective

elimination of this error can however be achieved by using the Boys–Bernardi counterpoise correction<sup>32</sup> in which all individual energy calculations are done using the same (dimer centered) basis set.

Using intermolecular perturbation theory, it was observed a while ago<sup>27,33,34</sup> that the supermolecular MP2 interaction energy of eq 1 contains certain correlation terms that are of second order in the intermolecular interaction, namely the uncoupled Hartree–Fock (UCHF) dispersion energy, the corresponding Hartree–Fock exchange-dispersion energy, and a deformation-correlation term. The exchange-dispersion contribution stems from exchange interactions between the monomers when the monomer distance decreases and vanishes for larger distances while the deformation correlation term includes exchange-penetration, induction, and charge-transfer interactions. The uncoupled Hartree–Fock dispersion energy can exactly be written in terms of the monomer Hartree–Fock orbitals and orbital energies:

$$E_{\text{disp}}^{(2)}(\text{UCHF}) = -4 \sum_{ia,jb} \frac{|(i^A a^A | j^B b^B)|^2}{\varepsilon_a^A - \varepsilon_i^A + \varepsilon_b^B - \varepsilon_j^B} \quad (2)$$

where A, B label monomer A or B, indices  $i, j$  denote occupied orbitals,  $a, b$  denote unoccupied orbitals,  $(ia|jb)$  is a two-electron repulsion integral in chemist's notation and  $\varepsilon_i$  is the orbital energy of orbital  $i$ . Note that closed-shell formalism will be used throughout. Using the Casimir–Polder integral transform, it can easily be seen that eq 2 can be rewritten as:

$$E_{\text{disp}}^{(2)}(\text{UCHF}) = -\frac{1}{2\pi} \int_0^\infty d\omega \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4 \chi_0^A(\mathbf{r}_1, \mathbf{r}_3, \omega) \chi_0^B(\mathbf{r}_2, \mathbf{r}_4, \omega) \frac{1}{r_{12} r_{34}} \quad (3)$$

where the  $\omega$ -integral runs over imaginary frequencies  $\omega$  and  $\chi_0^A$  and  $\chi_0^B$  are the uncoupled Hartree–Fock response functions of monomers A and B given by:

$$\chi_0(\mathbf{r}_1, \mathbf{r}_2, \omega) = 4 \sum_{ia} \frac{\varepsilon_{ia}}{\varepsilon_{ia}^2 + \omega^2} \phi_{ia}(\mathbf{r}_1) \phi_{ia}(\mathbf{r}_2) \quad (4)$$

with the occupied-virtual orbital products  $\phi_{ia}(\mathbf{r}) = \phi_i(\mathbf{r})\phi_a(\mathbf{r})$  and  $\varepsilon_{ia} = \varepsilon_a - \varepsilon_i$ . It is well-known that the dispersion energy on the UCHF level (eq 2) often poorly describes the dispersion energy, e.g., in case of stacked  $\pi$ – $\pi$  interactions the uncoupled HF dispersion energy can overestimate the dispersion energy by 15% and more.<sup>22,35,36</sup> Because of this, in refs 22 and 23 a correction was introduced to the supermolecular MP2 interaction energy that replaces the implicitly included UCHF dispersion contribution with the coupled dispersion energy on the time-dependent Hartree–Fock (TDHF) or time-dependent density-functional (TDDFT) level. Using coupled TDDFT dispersion energies, the corrected MP2 interaction energies are thus obtained with:

$$E_{\text{int}}^{\text{MP2C}} = E_{\text{int}}^{\text{MP2}} - E_{\text{disp}}^{\text{UCHF}} + E_{\text{disp}}^{\text{TDDFT}} \quad (5)$$

with the acronym MP2C denoting MP2 ‘coupled’. The decomposition of the supermolecular MP2 energy into the

uncoupled HF dispersion energy and a remainder term is based on an explicit decomposition of the total system into two subsystems. Because of this, the approach described in this work is not capable to improve the MP2 method also for intramolecular dispersion effects. The dispersion energies from the TDDFT method can be obtained from:

$$E_{\text{disp}}^{(2)}(\text{TDDFT}) = -\frac{1}{2\pi} \int_0^\infty d\omega \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4 \chi_{\text{coup}}^A(\mathbf{r}_1, \mathbf{r}_3, \omega) \chi_{\text{coup}}^B(\mathbf{r}_2, \mathbf{r}_4, \omega) \frac{1}{r_{12} r_{34}} \quad (6)$$

where  $\chi_{\text{coup}}^{A,B}$  denote the coupled response functions of monomers A and B which can be obtained from the Dyson-type equation:

$$\chi_{\text{coup}}(\mathbf{r}_1, \mathbf{r}_2, \omega) = \chi_0(\mathbf{r}_1, \mathbf{r}_2, \omega) + \int d\mathbf{r}_3 d\mathbf{r}_4 \chi_0(\mathbf{r}_1, \mathbf{r}_3, \omega) W(\mathbf{r}_3, \mathbf{r}_4, \omega) \chi_{\text{coup}}(\mathbf{r}_4, \mathbf{r}_2, \omega) \quad (7)$$

with  $W$  denoting the interelectronic interaction operator comprising Coulomb, exchange, and correlation effects:

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{1}{r_{12}} + f_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2, \omega) \quad (8)$$

where  $1/r_{12}$  is the Coulomb-operator and  $f_{\text{xc}}$  is the exchange-correlation (xc) kernel that in general is nonlocal and frequency-dependent. In the framework of TDDFT the xc-kernel almost always is approximated by the adiabatic local density approximation (ALDA) kernel:

$$f_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2, \omega) \approx f_{\text{xc}}^{\text{ALDA}}(\mathbf{r}_1) \delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (9)$$

that is the frequency-independent second derivative of the LDA xc-functional. Note that in order to obtain the coupled response functions via eq 7 an iterative procedure has to be used, as  $\chi_{\text{coup}}$  appears on both sides of the equation.

It has to be noted here that the MP2C approach (eq 5) does not account for also correcting the corresponding exchange-dispersion energy term ( $E_{\text{exch-disp}}^{(2)}$ ) that is also described on an uncoupled Hartree–Fock level only in the supermolecular MP2 method. The calculation of  $E_{\text{exch-disp}}^{(2)}$  is much more computationally demanding than the calculation of the dispersion energy and would therefore lead to a much more expensive method if explicitly corrected in addition. However, the exchange-dispersion energy is generally much smaller than the dispersion energy itself and decreases exponentially for larger intermonomer distances. It has been found that in the intermediate distance range the ratio between coupled and uncoupled exchange-dispersion energies is about the same as with the corresponding dispersion energies. Because, in contrast to the dispersion energy, the  $E_{\text{exch-disp}}^{(2)}$  contribution is always positive, it is found that in most cases the uncoupled exchange-dispersion energy overestimates the coupled one and leads to slightly higher total intermolecular interaction energies. This may be the reason why it has been found that the correction scheme of eq 5 works best if the TDDFT dispersion energy is calculated with an exchange-only approach and not by using more accurate exchange-correlation (xc) potentials and kernels for the calculation of the coupled response functions of eq 7.<sup>23</sup> The

TDDFT dispersion energies on the exchange-only level usually are only slightly less negative than the uncoupled HF dispersion energies, while the corresponding dispersion energies from more accurate xc potentials deviate more from the uncoupled HF dispersion energies; see ref 23. Thus the error from the coupled exchange-only approximation in the dispersion energy is reduced by accounting for the exchange-dispersion contribution on an uncoupled Hartree–Fock level.

