

# Dispersion Energy Based on Locally Projected Occupied and Excited Molecular Orbitals for Molecular Interaction

Suehiro Iwata\*

The abstract

To avoid the configuration basis inconsistency (CBI) in evaluating the molecular interaction energy, only the dispersion terms are calculated with the second order perturbation expansion based on the locally projected molecular orbitals (LP MO). Because of the local nature of both occupied and excited MOs, the electron configurations for the dispersion terms are generated distinctly from the configurations for the intramolecular electron correlation. Several test calculations are performed for rare gas dimers, water clusters, HF clusters as well as some of S22 sets. In addition, the ion -  $\pi$  interaction is examined.

## 1 Introduction

Efficient and reliable evaluation of the interaction energy between non-covalent atoms and molecules is one of the most important research subjects in quantum chemistry.[1] Last twenty years dramatic progress has been made in theories based both on molecular orbital and on density functional for study of molecular interaction. The theoretical computations become indispensable tools in planning the experiments of the atomic and molecular clusters and in analyzing their results. Most of experimental papers of atomic and molecular clusters contain the computational results of their own and/or of the theoretical collaborators. One of the difficulties in evaluating the weak molecular interaction both with the ab initio molecular orbital (MO) theory and with the density functional theory (DFT) is the basis set superposition error (BSSE), which is inherent in the basis set expansions. In almost all of MO theories and in most of DFT, the one-electron functions (orbitals) are expanded in terms of the Gaussian type functions. For MO theories, the many-electron wave function is constructed with a linear combination of the electron configurations (or Slater determinants); in the simplest case, the closed shell Hartree-Fock wave function is described by a single Slater determinant. Because BSSE results from the imbalance in the approximations for the composite system and for its isolated component systems, the inconsistency both of the one-electron basis set (orbital) and of many-electron Slater determinants (configurations) has to be examined. Liu and McLean, when they studied the interaction potential energy between two He atoms, defined Orbital Basis Inconsistency (OBI) and Configuration Basis Inconsistency (CBI).[2] Their wave function is the multi-reference SCF and first order configuration interaction (CI). To remove or to avoid BSSE, this distinction of OBI and CBI is crucial, although not many papers are aware of it. It should be emphasized that size-consistent and size-extensive many-electron theory[3] do not implies that they are free of CBI, though they are the necessary condition. The counterpoise (CP) procedure of Jansen and Ros[4] and of Boys and Bernardi[5] is widely used to make correction of BSSE, and the procedure can be run routinely with many quantum chemistry packages. However, it is time-consuming and there are some ambiguity in the procedure for polymers consisting of more than two components as noted by White and Davidson[6] and by Valiron and Mayer[7]. Because of the unawareness of the distinction of OBI and CBI, the CP procedure is often implicitly assumed to be equally applicable to remove the BSSE caused by both OBI and CBI. Originally the CP procedure by Jansen and Ros and by Boys and

Bernardi is for the single determinant self-consistent field (SCF) wave function, and so it is intended to remove the OBI.

The alternative to the CP procedure is the extrapolation to the basis set limit, advocated by Varandas,[8] and this procedure is applicable to the correlated wave functions and to any polymers to remove both OBI and CBI. It is very expensive and can be used only for the benchmark calculations. The approximate version of the extrapolation is extensively used by Hobza and his coworkers.[9][10] The symmetry adapted perturbation theory (SAPT) is accurate and BSSE-free.[11]

Another alternative is to use the localized orbitals. The local MP2 of Pulay is one of them,[12] but because the occupied MOs are the canonical MOs localized on atoms, the OBI in the Hartree-Fock level cannot be removed. Mata and Werner reported that the basis set convergence is better than the CP corrected MP2.[13] Last ten years we have been developing the perturbation expansion theory using the locally projected molecular orbital (LP MO).[14][15][16][17][18] The set of equations for LP MO is reformulated in a more transparent form using projection operators from those of "self-consistent field MO for molecular interaction (SCF MI)" of Gianinetti and his coworkers.[19][20] In recent papers,[21][22] we demonstrate that the corrected SCF energy by the third order single excitation perturbation (3rd SPT) is close to the CP corrected SCF energy, in particular with augmented basis sets. For smaller basis sets, by removing some of the excited MOs[21] or by modifying the related matrix elements,[22] the 3rd SPT can be used in place of the CP procedure in estimating the interaction energy of large clusters. In the perturbation theory, the excited MOs, which are local but orthogonal to all of the occupied MOs, are determined through a projection operator. The required computational resource is nearly equal to or less than that of a single cycle of the LP SCF calculation and of the ordinal closed shell SCF calculation.

The dispersion force (Van der Waals force) plays an important role in molecular interaction. To evaluate the dispersion energy within the MO based theory, the multi-determinant description of the wave function is required. Thus, we have to face the CBI. Because the dispersion energy is the intermolecular electron correlation, the separate calculations might be possible. The empirical and semi-empirical methods for estimating the dispersion energy have a long history.[23] Recently the separate calculation of the dispersion energy for SAPT is successfully applied.[24] With the DFT, after a pioneering work by Kamiya et al,[25] many papers report the separate evaluation of the contribution from the dispersion force. Because of theoretical and conceptional difficulties to split the exchange-correlation terms into the intra- and inter-molecular functionals, there are many versions. Sherrill reviews the present status of the methods.[26]

In the MO based quantum chemical calculations, the supermolecule approach is the most common. To evaluate the dispersion energy within the supermolecule approach, the multi-determinant description is required. Now MP2 and its variants are extensively used even among the experimentalists as mentioned above. Because the excited configuration space within the supermolecule approach cannot be divided to identify the dispersion type excitation, the separate calculation of the dispersion energy is not possible. Therefore, the MP2 calculation by taking into account all of the double excitations is carried out and after that, the CP procedure or the extrapolation on the basis sets is performed to make correction of the error. In the LP MO perturbation expansion, on the other hand, the dispersion type excitations can clearly be defined, because both occupied and excited MOs are assigned to each molecular component. So the separate calculations of the contribution from the dispersion terms can be evaluated. The purpose of the present paper is to numerically examine the possibility and to find the deficiency of this method.