In the following, a density-fitting implementation for the calculation of uncoupled and coupled dispersion energies will be described that scales only with  $\mathcal{N}^3 - \mathcal{N}^4$  with the molecular size  $\mathcal{N}$  if local xc-kernels are used, and that therefore can be used for relatively large molecular systems. We start by transforming the uncoupled response function of eq 4 in a local auxiliary basis set  $g_P$ :

$$(\chi_0)_{PQ} = \sum_{ia} (P|\eta|ia)\lambda_{ia}(ia|\eta|Q) \quad (10)$$

where  $\lambda_{ia} = 4\varepsilon_{ia}/(\varepsilon_{ia}^2 + \omega^2)$  and  $(P|\eta|ia)$  is a three-index integral over an auxiliary basis function  $g_P$  (indices  $P, Q, R, \dots$  are labeling auxiliary basis functions) and the occupied-virtual orbital product  $\phi_{ia}$ . The operator  $\eta$  conventionally is chosen as the Coulomb-operator:  $\eta = 1/r_{12}$ . The idea is now that the number of auxiliary basis functions  $N_{\text{aux}}$  that are used to span the occupied-virtual space is much smaller than the product of occupied times virtual orbitals:  $N_{\text{aux}} \ll N_{\text{occ}} \times N_{\text{virt}}$ . As the number of auxiliary functions only increases linearly with the system size, it can be shown that the computation of the matrix  $\chi_0$  scales only as  $\mathcal{N}^4$ . Using eq 10, the uncoupled dispersion energy of eq 3 can be written as:

$$E_{\text{disp}}^{\text{UCHF}} = -\frac{1}{2\pi} \int_0^\infty d\omega (\mathbf{S}^{-1} \chi_0^{\text{A}} \mathbf{S}^{-1}) \mathbf{J} (\mathbf{S}^{-1} \chi_0^{\text{B}} \mathbf{S}^{-1}) \mathbf{J} \quad (11)$$

where  $S_{P,Q} = (P|\eta|Q)$  is the metric matrix in the auxiliary basis set and  $J_{PQ} = (P|1/r_{12}|Q)$  is a two-indexed Coulomb matrix in the auxiliary basis. It can readily be seen that the computation of the dispersion energy using eq 11 scales only with  $\mathcal{N}^3$ .

In order to obtain the coupled response functions in the auxiliary basis set, the Dyson eq 7 is expanded in this basis and one obtains:

$$\chi_{\text{coup}} = \chi_0 + \chi_0 \mathbf{W} \chi_{\text{coup}} \quad (12)$$

where  $\chi_0$  is defined in eq 10 and

$$W_{PQ} = \left( P \left| \frac{1}{r_{12}} \right| Q \right) + (P|f_{\text{xc}}|Q) \quad (13)$$

is the interaction operator in the auxiliary basis set. The solution to eq 12 can easily be calculated:

$$\chi_{\text{coup}} = \chi_0 \mathbf{S}^{-1} \mathbf{W} (\mathbf{S} - \chi_0 \mathbf{S}^{-1} \mathbf{W})^{-1} \chi_0 \quad (14)$$

requiring only an inversion of a matrix of the dimension  $N_{\text{aux}}$ . The dispersion energy can then be obtained analogous to eq 11 by:

$$E_{\text{disp}}^{\text{TDDFT}} = -\frac{1}{2\pi} \int_0^\infty d\omega (\mathbf{S}^{-1} \chi_{\text{coup}}^{\text{A}} \mathbf{S}^{-1}) \mathbf{J} (\mathbf{S}^{-1} \chi_{\text{coup}}^{\text{B}} \mathbf{S}^{-1}) \mathbf{J} \quad (15)$$

It has been found that the computation of the two-indexed xc-kernel integrals  $(P|f_{\text{xc}}|Q)$  turns out to be the computational bottleneck in the calculation of the dispersion energy. As this contribution usually is determined using numerical quadrature, the total cost of its computation is  $N_{\text{grid}} \times N_{\text{aux}}^2$  where  $N_{\text{grid}}$  is the number of grid points. Though the scaling is only of the order  $\mathcal{N}^3$ , the prefactor is rather high because  $N_{\text{grid}}$  usually has values of  $10^5$  to  $10^6$  for larger molecular systems. Therefore, here an alternative way to compute the xc-kernel integrals is presented that reduces the computational cost of numerical quadratures by introducing a gridfree-based algorithm identical to gridfree DFT methods introduced by Almlöf and others:<sup>37–39</sup> we start by determining the matrix  $(P|\rho|Q)$  where  $\rho$  is the electron density and  $P, Q$  are auxiliary basis functions by:

$$M_{PQ} = (P|\rho|Q) = (PQ|R)[\mathbf{S}^{-1}]_{RS} (S|\eta|\mu\nu) \gamma_{\mu\nu} \quad (16)$$

with  $(PQ|R)$  being a three-index overlap integral over three auxiliary functions,  $\mathbf{S}^{-1}$  is the inverse of the metric,  $(S|\eta|\mu\nu)$  is a three-index integral over one auxiliary function and two atomic-orbital (AO) basis functions, and  $\gamma_{\mu\nu}$  is the density-matrix in the AO basis. The matrix  $\mathbf{M}$  is then transformed into a new matrix  $\tilde{\mathbf{M}}$  using an orthonormal basis set:

$$\tilde{\mathbf{M}}[\rho] = \mathbf{V}^T \mathbf{M}[\rho] \mathbf{V} \quad (17)$$

with  $\mathbf{V}^T \mathbf{S} \mathbf{V} = \mathbf{1}$  (note that here  $S_{PQ} = (P|\delta(\mathbf{r} - \mathbf{r}')|Q)$ ) and diagonalization of  $\tilde{\mathbf{M}}$  gives  $\tilde{\mathbf{M}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$  where  $\mathbf{U}$  contains the eigenvectors and  $\mathbf{\Lambda}$  is a diagonal matrix containing the eigenvalues of  $\tilde{\mathbf{M}}$ . One can then write any matrix  $(\tilde{P}|f(\rho)|\tilde{Q})$  of the orthogonal auxiliary basis and functions  $f(\rho)$  as follows:

$$\tilde{\mathbf{M}}[f(\rho)] = \mathbf{U} f(\mathbf{\Lambda}) \mathbf{U}^T \quad (18)$$

where the matrix  $f(\mathbf{\Lambda})$  is a diagonal matrix containing the function values  $f(\Lambda_i)$  for each eigenvalue  $\Lambda_i$  in its diagonal. In case of  $f = f_{\text{xc}}^{\text{ALDax}}$  (exchange-only ALDA kernel) the function  $f$  is given by  $f_{\text{xc}}^{\text{ALDax}}(\rho) = -C_x \rho^{-2/3}$  and  $C_x$  is the Slater–Dirac constant. Finally the matrix  $\tilde{\mathbf{M}}[f(\rho)]$  has to be backtransformed to the original nonorthogonal auxiliary basis set using:

$$\mathbf{M}[f(\rho)] = \mathbf{S} \mathbf{V} \mathbf{U} f(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{V}^T \mathbf{S} \quad (19)$$

It can be seen that the computational cost of the gridfree-based algorithm is only  $N_{\text{aux}}^3$  and  $N_{\text{aux}} \times N_{\text{orb}}^2$  ( $N_{\text{orb}}$ : number of AO basis functions) with both  $N_{\text{aux}} \ll N_{\text{grid}}$  and  $N_{\text{orb}} \ll N_{\text{grid}}$ .

### 3. Computational Details

Geometries of complexes from the S22 database<sup>31</sup> as well as the geometries of all conformers of the benzene dimer<sup>5</sup> (“TT”, “T”, “PD”, and “S”), the uracil dimer<sup>7</sup> (“HB” and “S”), and the stacked methyl-adenine...methyl-thymine dimer<sup>8</sup> (“mAmT”) were taken from respective original references.



**Table 1.** Total MP2, MP2C, and CCSD(T) Interaction Energies (in kcal/mol) for the S22 Complexes. MP2/CBS S22<sup>a</sup> and CCSD(T)/CBS Values Were Taken from Ref 31<sup>a</sup>

structure	MP2				MP2C				CCSD(T)
	aDZ	aTZ	aQZ	CBS (CBS S22)	aDZ	aTZ	aQZ	CBS	CBS
(NH <sub>3</sub> ) <sub>2</sub>	-2.68	-2.99	-3.09	-3.16(-3.20)	-2.73	-3.11	-3.23	-3.32	-3.17
(H <sub>2</sub> O) <sub>2</sub>	-4.37	-4.69	-4.86	-4.98(-5.03)	-4.38	-4.77	-4.96	-5.10	-5.02
(formic acid) <sub>2</sub>	-15.99	-17.55	-18.14	-18.57(-18.60)	-15.90	-17.62	-18.28	-18.76	-18.61
(formamide) <sub>2</sub>	-13.95	-15.03	-15.50	-15.84(-15.86)	-13.97	-15.20	-15.72	-16.10	-15.96
(uracil) <sub>2</sub> HB <sup>b</sup>	-18.41	-19.60	-20.07	-20.41(-20.43)	-18.22	-19.55	-20.06	-20.43	-20.47
2-PO-2-AP <sup>c</sup>	-15.55	-16.64	-17.06	-17.37(-17.37)	-15.25	-16.48	-16.95	-17.29	-16.71
A...T <sup>d</sup> HB	-14.70	-15.80	-16.23	-16.54(-16.54)	-14.52	-15.76	-16.24	-16.59	-16.46 <sup>e</sup>
(CH <sub>4</sub> ) <sub>2</sub>	-0.39	-0.46	-0.48	-0.49(-0.51)	-0.44	-0.52	-0.55	-0.57	-0.53
(ethene) <sub>2</sub>	-1.17	-1.46	-1.53	-1.58(-1.62)	-1.18	-1.50	-1.58	-1.64	-1.51
benzene...CH <sub>4</sub>	-1.47	-1.71	-1.77	-1.81(-1.86)	-1.17	-1.44	-1.51	-1.56	-1.50
(benzene) <sub>2</sub> PD <sup>f</sup>	-4.25	-4.70	-4.85	-4.96(-4.95)	-1.93	-2.48	-2.65	-2.77	-2.73
(pyrazine) <sub>2</sub>	-6.00	-6.56	-6.76	-6.91(-6.90)	-3.46	-4.04	-4.28	-4.46	-4.42
(uracil) <sub>2</sub> S <sup>g</sup>	-9.81	-10.63	-10.90	-11.10(-11.15)	-8.10	-8.99	-9.32	-9.56	-9.88
indole...benzene S	-7.14	-7.74	-7.94	-8.09(-8.12)	-3.56	-4.27	-4.51	-4.69	-4.66 <sup>e</sup>
A...T S	-13.24	-14.26	-14.59	-14.83(-14.93)	-9.71	-10.81	-11.21	-11.50	-12.23
ethene...ethine	-1.39	-1.58	-1.63	-1.67(-1.69)	-1.31	-1.52	-1.57	-1.61	-1.53
benzene...H <sub>2</sub> O	-2.98	-3.35	-3.46	-3.54(-3.61)	-2.69	-3.12	-3.24	-3.33	-3.28
benzene...NH <sub>3</sub>	-2.21	-2.52	-2.60	-2.66(-2.72)	-1.87	-2.21	-2.30	-2.37	-2.35
benzene...HCN	-4.38	-4.92	-5.06	-5.16(-5.16)	-3.85	-4.44	-4.59	-4.70	-4.46
(benzene) <sub>2</sub> T <sup>h</sup>	-3.10	-3.46	-3.56	-3.63(-3.62)	-2.32	-2.72	-2.82	-2.89	-2.74
indole...benzene T	-6.10	-6.71	-6.86	-6.97(-7.03)	-5.00	-5.67	-5.84	-5.96	-5.73
(phenol) <sub>2</sub>	-6.79	-7.36	-7.59	-7.76(-7.76)	-6.19	-6.85	-7.11	-7.30	-7.05

<sup>a</sup> aXZ stands for Dunning's<sup>42</sup> aug-cc-pVXZ basis sets, X = D, T, Q. MP2/CBS and MP2C/CBS values correspond to Helgaker's<sup>47</sup> extrapolations from the aTZ and aQZ basis sets. <sup>b</sup> HB = hydrogen-bonded. <sup>c</sup> 2-PO-2-AP = 2-pyridoxine-2-aminopyridine. <sup>d</sup> A...T = adenine...thymine. <sup>e</sup> In ref 31 for A...T HB and indole...benzene S complexes, -16.37 and -5.22 kcal/mol estimated CCSD(T)/CBS results were calculated as MP2/CBS +  $\Delta$ CCSD(T)/dz. Because the  $\Delta$ CCSD(T)/dz values, 0.21 and 2.90 kcal/mol, are significantly underestimated for these systems, more accurate  $\Delta$ CCSD(T)/aDZ values, of 0.08 and 3.46 kcal/mol, are used instead. <sup>f</sup> PD = parallel displaced. <sup>g</sup> S = stacked. <sup>h</sup> T = "T"-shaped -16.37.

PES studies were based on geometries taken from the work of Grimme et al.<sup>40</sup> ("PD-to-IP" and "T-to-S") and on geometries taken from the work of Wang and Hobza,<sup>41</sup> with modified intermonomer distances as explained in detail in section 4.3.

All MP2, MP2C, and CCSD(T) calculations were done in the frozen-core approximation and using the (aug)-cc-pVXZ basis sets ((a)XZ) of Dunning and co-workers.<sup>42</sup> MP2 and MP2C results were obtained with MOLPRO<sup>43</sup> using corresponding JKfit density-fitting basis sets of Weigend<sup>44</sup> for the Hartree-Fock and Localized Hartree-Fock calculations and the MP2fit basis sets from Weigend et al.<sup>45</sup> for the MP2 and TDDFT calculations. For the LHF calculations, the computational-efficient scheme described in ref 46 was used.

CCSD(T) calculations were carried out with the MOLCAS 7<sup>1</sup> package using a Cholesky decomposition of two-electron integrals with a threshold of 10<sup>-6</sup> Hartree on both SCF and the CCSD(T) levels. This calculation setup was previously validated to be in agreement with the exact two-electron integral-based calculations beyond 0.01 kcal/mol in interaction energies.

## 4. Results and Discussion

**4.1. S22 Test Set.** Tables 1–3 show results obtained for the S22 data set in two different ways. In Table 1 total MP2 and MP2C interaction energies with the respective aTZ→aQZ extrapolations according to Helgaker et al.<sup>47</sup> are presented along with the estimated CCSD(T)/CBS reference values from ref 31. In Table 2 the total MP2 interaction energies are displayed along with the " $\Delta$ MP2C" and " $\Delta$ CCSD(T)"

energies, defined as  $\Delta X = X - \text{MP2}$ , where X stands for MP2C or CCSD(T). The last table on the S22 results, Table 3, shows a statistical evaluation of errors of estimated MP2/CBS and MP2C/CBS results with respect to the estimated CCSD(T)/CBS for each type of interactions separately as well as for the whole test set.