## 2 Theoretical

### 2.1 Perturbation expansion theory in terms of locally projected molecular orbitals

The locally projected (LP) self-consistent field (SCF) molecular orbital (MO) method is already described in several literatures.[19][20][14][16][17][18] [21][22][27][28] Below, the equations only relevant to the present work are summarized. The occupied molecular orbitals are locally expanded in terms of the basis sets defined on each molecule (component) which constitutes the cluster. The MO coefficients for the occupied orbitals are variationally determined under this restriction. The restriction ensures that the calculated SCF interaction energy is free of BSSE. But because the wave function is absolutely local, it fails to describe the electron delocalization (charge-transfer, CT) among the molecules in the cluster. The lack of the electron delocalization (CT) causes the large underestimation for the binding energy, [14][17][18][21][27], and therefore the perturbation correction to the energy and wave function is required to incorporate the charge-transfer terms. To evaluate the contribution from the CT terms, the well-defined local excited (virtual) orbitals are required in the expansion. The locally projected excited orbitals are determined by solving the eigenvalue problem,

$$(1 - \hat{P}_{occ})\chi_A t_{Ak}^{ex} = \chi_A t_{Ak}^{ex} \eta_k \quad (1)$$

where column vector  $t_{Ak}^{ex}$  is the MO coefficients for  $\varphi_{Ak}$ , and row vector  $\chi_A$  is the basis functions defined only on molecule A. The projection operator  $\hat{P}_{occ}$  is for the space spanned by all of the occupied MOs of the cluster. If  $\eta_k = 1$ , MO  $\varphi_{Ak} \equiv \chi_A t_{Ak}^{ex}$  is orthogonal to all of the occupied MOs of the cluster under the strong local restriction on the basis sets. The orthogonality to the occupied MOs is required for the perturbation expansion. These MOs can be called the absolutely local excited MO (ALEx MO, or strictly monomer basis excited MO). In the present study, when  $\eta_k \geq 0.99999$ , the orbitals are classified to ALExMO. For  $\eta_l < 0.99999$ , the excited MOs are partially delocalized as

$$\varphi_{Al} = N_{Al}(1 - \hat{P}_{occ})\chi_A t_{Al}^{ex} \quad (2)$$

where  $N_{Al}$  is a normalization factor. Because the vector  $t_{Al}^{ex}$  is an eigenvector of (1), the partially delocalized excited MO is also characterized by the eigenvalue  $\eta_l$ .

Although the MOs are not the canonical orbitals for the full Fock operator, the zero order Hamiltonian  $\hat{H}^0$  is defined similar to the Møller-Plesset form as [17][18][27]

$$\hat{H}^0 = \sum_{b,c}^{occ} \hat{a}_b^\dagger \langle b | \hat{F} | c \rangle \hat{a}_c + \sum_{r,s}^{exct} \hat{a}_r^\dagger \langle r | \hat{F} | s \rangle \hat{a}_s \quad (3)$$

and therefore the perturbation term is split to the one- and two-electron parts as

$$\hat{H} = \hat{H}^0 + \lambda (\hat{V}_1 + \hat{V}_2) \quad (4)$$

$$= \hat{F} + \lambda \hat{V}_2 \quad (5)$$

$$\lambda \hat{V}_1 = \sum_b^{occ} \sum_s^{exct} \hat{a}_b^\dagger \langle b | \hat{F} | s \rangle \hat{a}_s + c.c. \quad (6)$$

We first took into account only term  $\lambda \hat{V}_1$ , [17][18] within the single excitation perturbation (SPT). Then, the contribution from  $\lambda \hat{V}_2$  was added with the third and fourth order perturbation expansion.[21] Several test calculations showed that the fourth order correction is small and is not necessary to be evaluated. To compute the third order correction energy  $\Delta E^{3SPT}$ , the MO integral transformation of two-electron integrals is not required, and  $\Delta E^{3SPT}$  can be evaluated directly from the "AO" integrals; the required computer time is shorter than for a single SCF cycle.[22]

Because both occupied and excited orbitals are local on each component of the clusters, the single and double excitations are grouped to several types. The single excitations are the local excitation (LE) and charge-transfer (CT) types, and the CT terms make correction of the SCF binding energy by allowing the electron delocalization among the components. It is demonstrated that for smaller basis sets the CT terms introduce the BSSE if all of the excited orbitals are used in the perturbation expansion. The excited orbitals originated from the orbitals of small  $\eta$  in eq.(1) causes the error. For these cases, the restriction of the use of the excited orbitals or the modification of the matrix elements is required.[22] For the extensive basis sets such as aug-cc-pVTZ (apvtz) and aug-cc-pVQZ (apvqz), all of the excited orbitals can be used; the CT excitations to the orbitals originated from small  $\eta$  do not contribute to the second and third order correction. With these extensive basis sets, the 3rd order SPT corrected binding energy  $E_{BE}^{3SPT}$  is very close to the CP corrected SCF binding energy  $E_{BE}^{CP-SCF}$ .

The double excitations are classified to several types.[18] The basic excitations are the intra-molecular pair excitations  $|a_A b_A \rightarrow r_A s_A\rangle$  and the dispersion type excitations  $|a_A b_B \rightarrow r_A s_B\rangle$ , where  $a_A$  ( $r_A$ ) is an occupied (excited) orbital localized on molecule A. The other types involve the charge-transfer such as  $|a_A b_B \rightarrow r_B s_B\rangle$ . In the previous work,[18] the preliminary calculations are carried out only for small basis sets, cc-pVDZ(pvdz) and apvdz, and the double excitations involving the CT tends to overestimate the binding energy.

The first order wave function in terms of the non-orthogonal many-electron basis functions  $\Theta_j$  is

$$\Psi^{(1)} = \sum_{j=1} \Theta_j T_j = \Psi_{SPT} + \Psi_{DPT} \quad (7)$$

where  $\Theta_j$  is, for example,  $|a_A \rightarrow r_B\rangle$  for the CT excitation, and  $|a_A b_B \rightarrow r_A s_B\rangle$  for the dispersion type excitation. Because of the non-orthogonality among the occupied orbitals and among the excited orbitals, to obtain the first order wave function vector  $\mathbf{T}$ , a set of linear equations has to be solved,

$$\sum_j \langle \Theta_k | (\hat{H}^0 - E_0) | \Theta_l \rangle T_l = - \langle \Theta_k | \lambda (\hat{V}_1 + \hat{V}_2) | \Phi_{LP-MO} \rangle \quad (8)$$

In a matrix form

$$(\mathbf{\Xi} + \mathbf{W}) \mathbf{T} = -\mathbf{v} \quad (9)$$

where  $\mathbf{\Xi}$  is a diagonal matrix of

$$\Xi_{k,l} = \delta_{k,l} \langle \Theta_k | (\hat{H}^0 - E_0) | \Theta_k \rangle \quad (10)$$

and

$$\mathbf{W}_{kl} = (1 - \delta_{k,l}) \langle \Theta_k | (\hat{H}^0 - E_0) | \Theta_l \rangle \quad (11)$$

The inhomogeneous term is

$$v_k = \langle \Theta_k | \lambda \hat{V}_1 | \Phi_{LP-MO} \rangle \quad \text{for single excitations} \quad (12)$$

$$v_k = \langle \Theta_k | \lambda \hat{V}_2 | \Phi_{LP-MO} \rangle \quad \text{for double excitations} \quad (13)$$

Because of the orthogonality between the occupied and excited orbitals, the linear equations for the single and double excitations are decoupled. For the possible application of spin-component-scaled Møller-Plesset theory,[29] the new code does not use the spin-symmetry adapted form of the many-electron basis functions. The matrix  $\mathbf{W}$  for the parallel- an anti-parallel-spin components is decoupled. The formula of  $\mathbf{W}_{kl}$  are given in Appendix of the previous paper.[18]

The present version of codes is developed using the Intel compiler on Intel Mac Pro (quodra-duo, OS X 10.6.6) and on Linux (quodra-duo, Redhat) as one of the tasks of MOLYX package, which is a private ab initio MO package and uses the integral package of 1997 version of GAMESS.[30] The counterpoise corrections are evaluated with GAUSSIAN 03.[31] All of the computations use the cartesian gaussian functions.

## 2.2 Linear equation solver

The homogeneous part of the linear equation (9) is diagonal-dominant, the equation can be preconditioned as

$$\left(\mathbf{1} + \Xi^{-\frac{1}{2}} \mathbf{W} \Xi^{-\frac{1}{2}}\right) \Xi^{\frac{1}{2}} \mathbf{T} \equiv (\mathbf{1} + \mathbf{\Omega}) \boldsymbol{\varpi} = -\Xi^{-\frac{1}{2}} \mathbf{v} \equiv -\mathbf{v} \quad (14)$$

The solution  $\boldsymbol{\varpi}$  may be approximated by

$$\boldsymbol{\varpi} \simeq c_0 \mathbf{v} + c_1 \mathbf{\Omega} \mathbf{v} + c_2 \mathbf{\Omega}^2 \mathbf{v} + c_3 \mathbf{\Omega}^3 \mathbf{v} + \dots$$

and the coefficients  $c_j$  can be determined by minimizing the error  $|(\mathbf{1} + \mathbf{\Omega}) \boldsymbol{\varpi} + \mathbf{v} - \boldsymbol{\varpi}|$ . Because of a large available core memory, a set of vectors  $\boldsymbol{\rho}^{(j)} \equiv \mathbf{\Omega}^j \mathbf{v}$  are stored on core. The equation for  $c_j$  is

$$\sum_{j=0}^{iter-1} \left(\tilde{\boldsymbol{\rho}}^{(k)} + \tilde{\boldsymbol{\rho}}^{(k+1)}\right) \left(\boldsymbol{\rho}^{(j)} + \boldsymbol{\rho}^{(j+1)}\right) c_j = -\left(\tilde{\boldsymbol{\rho}}^{(k)} + \tilde{\boldsymbol{\rho}}^{(k+1)}\right) \mathbf{v} \quad \text{for } k = 0, \dots, iter$$

where *iter* is the number of the iteration. When the error is smaller than a threshold value, the iteration is stopped. In most case, the maximum iteration is less than 10.

In the present version, the matrix elements  $\mathbf{\Omega}$  larger than a threshold value are stored on RAID0 disk. The rate-determining step is to generate these matrix elements. BLAS, sparse BLAS and LAPACK are used, if they are applicable.

## 3 Computational Results

### 3.1 The third order single excitation perturbation

In the previous work the counterpoise corrected scf binding energy  $E_{BE}^{CP-SCF}$  for various clusters is found to be bracketed by the third order SPT energies,  $E_{BE}^{3SPT}(all)$  and  $E_{BE}^{3SPT}(ALEx)$ , where the former is evaluated using all of the excited orbitals and the latter uses only the absolutely local excited (*ALEx*) MOs. Always  $E_{BE}^{3SPT}(all)$  overestimates  $E_{BE}^{CP-SCF}$  for a given basis set. This is theoretically justified; the excited MO  $\varphi_{Al}$  generated by eq.(2) from the vector of a small eigenvalue  $\eta$  in eq.(1) has a contribution from the basis sets on the components other than component A, and therefore, the CT to this orbital may introduce the BSSE caused by OBI. It should be emphasized that for small basis sets the CT contribution to the binding energy cannot be separated from BSSE caused by OBI. For large clusters, in particular, *ALEx* spans a too restricted excited orbital space, and  $E_{BE}^{3SPT}(ALEx)$  underestimates  $E_{BE}^{CP-SCF}$ . In several model clusters, for the extensive basis sets such as apvtz and apvqz,  $E_{BE}^{3SPT}(all)$  is a good approximation to  $E_{BE}^{CP-SCF}$  even for large clusters, which is close to the extrapolated energy  $E_{BE}^{CP-SCF-extrp}$ . The difference  $(E_{BE}^{3SPT}(all) - E_{BE}^{3SPT}(ALEx))$  is a good measure of the approximation of  $E_{BE}^{3SPT}(all)$  to the counterpoise SCF binding energy  $E_{BE}^{CP-SCF}$  for a given basis set. In this paper, all of the 3rd SPT calculations use all of the excited MOs, otherwise mentioned.