Let us first analyze the performance of the MP2C method on total interaction energies shown in Table 1. MP2 and MP2C values obtained in the series of aXZ (X = D, T, and Q) basis sets are presented to demonstrate the similar rate of convergence of these methods toward the CBS. Two different numbers appear in the CBS column for MP2, one being obtained by us from the extrapolation from aTZ and aQZ basis sets, another one in parentheses from the original S22 paper.<sup>31</sup> These numbers differ slightly, typically a few hundreds of kcal/mol, at most by 0.1 kcal/mol for the stacked adenine...thymine complex, which is because mostly un-augmented, but of one order of cardinality higher, cc-pVXZ basis sets were used. Another, but less important, deviation could also arise from using different density-fitting basis sets. The performance of MP2C for H-bonded complexes is excellent, just like the performance of uncorrected MP2 itself, with an average error of  $\sim 0.2$  kcal/mol. The only exception is the 2-pyridoxine-2-aminopyridine, for which the error of MP2 and MP2C is 0.66 and 0.58 kcal/mol, respectively. The performance of MP2 and MP2C is, however, dramatically different for dispersion-dominated complexes, as expected. Here the maximum error of MP2C is obtained for the adenine...thymine stacked complex, being -0.73 kcal/mol. Note that the actual error is probably by  $\sim 0.1$  kcal/mol less negative, due to the inconsistency with the MP2/CBS value

**Table 2.** Total MP2 and  $\Delta$ MP2C and  $\Delta$ CCSD(T) Corrections to the MP2 Interaction Energies (in kcal/mol) for the S22 Complexes<sup>a</sup>

structure	MP2			$\Delta$ MP2C					$\Delta$ CCSD(T)	
	aDZ	aTZ	aQZ	aDZ	aTZ	aQZ	S22 <sup>b</sup>		aDZ	S22 <sup>b</sup>
(NH <sub>3</sub> ) <sub>2</sub>	-2.68	-2.99	-3.09	-0.06	-0.12	-0.14	-0.04	(qz)	0.06	0.03
(H <sub>2</sub> O) <sub>2</sub>	-4.37	-4.69	-4.86	-0.02	-0.09	-0.11	-0.01	(qz)	0.03	0.00
(formic acid) <sub>2</sub>	-15.99	-17.55	-18.14	0.09	-0.07	-0.14	0.09	(tz)	0.03	-0.02
(formamide) <sub>2</sub>	-13.95	-15.03	-15.50	-0.02	-0.17	-0.22	-0.02	(tz)	-0.02	-0.11
(uracil) <sub>2</sub> HB <sup>c</sup>	-18.41	-19.60	-20.07	0.19	0.05	0.00	0.22	(tz-fd)	-0.06	-0.03
2-PO-2-AP <sup>d</sup>	-15.55	-16.64	-17.06	0.31	0.16	0.11	0.34	(tz-fd)	0.63	0.66
A...T <sup>e</sup> HB	-14.70	-15.80	-16.23	0.18	0.04	-0.01	0.18	(dz)	0.08	0.21
(CH <sub>4</sub> ) <sub>2</sub>	-0.39	-0.46	-0.48	-0.05	-0.06	-0.07	-0.01	(qz)	-0.03	-0.02
(ethene) <sub>2</sub>	-1.17	-1.46	-1.53	-0.00	-0.04	-0.05	0.05	(qz)	0.10	0.11
benzene...CH <sub>4</sub>	-1.47	-1.71	-1.77	0.30	0.27	0.26	0.35	(tz-fd)	0.35	0.36
(benzene) <sub>2</sub> PD <sup>f</sup>	-4.25	-4.70	-4.85	2.32	2.23	2.20	2.32	(adz)	2.22	2.22
(pyrazine) <sub>2</sub>	-6.00	-6.56	-6.76	2.54	2.52	2.48	2.52	(tz-fd)	2.57	2.48
(uracil) <sub>2</sub> S <sup>g</sup>	-9.81	-10.63	-10.90	1.71	1.63	1.58	1.69	(tz-fd)	1.26	1.28
indole...benzene S	-7.13	-7.74	-7.94	3.57	3.47	3.42	3.08	(dz)	3.46	2.90
A...T S	-13.24	-14.26	-14.59	3.54	3.45	3.38	2.91	(dz)	2.76	2.70
ethene...Ethine	-1.39	-1.58	-1.63	0.08	0.06	0.05	0.10	(tz)	0.16	0.18
benzene...H <sub>2</sub> O	-2.98	-3.35	-3.46	0.29	0.23	0.22	0.30	(tz-fd)	0.28	0.33
benzene...NH <sub>3</sub>	-2.21	-2.52	-2.60	0.35	0.31	0.30	0.37	(tz-fd)	0.33	0.37
benzene...HCN	-4.38	-4.92	-5.06	0.53	0.49	0.47	0.56	(tz-fd)	0.64	0.70
(benzene) <sub>2</sub> T <sup>h</sup>	-3.10	-3.46	-3.56	0.78	0.74	0.73	0.78	(adz)	0.88	0.88
indole...benzene T	-6.10	-6.71	-6.86	1.10	1.05	1.02	0.93	(dz)	1.34	1.30
(phenol) <sub>2</sub>	-6.79	-7.36	-7.59	0.60	0.51	0.48	0.61	(tz-fd)	0.69	0.71

<sup>a</sup>  $\Delta$ CCSD(T)/S22 values were taken from the ref 31. <sup>b</sup> Basis sets used in ref 31 for calculation of  $\Delta$ CCSD(T), i.e., (a)XZ = (aug-)cc-pVXZ; tz-fd = cc-pVTZ with less diffuse d- and all f-functions removed. <sup>c</sup> HB = hydrogen-bonded. <sup>d</sup> 2-PO-2-AP = 2-pyridoxine-2-aminopyridine. <sup>e</sup> A...T = adenine...thymine. <sup>f</sup> PD = parallel displaced. <sup>g</sup> S = stacked. <sup>h</sup> T = "T"-shaped.

**Table 3.** Root Mean Squared Errors (RMS), Mean Absolute Deviations (MAD), Mean Signed Deviations (MSD), and Maximum Absolute Errors (MAX) (in kcal/mol) to the Estimated CCSD(T) Interaction Energies of the S22 Complexes from Table 1<sup>a</sup>

	H-bonded	dispersion	mixed	all
rms	0.26/0.25	1.86/0.34	0.71/0.17	1.20/0.24
MAD	0.16/0.18	1.48/0.24	0.61/0.15	0.78/0.17
MSD	0.08/0.17	1.47/-0.16	0.61/0.15	0.75/0.07
MAX	0.66/0.58	2.87/0.73	1.24/0.25	2.87/0.73

<sup>a</sup> MP2 values are the first and the MP2C errors are the second number in each column.

from S22, as mentioned above. The second largest deviation with respect to the S22 CCSD(T)/CBS was found for the stacked complex of indole...benzene, being -0.53 kcal/mol, which almost completely diminishes (to -0.03 kcal/mol) after improving the quality of the estimated CCSD(T)/CBS benchmark, from -5.22 to -4.66 kcal/mol, by calculating  $\Delta$ CCSD(T) in the aDZ instead of the DZ basis set. The description of "mixed" type complexes by MP2C is the most balanced with average errors of ~0.15 kcal/mol and the largest error occurring for the phenol dimer, with only 0.25 kcal/mol (the error of MP2 is ~0.7 kcal/mol for this system).

A deeper insight into the magnitude and the basis set dependence of the TDDFT dispersion contribution can be acquired from Table 2. Because the MP2C and the CCSD(T) method can both be viewed as "corrections" to MP2, we can formally separate the respective correction term, as already mentioned. For MP2C it would be more methodologically correct to show the subtracted UCHF and the added TDDFT dispersion energy, but this would not serve the assessment we would like to address. The separation of the  $\Delta$ CCSD(T) correction was shown to be extremely helpful

especially in benchmark calculations,<sup>48</sup> because it was observed that the  $\Delta$ CCSD(T) converges typically much faster with the basis set size than the MP2 interaction energy itself, thus making it sufficient to be calculated in small- to medium-sized diffuse basis sets. It would also be advantageous to observe a similar feature in MP2C, so we could avoid TDDFT calculations (though not more expensive than the supermolecular MP2 itself) in large basis sets.