## 3.2 Counterpoise procedure for the correlated levels of theories

The CP procedure is uniquely defined for the dimer. On the other hand, for the polymers, in addition to the original procedure by Jansen and Ros[4](Note that most of the papers, which refer to this pioneering work on the protonation energy of CO, misspell one of the authors name as Ross; the correct one is Ros) and by Boyes and Bernardi,[5] there are other versions; one is the functional counterpoise procedure by White and Davidson[6] and generalized by Valiron and Mayer.[32] Kamiya, Hirata and Valiev adopt a truncated version.[33] The other is the pair-wise additive function scheme by Wells and Wilson.[34] The difference in the correction energies among the procedures may not be so serious, compared with the size of the error itself, as examined by Kamiya et al [33] and by Salvador and Szczesniak.[35] The CP procedure and its variants are assumed to be equally applicable to any MO-based theories and DF theories. It is also assumed that the procedures can be used equally for the single determinant wave function theory as well as for the multi-determinant function MO theories. In other words, it is implicitly assumed that the inconsistency both in the orbital basis (OB) and in the configuration basis (CB) can be corrected by the CP procedures. It is known that the convergence of the CP corrected MP2 binding energy on the basis sets is much slower than that of the CP corrected SCF binding energy. At the level of aug-cc-pVQZ, the latter almost converges, but the former does not. The examples are shown in the following subsections. At least we have to be aware of the different convergence behavior of the CP corrected binding energy for OBI and for CBI. In the extrapolation to obtain the CBS (Complete Basis Set limit, which is a very misleading term, though), the distinct formula for the SCF energy and for the electron correlation energy should be used not only because of the different convergence behaviors based on the nature of the energies[9] but also because of the different origins of the BSSE.

Plaizs et al examine the basis set dependence of the geometry change for MP2 energy with and without CP correction for  $(\text{H}_2\text{O})_2$  and  $(\text{HF})_2$ , and they find occasionally the different converged geometric parameters.[36]

Mayer and Valiron report that the binding energy evaluated by MP2 based on the chemical Hamiltonian approach (CHA-MP2) agree well with those of the CP corrected MP2.[37] It looks to suggest that the CP procedure for MP2 is proved to properly remove the BSSE. But, in terms of the many-electron configuration space, that of CHA-MP2 is same with that of the CP procedure. The CHA modifies only the AO integrals, and thus the BSSE resulting from the OBI can be removed. But it does not imply that the error caused by the CBI is removed. They also shows the examples that the larger basis sets increase the difference between the CP corrected and uncorrected binding energies.[37]

## 3.3 Rare gases

### 3.3.1 He<sub>2</sub>

The interaction potential energy of He dimer is a basic problem in Van der Waals interaction. From the seminar work using quantum Monte Carlo simulations[38] to the latest accurate calculations,[39] there are numerous papers. Here, for consistency with the other rare gas dimers, we examine the basis set dependence of the potential energy curves with the counterpoise corrected CCSD(T) and MP2, and with our 3rd SPT + 2nd order DPT(Dispersion + Pair correlation),  $E_{BE}^{3SPT+Pair+Disp}(all) \equiv E_{BE}^{3SPT}(all) + \Delta E^{Pair+Disp}$ . Figure 1a shows the curves of the CP corrected and uncorrected CCSD(T) with aug-cc-pvXz (X=D, T, Q). Hereafter aug-cc-pvXz is abbreviated to apvXz. Figure 1b shows those for MP2. The two-point extrapolation scheme

$$E_X = E_{extrp} + BX^{-3} \quad (15)$$

of Halkier et al [40] is used for the total energy. In both figures, the curves of apvdz are extremely underestimated for the CP corrected energy and overestimated for the uncorrected energy. The

equilibrium bond length with the CP corrected CCSD(T) and MP2 is too long. Also in both figures, the uncorrected curves converge much faster to the extrapolated curve than the CP corrected one. The two extrapolated curves for He dimer are close to each other. The equilibrium bond length and the binding energy in the accurate calculation is 2.97Å and -11.0 K.[39] The extrapolated curve of CP corrected and uncorrected CCSD(T) has the potential minimum close to the accurate one, but the equilibrium bond length of MP2 is substantially longer than the accurate one. Figure 1c shows the curves of the present  $E_{BE}^{3SPT+Pair+Disp}(all)$  with the apvtz, apvqz and apv5z basis sets. Although the binding energies are substantially underestimated, the equilibrium bond length of these curves are around 3.0Å. By following Hobza’s recommendation,[9] the apv5z curve is improved by adding the correction term estimated by CCSD(T) with apvtz.