Generally, the  $\Delta$ CCSD(T) correction for typical H-bonded complexes (e.g., the first four complexes in Table 2) is small and clustered around zero kcal/mol (see the results in Table 2), and this trend is slightly underestimated on the MP2C/CBS level. The last three H-bonded complexes have a significantly larger contribution from the dispersion energy, from which the uracil dimer and the adenine...thymine complexes are described fairly accurately, errors being -0.04 and 0.13 kcal/mol, respectively, after recalculating the  $\Delta$ CCSD(T) for the adenine...thymine complex in the aDZ basis set (the error with respect to the original S22 CCSD(T)/CBS value is also acceptable with a value of 0.23 kcal/mol). The most problematic H-bonded complex, with an error of 0.58 kcal/mol, is the already mentioned 2-pyridoxine-2-aminopyridine. So far we do not have an explanation or numerical evidence for such a deviation of the MP2C method for this complex. Perhaps the intermolecular correlation beyond the second order is important for electrostatics, exchange, and deformation contributions. The basis set convergence of  $\Delta$ MP2C for these complexes in percentile scale is rather slow; changes of  $\Delta$ MP2C from the aTZ to the aQZ basis set account for 20 to 125%. These large numbers might be a bit misleading, because the  $\Delta$ MP2C correction changes in absolute values only by -0.04 kcal/mol on average.

**Table 4.** Total MP2C and CCSD(T) Interaction Energies and Respective  $\Delta$ MP2C and  $\Delta$ CCSD(T) Energy Corrections (in kcal/mol) for a Few Selected Benchmark Complexes

structure	MP2C				CCSD(T)				$\Delta$ MP2C			$\Delta$ CCSD(T)		
	aDZ	aTZ	aQZ	CBS	aDZ	aTZ	aQZ	CBS	aDZ	aTZ	aQZ	aDZ	aTZ	aQZ
B <sub>2</sub> <sup>a</sup> TT	−2.53	−2.81	−2.90	−2.97	−2.44	−2.66	−2.75	−2.78 ± 0.03	0.68	0.65	0.63	0.75	0.80	0.79
B <sub>2</sub> T	−2.37	−2.72	−2.82	−2.89	−2.28	−2.57	−2.65	−2.69 ± 0.02	0.72	0.69	0.68	0.82	0.85	0.85
B <sub>2</sub> PD	−2.06	−2.58	−2.74	−2.85	−2.15	−2.49	−2.63	−2.70 ± 0.04	2.20	2.11	2.09	2.11	2.20	2.19
B <sub>2</sub> S	−1.19	−1.58	−1.70	−1.79	−1.27	−1.51	−1.61	−1.64 ± 0.04	1.70	1.61	1.58	1.62	1.68	1.67
U <sub>2</sub> <sup>b</sup> HB	−18.22	−19.55	−20.06	−20.41	−18.43	−19.81	—	−20.50 ± 0.14	0.19	0.05	0.00	−0.02	−0.21	—
U <sub>2</sub> S	−8.10	−8.99	−9.32	−9.55	−8.54	−9.33	—	−9.68 ± 0.11	1.71	1.63	1.58	1.26	1.29	—
mAmT <sup>c</sup>	−11.07	−12.74	—	−13.45	−11.61	−13.06	—	−13.70 ± 0.04	4.71	4.57	—	4.16	4.25 <sup>d</sup>	—

<sup>a</sup> B<sub>2</sub> = benzene dimer, TT = “T”-shaped tilted, T = “T”-shaped, PD = parallel displaced, S = sandwich. For details on structures, see ref 5. <sup>b</sup> U<sub>2</sub> = uracil dimer, HB = hydrogen-bonded, S = stacked. For details on structures, see ref 7. <sup>c</sup> mAmT = methyl-adenine...methyl-thymine. For details on structure, see ref 8. <sup>d</sup>  $\Delta$ SCS(MI)-MP2/cc-pV(DT)Z and  $\Delta$ SCS(MI)-MP2/cc-pVQZ values are 2.87 and 2.65 kcal/mol, respectively.

Trends in the basis set convergence of the  $\Delta$ MP2C term can be better observed for the dispersion-dominated complexes, because of much larger absolute values of the interaction energies, especially for the  $\pi$ - $\pi$  stacked complexes. First of all, in all  $\pi$ - $\pi$  stacked complexes investigated by us<sup>5-8</sup> or other authors (for instance see ref 49), the  $\Delta$ CCSD(T) (thus  $\Delta$ MP2C as well) values are repulsive. What is different is that the  $\Delta$ CCSD(T), unlike the  $\Delta$ MP2C, converges with the basis set from below, i.e., the repulsion is increasing with the basis set size. Surprisingly (see, for instance, results for the stacked uracil dimer in Table 2), the  $\Delta$ MP2C correction decreases when going from the aDZ to the aQZ basis set. This means that for these kinds of complexes the total MP2C interaction energy (in attractive equilibrium complex geometries) is increasing with the basis set size, unlike the total CCSD(T) interaction energy, which is decreasing. Because we do not have systematic CCSD(T) results for the whole S22 set, few selected complexes, shown in Table 4, will be analyzed separately in section 4.2. The  $\Delta$ MP2C values for dispersion-dominated complexes converge rather fast, both in percentile and absolute scale. The average change of  $\Delta$ MP2C for the dispersion-bound complexes (methane and ethene dimers, benzene...methane complex) from the aDZ to the aTZ basis set is −0.03 kcal/mol (~15%), while from the aTZ to the aQZ basis set it is only −0.01 kcal/mol (~13%). For the stacked complexes (benzene, pyrazine, and uracil dimers and adenine...thymine complex), the average change of the  $\Delta$ MP2C from aDZ to aTZ basis set is only 0.08 kcal/mol (~3%), while from aTZ to aQZ it is even less, 0.05 kcal/mol (~2%). This is similar to the rate of convergence of the  $\Delta$ CCSD(T) term, as discussed further. The largest error from the dispersion dominated complexes as well as for the whole S22 set is obtained for the adenine...thymine stacked complex, estimated to be about 0.6 kcal/mol. As in case of 2-pyridoxine-2-aminopyridine, we have no clear explanation where this discrepancy stems from. We can only hypothesize that here maybe contributions to the intermolecular interaction other than the dispersion energy are approximated poorly on the intra- and/or intermolecular MP2 level. The convergence of  $\Delta$ MP2C for the “mixed” complexes is also monotonous and “from above” (the repulsion due to the corrected dispersion energy decreases with the basis set size). Changes of  $\Delta$ MP2C from aDZ to aTZ are on average 0.05 kcal/mol (~9%), while

from the aTZ to aQZ basis sets, the average change decreases to only 0.02 kcal/mol (~4%).

A statistical evaluation of errors via rms (root-mean-squared deviation), MAD (mean absolute deviation), MSD (mean signed deviation), and MAX (maximum absolute error) is summarized in Table 3. It is found that errors similar to those of the SCS(MI)-MP2<sup>14</sup> or the MP2.5<sup>9</sup> are obtained. However, both of these two methods contain empirical parameters, and furthermore the spin energy component scaling parameters utilized in the SCS(MI)-MP2 method were optimized exactly for the S22 data set. The performance of the first of these methods on complexes outside its training set is clearly deteriorating, as shown in the footnote of Table 4. Results for several systems from the S22 can also be compared with those used for the testing of the MP2+ $\Delta$ vDW method in ref 29. For water and ammonia H-bonded dimers errors of the best and the worst performing damping functions are 0.09–0.21 kcal/mol and 0.09–0.23 kcal/mol, respectively. For “mixed” and dispersion bound complexes such as “T”-shaped benzene dimer and parallel-displaced (PD) conformations, errors of −0.09–0.35 kcal/mol and −0.05–0.60 kcal/mol were obtained.