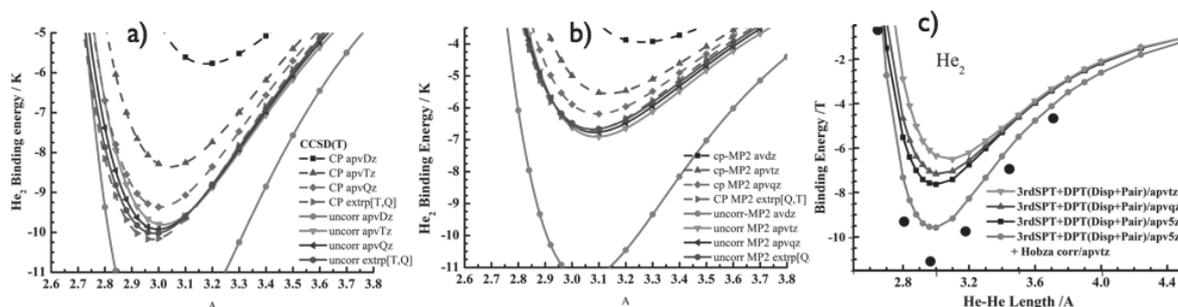


Figure 1. The basis set dependence of the potential energy curve of He<sub>2</sub> with and without CP correction. The energy unit is K. a) CCSD(T), b) MP2, c) 3rd SPT(all) + DPT(Disp+Pair). The reference curve is from [38]

### 3.3.2 Ne<sub>2</sub>

Figures 2a and 2b show the CP corrected and uncorrected potential energy curves of CCSD(T) and MP2. In Figure 2a, the curve deduced from the experiment is included for comparison.[41] The extrapolation is carried out separately for the correlation energy with eq.(15) and the SCF total energy with the three-point form[42]

$$E_X^{SC-SCF} = E_{extrp}^{SC-SCF} + A \exp(-\alpha X) \quad (16)$$

in the figures. Hereafter, [QTP] stands for the three-point extrapolation using the energy obtained with apvqz, apvtz and apvdz, and [QT] for the two-point extrapolation using the energy with apvqz and apvtz. The characteristics similar to the curves of He<sub>2</sub> are found for the CP corrected and uncorrected curves of apvdz both in CCSD(T) and MP2. The CP corrected and uncorrected curves converge to the substantially different curves both with CCSD(T) and MP2. The experimentally deduced curve is bracketed by the CP corrected and uncorrected CCSD(T) curves; the difference of the three curves at the bottom is about 0.04kJ/mol. The extrapolated uncorrected MP2 curve shows an awkward shape, which results from the apvdz curve. Figure 2c shows the curves of the 3rd SPT + 2nd DPT(Disp+Pair),  $E_{BE}^{3SPT+Pair+Disp}(all)$ , and 3rd SPT + 2nd DPT(Disp only),  $E_{BE}^{3SPT+Disp}(all)$ , with apvqz, and the curves of the corresponding extrapolation. They are close to each other and to the CP corrected MP2 in Figure 2b. In the figure the experimentally deduced curve and the CCSD(T) curve extrapolated from cc-pVXZ by Ruedenberg et al are included for comparison.[41] The 3rd SPT + 2nd DPT curves with and without the pair correlation are underestimated by about 0.1kJ/mol, but their equilibrium bond distance and the shape of the curves agree reasonably well with the experimentally derived one. Hobza’s correction improves the agreement.

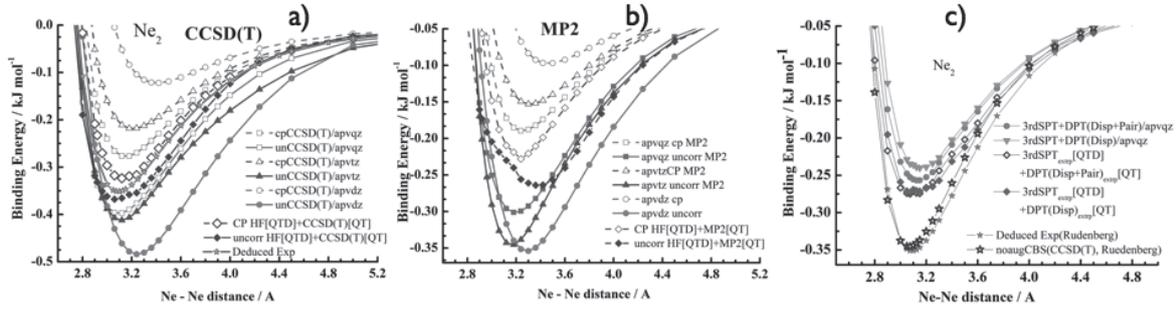


Figure 2. The basis set dependence of the potential energy curve of  $\text{Ne}_2$  with and without CP correction. The energy unit is  $\text{kJ mol}^{-1}$ . a) CCSD(T), b) MP2, c) 3rd SPT(all) + DPT(Disp+Pair) and 3rd SPT(all) + DPT(Disp). The reference curves are from [41].

### 3.3.3 $\text{Ar}_2$

Figures 3a and 3b are the CP corrected and uncorrected potential energy curves of CCSD(T) and MP2. In both figures, the experimentally deduced curve and the accurate curve reported by Koch are included.[43] There are also numerous works on  $\text{Ar}_2$ . [44][45] The characteristics of the curves in Figure 3a and 3b are similar to those of  $\text{Ne}_2$ , but there are some abrupt points of the curves, which might be caused by the improper choice of the parameters for convergence. Because of the unsmooth curves, the extrapolation is not attempted. The uncorrected CCSD(T) curves with apvtz and apvqz are close to each other, and they differ from the experimental curve by about  $5\text{cm}^{-1}$  at the bottom of the curves. The uncorrected MP2 curves with apvtz and apvqz are also very close to each other, but the curves overshoot the accurate curve by about  $15\text{cm}^{-1}$ . It is clearly seen that the CP corrected CCSD(T) and MP2 curves converge much slowly than the uncorrected ones. Figure 3c shows the curves of the  $E_{BE}^{3SPT+Disp}(all)$  with apvdz, apvtz and apvqz, and the curves of the corresponding extrapolation together with the reference curves. The extrapolated curve of  $E_{BE}^{3SPT}(all)[QTD]+\Delta E^{Disp}[QT]$  is, probably fortuitously, very close to the reference curves, and slightly overshoots them, which might be related to the overshoot seen in Figure 3b, because  $\Delta E^{Disp}$  is the MP2 level of theory.