**4.2. Other Benchmarks.** The growing number of highly accurate, systematically calculated CCSD(T)/CBS benchmarks for medium-sized (according to the applicability of the coupled-cluster theory) noncovalent complexes allows us to assess the accuracy of the MP2C approach even more critically. Table 4 shows the comparison of MP2C with the CCSD(T)/CBS results gathered from a few publications.<sup>5,7,8</sup> The first four lines show the performance of the MP2C method on various conformers of the benzene dimer. The rates of convergence of the  $\Delta$ MP2C and  $\Delta$ CCSD(T) terms are very similar, values being practically converged toward the CBS already in the aTZ basis set. A further increase of the basis set size amounts to changes of only 0.01–0.03 kcal/mol. Comparing the MP2C and the CCSD(T) values in the CBS limit, an almost uniform overestimation of 0.15–0.20 kcal/mol by the MP2C method is observed. When  $\Delta$ MP2C and  $\Delta$ CCSD(T) values are compared, for instance, in the aQZ basis set, a similar underestimation of the  $\Delta$ CCSD(T) by 0.09–0.17 kcal/mol is found. What is, however, of key importance is that the CCSD(T)/CBS relative stability of different structures is well reproduced by the MP2C method. PD and T, almost isoenergetic structures according to the

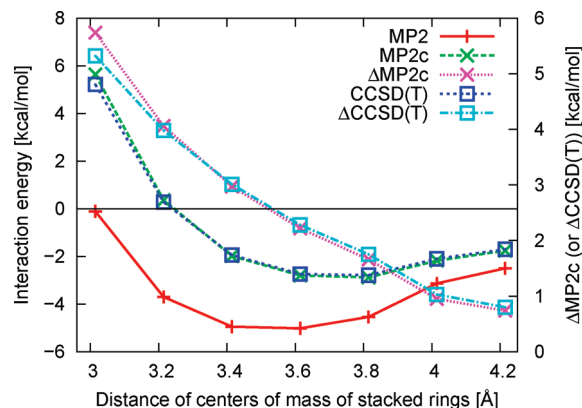


CCSD(T)/CBS results, differ only by 0.04 kcal/mol with MP2C/CBS. The energetic separation of the TT structure from both PD and T, being  $\sim 0.1$  kcal/mol on the CCSD(T)/CBS level, is well reproduced on the MP2C/CBS level, with values of 0.08–0.12 kcal/mol.

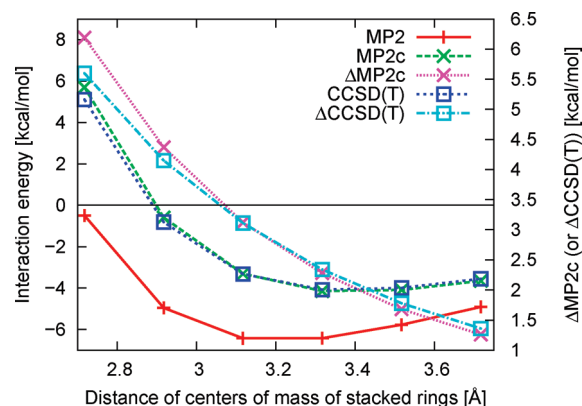
An excellent agreement of the MP2C/CBS and the CCSD(T)/CBS within  $\sim 0.1$  kcal/mol is achieved also for the H-bonded and stacked uracil dimers. For these complexes the CCSD(T)/aQZ results are not available because of the enormous computational requirements. At least for the stacked structure, judged according to the convergence of the  $\Delta$ CCSD(T) correction in aDZ and aTZ basis sets, the estimated CCSD(T)/CBS value is supposed to be very close to its exact basis set limit. The accuracy of the CCSD(T)/CBS value for the H-bonded dimer is, however, a bit more uncertain.

The last of the complexes in Table 4, the stacked methyl-adenine...methyl-thymine dimer, was thoroughly investigated in ref 8, motivated by its previously estimated<sup>31</sup> large repulsive higher-order correlation contribution to the interaction energy. The difference between  $\Delta$ CCSD(T) and  $\Delta$ MP2C is decreasing from  $-0.55$  kcal/mol in the aDZ to  $-0.32$  kcal/mol in the aTZ basis set. Just like for the rest of the complexes in Table 4, the same inverse slope of convergence of  $\Delta$ MP2C compared to  $\Delta$ CCSD(T), as for the S22 data set, is observed. The total MP2C/CBS interaction energy is finally only slightly underestimated by  $\sim 0.25$  kcal/mol ( $\sim 2\%$ ) compared to the CCSD(T)/CBS value. This is an excellent result, especially taking into consideration that the total MP2C/aTZ calculation of the BSSE corrected interaction energy was done in 12 h on a single Intel Core2 Quad 2.40 GHz processor, while only the (T) part of the coupled-cluster calculation of this complex in the aTZ basis set took almost 7 days on 80 four-core Xeon E5430 2.66 GHz processors.

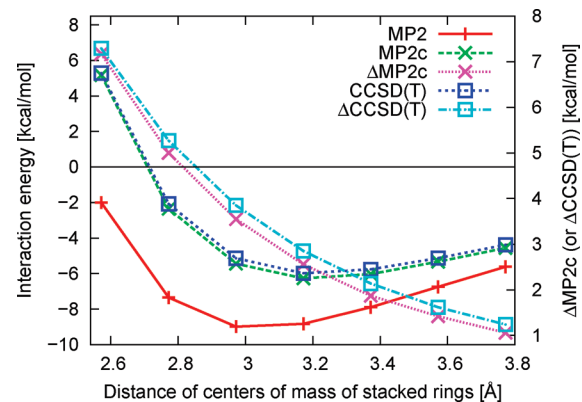
**4.3. PEC for Dimers of Aromatic Systems.** Generally applicable methods for calculating noncovalent interactions should deliver accurate results not only for the minima but also over the wide region of the PES. This is a strong requirement for applications, for instance in biology, where the interacting fragments of protein, DNA, etc., often appear in distorted geometries or where a superposition of weak but numerous middle- to long-range interactions occurs. The first three figures show the “dissociation” of the parallel-displaced conformation of the benzene dimer (Figure 1), benzene...pyrimidine (Figure 2), and benzene...1,2,4,5-pentazine (Figure 3) calculated with MP2C/CBS (aTZ $\rightarrow$ aQZ extrapolation) and estimated with CCSD(T)/CBS (MP2 from aTZ $\rightarrow$ aQZ extrapolation,  $\Delta$ CCSD(T) calculated in the aDZ basis set). For a more refined view,  $\Delta$ MP2C and  $\Delta$ CCSD(T) curves are plotted along the curves of total interaction energies. These systems were selected for known strong  $\pi$ – $\pi$  interactions<sup>41</sup> which increase with the number of nitrogen heteroatoms. The increase of polarity from benzene to 1,2,4,5-pentazine results in a decrease of the optimal vertical displacement of the rings in the complex and consequently an increase of the interaction energy from  $-2.32$  through  $-3.50$  to  $-5.13$  kcal/mol. Starting from the optimized structures of Wang and Hobza<sup>41</sup> (the acronyms “B0P”, “B2P”, and “B5P” were used in this publication), the distance



**Figure 1.** Vertical stretch of stacked benzene rings calculated on MP2C/CBS and estimated CCSD(T)/CBS levels.



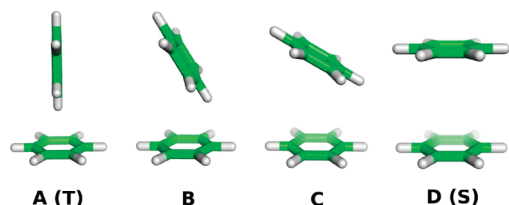
**Figure 2.** Vertical stretch of stacked benzene and pyrimidine rings calculated on MP2C/CBS and estimated CCSD(T)/CBS levels.