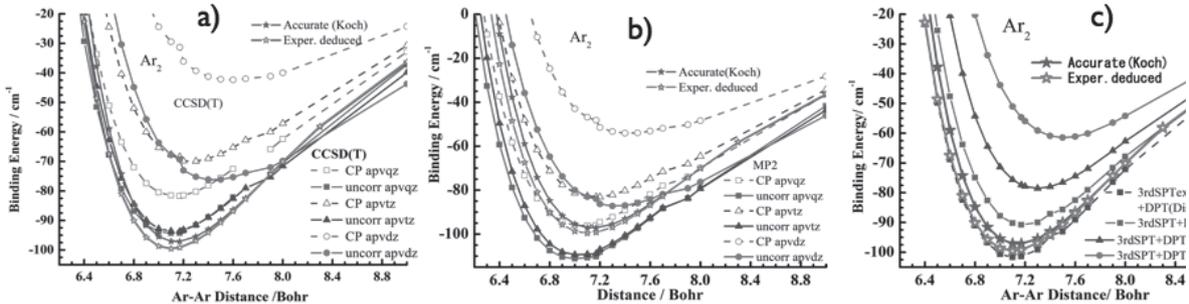


Figure 3. The basis set dependence of the potential energy curve of  $\text{Ar}_2$  with and without CP correction. The energy unit is  $\text{cm}^{-1}$ . a) CCSD(T), b) MP2, c) 3rd SPT(all) + DPT(Disp)

For three of the rare gas dimers, we can find a few common features in CCSD(T) and MP2 calculations, and in the present LP MO 3rd SPT+2nd DPT calculations. The uncorrected curves of apvtz and apvqz are close to each other, and those of CCSD(T) are close to the accurate one. The curves of apvdz both with and without the CP correction differ from the others, which sometimes make it impossible to use it for the extrapolation. The LP MO 3rd SPT + 2nd DPT calculations only with the dispersion terms behaves well with the basis sets, and the extrapolated curves are a reasonably good approximation to the accurate curves of the high quality levels of theory, in particular, if the required computational costs are taken into account.

Table 1. Comparison of the binding energy ( $\text{kJ mol}^{-1}$ ) of some of the S22 set [9]

dimer	CCSD(T) <sup>a</sup>	MP2 <sup>a</sup>	MP2 <sub>CP</sub> <sup>a</sup>	MP2 <sub>CP</sub> <sup>a</sup>	MP2 <sub>uncorr</sub>	$E_{BE}^{3SPT+Disp}$	$E_{BE-extrp}^{3SPT+Disp}$
basis set	CBS-CP	CBS-CP	apv5z	apvqz	apvqz	apvXz	[DTQ, TQ]
(NH <sub>3</sub> ) <sub>2</sub>	-13.26	-13.39	-12.97	-12.64	-13.38	-13.81(Q)	-14.24
(H <sub>2</sub> O) <sub>2</sub>	-21.00	-21.05	-20.46	-19.87	-21.71	-24.34(Q)	-25.13
(CH <sub>4</sub> ) <sub>2</sub>	-2.22	-2.13	-1.92	-1.76	-2.40	-3.30(Q)	–
(C <sub>2</sub> H <sub>4</sub> ) <sub>2</sub>	-6.32	-6.78	-6.57	-5.98	-7.13	-6.76(Q)	–
(C <sub>6</sub> H <sub>6</sub> )(CH <sub>4</sub> )	-6.28	-7.07	-6.78	-6.57	-7.40	-7.65(D)	–

a) Reference [9]

### 3.4 Some of the S22 set

Some of the dimers in the S22 set of Hobza *et al* [9] are examined and summarized in Table 1, where the last three columns are the results of the present study. In parentheses, X of the basis set, aug-cc-pvXz, is given. The  $E_{BE}^{3SPT+Disp}$  in the entry of Table 1 stands for the energy of LP MO 3rd SPT+2nd DPT with the dispersion terms. For comparison, the uncorrected MP2 energy is given. The errors of  $E_{BE}^{3SPT+Disp}$  and its extrapolated energy  $E_{BE-extrp}^{3SPT+Disp}$  from the CCSD(T)<sub>CBS-CP</sub> are less than  $1\text{ kcal mol}^{-1}$  ( $=4.184\text{ kJ mol}^{-1}$ ). The present  $E_{BE}^{3SPT+Disp}$  tends to overestimate the hydrogen bonds in (H<sub>2</sub>O)<sub>2</sub> and (NH<sub>3</sub>)<sub>2</sub>; the similar trend is seen in the following subsections. The uncorrected MP2 energy with apvqz is also close to the CP corrected CCSD(T)<sub>CBS-CP</sub> and MP2<sub>CBS-CP</sub>.

Figure 4 shows the basis set dependence of the binding energy of ammonia dimer (NH<sub>3</sub>)<sub>2</sub> in S22. In the 3rd SPT calculations, the calculations used only the absolutely local excited MO (ALEx) are included in the figure. For cc-pVDZ (pvdz) and cc-pVTZ (pvtz), the binding energy obtained with  $E_{BE}^{3SPT+Disp}(ALEx)$  almost coincides with the CP corrected MP2. For larger basis sets, the difference of the  $E_{BE}^{3SPT+Disp}(all)$  and  $E_{BE}^{3SPT+Disp}(ALEx)$  becomes too small to be seen in the figure. In Figure, the other type of basis sets, Sapporo(spX)-Xz and Sapporo(spX)-aXz, prepared by Sapporo group[46] are tested, and no significance difference in the binding energy from the aug-cc-pvXz series is found for this example. The remarkably small basis set dependence of  $E_{BE}^{3SPT+Disp}(all)$  is seen in the figure; they are between  $-15.7$  and  $-12.3\text{ kJ mol}^{-1}$ , within less than  $1\text{ kcal mol}^{-1}$  from pvdz to apvqz. This is contrast with the large basis set dependence of the CP corrected and uncorrected MP2.