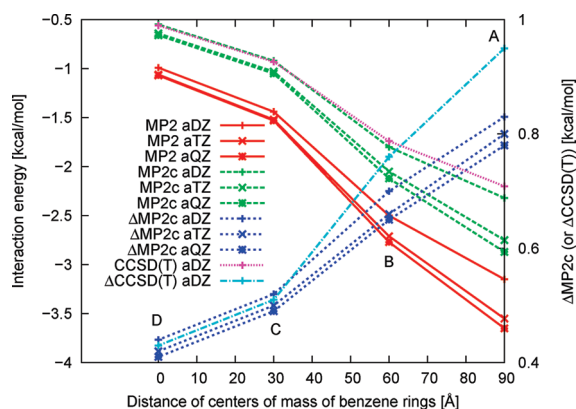
**Figure 3.** Vertical stretch of stacked benzene and 1,2,4,5-pentazine rings calculated on MP2C/CBS and estimated CCSD(T)/CBS levels.

between the centers of mass of the stacked rings was sampled by  $0.2 \text{ Å}$  to cover both repulsive and attractive regions around  $\sim 1.2 R_0$  ( $R_0$  being the equilibrium distance) in both directions. As can be clearly seen from Figures 1–3, MP2C well reproduces the CCSD(T) curve. Some deviations can be seen, for instance, in the strongly repulsive region for benzene...pyrimidine (more clearly visible on the  $\Delta$ MP2C and  $\Delta$ CCSD(T) curves) or as a small constant shift of  $\sim 0.2$  kcal/mol along the whole PEC for benzene...1,2,4,5-pentazine. Deviations in the repulsive region probably originate from

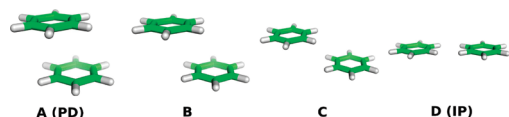




**Figure 4.** Structures on the “T-to-S” benzene dimer conversion potential energy curve.



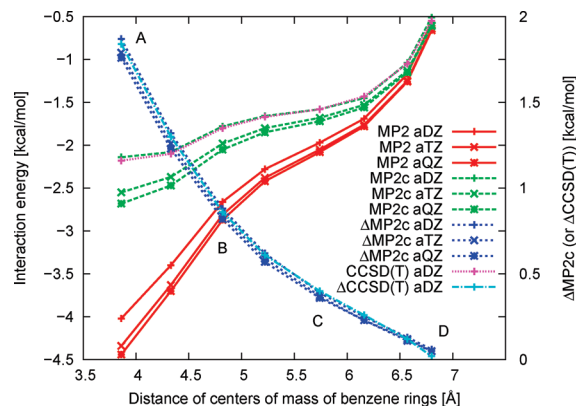
**Figure 5.** “T-to-S” benzene dimer conversion potential energy curve calculated on CCSD(T)/aug-cc-pVDZ and MP2C/aug-cc-pVXZ, X = 2, 3, 4, levels.



**Figure 6.** A few selected structures on the “PD-to-IP” benzene dimer conversion potential energy curve.

the intrinsic error of the MP2C method, while the constant shift on the curve for benzene...1,2,4,5-pentazine could be caused by a deficiency of the aDZ basis set used in the calculation of the estimated CCSD(T)/CBS reference.

The last two examples are PECs of the T-shaped benzene dimer conversion to the sandwich (S) structure, Figure 4 and Figure 5, and the parallel displaced (PD) benzene dimer conversion to the in-plane (IP) arrangement, Figure 6 and Figure 7. Geometries for all structures on these PECs were taken from the work of Grimme et al.<sup>40</sup> On both Figure 5 and Figure 7, MP2 and MP2C total interaction energies in the aDZ to aQZ basis sets are shown, along with the corresponding  $\Delta$ MP2C correction. As references, CCSD(T) and  $\Delta$ CCSD(T) in the aDZ basis set (unlike the estimated CCSD(T)/CBS used in previous figures, i.e., Figure 1–3) are plotted as well. A common feature in Figures 5 and 7 is the trend of convergence of the MP2 and the  $\Delta$ MP2C energies with an increase of the basis set size. For the total MP2 interaction energy, a convergence toward a stronger stabilization is observed and in the case of MP2C a decrease of the “destabilization” due to a decrescent value of  $\Delta$ MP2C occurs. As analyzed by Grimme et al.,<sup>40</sup> the contribution from the intermolecular interaction of  $\pi$ – $\pi$  orbitals is doubled (from 17% to 36% of the MP2/aDZ level) going from the T to the S structure. This can be tracked by more than a doubled rise of the  $\Delta$ CCSD(T) (and  $\Delta$ MP2C as well) in Figure 5.



**Figure 7.** “PD-to-IP” benzene dimer conversion potential energy curve calculated on CCSD(T)/aug-cc-pVDZ and MP2C/aug-cc-pVXZ, X = 2, 3, 4, levels.

The error of the MP2C is also following this trend, being the largest for the sandwich structure, 0.12 kcal/mol in the aDZ basis set. On the basis of the trends of convergence of the  $\Delta$ CCSD(T) and  $\Delta$ MP2C correction terms (see Table 4), this discrepancy has a tendency to further increase upon an improvement of the basis set quality, because they converge in opposite directions. The overall quality of the total MP2C interaction energy is still excellent, with an error compared to the CCSD(T)/aDZ being smaller than 6% along the whole PEC.

The character of interaction is also significantly changing along the PD-to-IP PEC: the PD structure (A) has a strong  $\sigma$ – $\pi$  (49%) and  $\pi$ – $\pi$  (34%) character,<sup>40</sup> while for the IP structure (D) the importance of the intermolecular  $\pi$ – $\pi$  interaction decreases only to 9% ( $\sigma$ – $\pi$  decreases only slightly to 36%). This change of interaction character can be tracked in Figure 7 by a decrease of the  $\Delta$ CCSD(T) and the  $\Delta$ MP2C term from 1.84 and 1.87 kcal/mol for PD to almost zero for the IP structure. However, at the same time the total CCSD(T) and MP2C interaction energies decrease by  $\sim 2$  kcal/mol, because the repulsive  $\Delta$ CCSD(T) and  $\Delta$ MP2C terms correct the overestimation of the interaction energy on the MP2 level occurring only for structures with significant overlap of the stacked rings. MP2C/aDZ accurately reproduces the CCSD(T)/aDZ values along the whole PEC within 0.04 kcal/mol, leading to percentile errors of at most 8% (obtained for the IP structure).

## 5. Conclusions

The recently developed<sup>23</sup> dispersion correction scheme (termed MP2C, i.e., MP2 ‘coupled’, in this work) for the supermolecular MP2 method has been extensively tested for different types of dimer systems containing both strong electrostatic and/or strong dispersion interaction energy contributions. While it is well-known that the supermolecular MP2 method is for some types of noncovalent interactions (such as hydrogen-bonding, etc.) capable of accurately reproducing CCSD(T) interaction energies, it has been shown in this work that the correction used in the MP2C approach does not worsen this performance. In fact it has been found that for the set of seven hydrogen-bonded complexes contained in the S22 benchmark set (see Table 1) the root

mean squared errors to the estimated CCSD(T)/CBS results for MP2 and MP2C are of about the same magnitude and amount to 0.25 kcal/mol.

In contrast to this, a clear difference between the MP2 and MP2C method is observed for the  $\pi$ - $\pi$  stacked complexes where the intermolecular dispersion interactions are dominant. Here the MP2 method does not give reliable estimates of the interaction energy because of large overestimations of the dispersion interaction component contained in the supermolecular MP2 energy. This stems from the fact that the MP2 method accounts for dispersion interactions only on an uncoupled Hartree-Fock level, and this can typically lead to errors of 15% and more in this interaction energy component. In the MP2C method the uncoupled Hartree-Fock dispersion energy contribution is replaced by the coupled dispersion energy calculated using the exchange-only time-dependent density functional theory employing local Hartree-Fock (LHF) orbitals and eigenvalues and the exchange-only adiabatic local density approximation kernel (ALDAx). It has been shown that this approach leads to strong improvements over the MP2 method for the dispersion-dominated and mixed complexes of the S22 dimer set. Root mean square errors of MP2 to the CCSD(T) interaction energies are reduced from 1.86 to 0.36 kcal/mol for the dispersion-dominated complexes and from 0.71 to 0.17 kcal/mol for the mixed type complexes. Therefore, in contrast to the supermolecular MP2 method, the MP2C approach yields accurate and balanced estimates of the intermolecular interaction energy for all diverse types of interactions.

While the S22 benchmark data set contains only complexes in their equilibrium geometries, we have extended the testing of the MP2C method on several cuts through the potential energy surface of the benzene dimer and benzene...pyrimidine and benzene...1,2,4,5-pentazine dimers. In all three cases the MP2C curves do not only agree well with the CCSD(T) reference curves in the region of the minimum of the potential, but they also nicely reproduce the CCSD(T) curves for repulsive and stretched dimer distances. Another challenging test presented here are the potential energy curves corresponding to the transition of the benzene dimer from "T"-shaped to sandwich arrangement and from the parallel-displaced to the in-plane arrangement. CCSD(T) reference curves are reproduced within  $\sim 0.1$  kcal/mol accuracy, despite the fact that the character of interaction, especially for the latter one, is changing from strong  $\pi$ - $\pi$  to  $\sigma$ - $\sigma$  dominated. These examples strongly support the validity of the MP2C approach even for the prediction of wide areas of potential energy surfaces of noncovalent complexes with a good accuracy.

A computer implementation of the correction term used in the MP2C method is presented that scales only as  $N^4$  with the system size and is therefore 1 order of magnitude lower than that of conventional density-fitting MP2 implementations. As an example, the additional computational cost over the standard supermolecular MP2 in the calculation of one of the largest dimers studied in this work, namely the adenine...thymine base pair, amounts to only  $\sim 27\%$  of the total CPU time. Moreover, it was found that the basis set dependence of the  $\Delta$ MP2C correction is less dependent on

the basis set size than the total interaction energy (compare Table 1 and Table 2) and therefore can be, if needed, accurately approximated using smaller basis sets.

**Acknowledgment.** This work was a part of the research project no. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and it was supported by Grants No. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. This work was also supported by the Slovak Research and Development Agency under the contract no. APVV-20-018405, and the Slovak Grant Agency VEGA under the contract no. 1/0428/09. A.H. gratefully acknowledges the funding of the German Research Council (DFG), which, within the framework of its 'Excellence Initiative', supports the Cluster of Excellence 'Engineering of Advanced Materials' ([www.eam.uni-erlangen.de](http://www.eam.uni-erlangen.de)) at the University of Erlangen-Nuremberg.

## References

- (1) Aquilante, F.; Vico, L. D.; Ferre, N.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T.; Pitoňák, M.; Reiner, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Vyzarov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224.
- (2) PQS, version 2.3, Parallel Quantum Solutions, 2013 Green Acres Rd., Fayetteville, AR 72703.
- (3) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
- (4) Lotrich, V.; Flocke, N.; Ponton, M.; Yau, A. D.; Perera, A.; Deumens, E.; Bartlett, R. J. *J. Chem. Phys.* **2008**, *128*, 194104.
- (5) Pitoňák, M.; Neogrády, P.; Rezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829.
- (6) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27.
- (7) Pitoňák, M.; Riley, K. E.; Neogrády, P.; Hobza, P. *Chem. Phys. Chem.* **2008**, *9*, 1636.
- (8) Pitoňák, M.; Janowski, T.; Neogrády, P.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1761.
- (9) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *Chem. Phys. Chem.* **2009**, *10*, 282.
- (10) Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.
- (11) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146.
- (12) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- (13) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80.
- (14) DiStasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073.
- (15) Janesko, B. G.; Scuseria, G. E. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9677.
- (16) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- (17) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.
- (18) Bartlett, R. J.; Lotrich, V. F.; Schweigert, I. V. *J. Chem. Phys.* **2005**, *123*, 062205.

- (19) Furche, F.; Van Voorhis, T. *J. Chem. Phys.* **2005**, *122*, 164106.
- (20) Goll, E.; Werner, H.-J.; Stoll, H.; Leininger, T.; Gori-Giorgi, P.; Savin, A. *Chem. Phys.* **2006**, *329*, 276.
- (21) Chakarova-Käck, S. D.; Schröder, E.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. Lett.* **2006**, *96*, 146107.
- (22) Cybulski, S. M.; Lytle, M. L. *J. Chem. Phys.* **2007**, *127*, 141102.
- (23) Heßelmann, A. *J. Chem. Phys.* **2008**, *128*, 144112.
- (24) Harl, J.; Kresse, G. *Phys. Rev. Lett.* **2009**, *103*, 056401.
- (25) Toulouse, J.; Gerber, I. C.; Jansen, G.; Savin, A.; Angyan, J. *Phys. Rev. Lett.* **2009**, *102*, 096404.
- (26) Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. *J. Chem. Phys.* **2009**, *131*, 034110.
- (27) Szabo, A.; Ostlund, N. S. *J. Chem. Phys.* **1977**, *67*, 4351.
- (28) Della Sala, F.; Görling, A. *J. Chem. Phys.* **2001**, *115*, 5718.
- (29) Tkatchenko, A.; DiStasio, R. A.; Head-Gordon, M.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 094106.
- (30) Tkatchenko, A.; DiStasio, R. A.; Head-Gordon, M.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 129901.
- (31) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (32) Bernardi, F.; Boys, S. F. *Mol. Phys.* **1970**, *19*, 553.
- (33) Chalasinski, G.; Szczesniak, M. M. *Mol. Phys.* **1987**, *63*, 205.
- (34) Cybulski, S. M.; Chalasinski, G.; Moszynski, R. *J. Chem. Phys.* **1990**, *92*, 4357.
- (35) Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
- (36) Heßelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.
- (37) Zheng, Y. C.; Almlöf, J. *Chem. Phys. Lett.* **1993**, *214*, 397.
- (38) Zheng, Y. C.; Almlöf, J. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 277.
- (39) Glaesemann, K. R.; Gordon, M. S. *J. Chem. Phys.* **1998**, *108*, 9959.
- (40) Grimme, S.; Mück-Lichtenfeld, C.; Anthony, J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3327.
- (41) Wang, W.; Hobza, P. *Chem. Phys. Chem.* **2008**, *9*, 1003.
- (42) Kendall, R. A.; Dunning, T. H. J.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
- (43) Werner, H.-J., et al. MOLPRO, version 2009.2, a package of ab initio programs, 2009, see <http://www.molpro.net>.
- (44) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- (45) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.
- (46) Hesselmann, A.; Manby, F. *J. Chem. Phys.* **2005**, *123*, 164116.
- (47) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.
- (48) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.
- (49) Min, S. K.; Lee, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D.; Kim, K. S. *J. Comput. Chem.* **2008**, *29*, 1208.

CT9005882