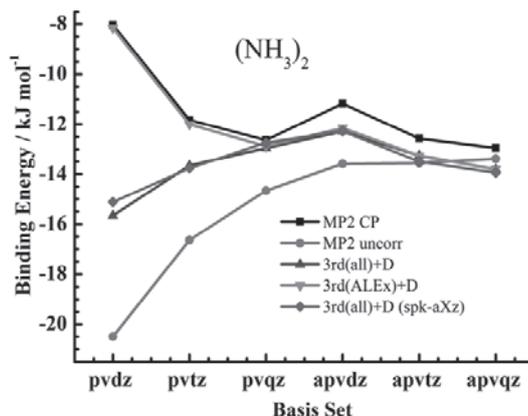


Figure 4. The basis set dependence of the binding energy of (NH<sub>3</sub>)<sub>2</sub> for CP corrected and uncorrected MP2. The 3rd SPT(ALEx)+DPT(Disp) is shown together with the 3rd SPT(all)+DPT(Disp). The Sapporo (spX-aXz) basis sets is also examined.

### 3.5 Water and hydrogen fluoride dimers

The potential energy curves of water dimer ( $\text{H}_2\text{O}$ )<sub>2</sub> and of hydrogen fluoride dimer ( $\text{HF}$ )<sub>2</sub> are examined with CCSD(T) and MP2 as examined for rare gas dimers. The geometric parameters other than the O-O or F-F distance are optimized with the uncorrected MP2/apvdz.[18] The computations to draw the curves with CCSD(T)/apvqz are given up after losing the extensive computer time, because of the ill conversion. Contrarily to the rare gas dimers, the uncorrected CCSD(T) and MP2 curves with apvdz are close to those with apvtz as are seen in Figure 5a and 6a. In the CP corrected CCSD(T) and MP2, on the other hand, the difference of the apvdz and apvtz is substantial. For MP2, the extrapolated curves are inserted in Figure 5b and 6b. The extrapolated curve of the CP corrected MP2 is slightly unstable than the extrapolated MP2 curve by about 1kJ mol<sup>-1</sup> at the bottom. In these figures some of the curves obtained with cc-pvXz (pvxz) are also shown. The uncorrected pvqz curve is close to the curve of apvtz in CCSD(T) and to the curve of apvqz in MP2.

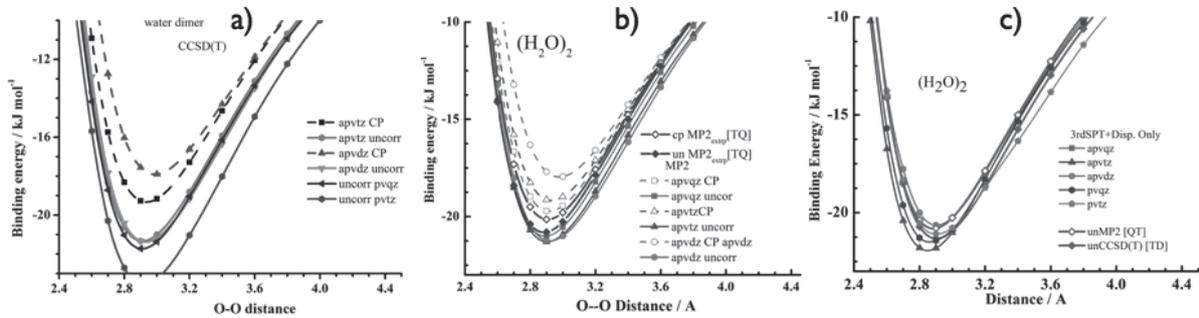


Figure 5. The basis set dependence of the potential energy curve of ( $\text{H}_2\text{O}$ )<sub>2</sub>. a) CCSD(T), b) MP2, c) 3rd SPT + 2nd DPT(Disp).

Figure 5c and 6c the curves of  $E_{BE}^{3SPT+Disp}(all)$  are shown. The extrapolated uncorrected MP2 and CCSD(T) are shown as references. The shape of the curves near the bottom are slightly different from each other, but the binding energy are all within 1kcal mol<sup>-1</sup> of the reference. One of the important findings in Figure 5c and 6c is that  $E_{BE}^{3SPT+Disp}$  curve with apvqz both for ( $\text{H}_2\text{O}$ )<sub>2</sub> and for ( $\text{HF}$ )<sub>2</sub> overshoots the reference curve, and the equilibrium bond distance is shorter than that of the reference. The similar trends are also found  $E_{BE}^{3SPT+Disp}$  curve with apvtz, though it is less than with apvqz. The hydrogen bonds in these dimer are strong, and the CT contribution to the 3rd order SPT is substantial. The separate evaluation of the dispersion energy assumes that the occupied orbitals are not deformed by molecular interaction. This assumption might be broken when the CT contribution to the wave function becomes large. Careful theoretical and numerical studies are further required, in particular, to make it clear why the error becomes large for the extended basis sets.

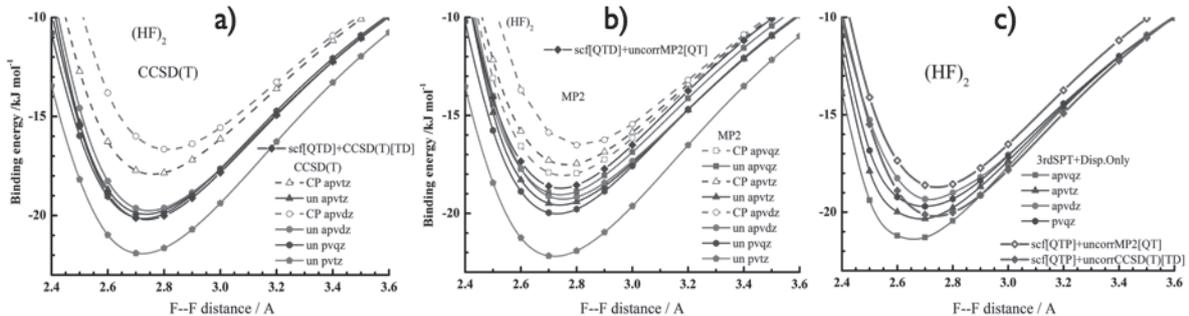


Figure 6. The basis set dependence of the potential energy curve of ( $\text{HF}$ )<sub>2</sub>. a) CCSD(T), b) MP2, c) 3rd SPT + 2nd DPT(Disp).

### 3.6 Hydrogen fluoride clusters

The hydrogen fluoride clusters  $(\text{HF})_n$  form a linear chain and a cyclic ring,[33] and the hydrogen bonds in these clusters are very strong; the bonding energy per bond of the cyclic isomer is larger than the averaged bond energy of the same size of linear isomer.[22] Figure 7a and 7b show the basis set dependence of the hydrogen bonding energy per bond for linear  $(\text{HF})_2$  and  $(\text{HF})_3$ , and for cyclic  $(\text{HF})_4$  and  $(\text{HF})_5$ . The geometric parameters are from Kamiya *et al.*[33] Without the augment functions  $E_{BE}^{3\text{SPT}+\text{Disp}}$  are close to the CP corrected MP2. The plots for both the CP corrected and uncorrected MP2 display the irregular points at apvdz (aug-cc-pvdz). On the other hand, the plots of  $E_{BE}^{3\text{SPT}+\text{Disp}}$  are less basis set dependent for all clusters examined, including clusters other than in Figure 7. The figures, together with Figure 5c and 5c, suggest the possible practical use of the 3rd SPT + DPT(Dispersion only),  $E_{BE}^{3\text{SPT}+\text{Disp}}$ , with modest sizes of basis sets in the study of a series of large molecular clusters.

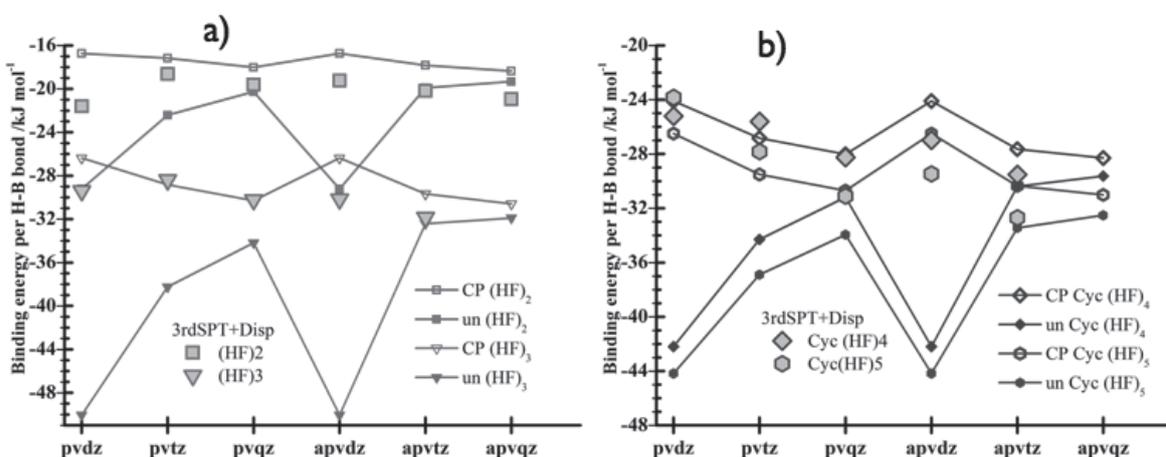


Figure 7. Basis set dependence of the hydrogen bond energy per bond in  $(\text{HF})_n$  for the CP corrected and uncorrected MP2, and the 3rd SPT+DPT(Disp). a) Linear  $(\text{HF})_2$  and  $(\text{HF})_3$ . b) cyclic  $(\text{HF})_4$  and  $(\text{HF})_5$ .

### 3.7 Ion - $\pi$ interaction

The atomic ions strongly interact with the surrounding neutral atoms and molecules in solids and in solution. Because of the coulombic nature, the ions influence the long-range interaction. In some of the ionic clusters and in the biological polymers, the specific local interaction of the atomic ions also play a key role in determining the structures and reactions. One of examples is the interaction with the conjugated  $\pi$  electrons of the aromatic molecules. Recently, Frontera and his coworkers [47] report the computational study for benzene -  $\text{M}^+$  ( $\text{M}=\text{Li}, \text{Na}, \text{K}$ ) and for *s*-triazine -  $\text{X}^-$  ( $\text{X}=\text{F}, \text{Cl}, \text{Br}$ ) with the MP2 level of theory. One of their finding is the erroneous behavior of basis set dependence of MP2(full) energy with aug-cc-pvXz (apvxz), which makes the extrapolation (CBS) energy improper. They demonstrate that if the core electrons are frozen in the MP2 calculations, the erroneous behavior disappears. Further they examined the basis sets, aug-cc-pCXz (apCXz), which contains the basis sets to correlate the core electrons, in the full MP2 calculations, and then, as expected, the smooth basis set dependence with and without counterpoise correction is found. Using their geometry optimized with the uncorrected MP2/apvdz, the LP MO 3rd SPT+2nd DPT(Disp) is applied to the ion -  $\pi$  interaction. Though in our calculations the intermolecular correlations are evaluated in the "full-electron" option, no erroneous behaviors are found in the basis set dependence of  $E_{BE}^{3\text{SPT}+\text{Disp}}$ . The energy at the 8th row in Table 2 is close to the energy of CBS(full) (apCXz[D-T]) which is evaluated at a slight different geometry. In the table, the calculated dispersion energy and the CT energy are also given. The CT energy is defined by the difference of  $E_{BE}^{3\text{SPT}} - E_{BE}^{\text{LPMO}}$  in the previous work.[22]

Table 2. Comparison of the binding energy ( $\text{kJ mol}^{-1}$ ) of benzene -  $\text{M}^+$  ( $\text{M}=\text{Li}, \text{Na}$ ) and triazine -  $\text{X}^-$  ( $\text{X}=\text{F}, \text{Cl}, \text{Br}$ )

molecule	Benzene	Benzene	Triazine	Triazine	Triazine
ion	$\text{Li}^+$	$\text{Na}^+$	$\text{F}^-$	+ $\text{Cl}^-$	+ $\text{Br}^-$
Exper	-160.25 <sup>e)</sup>	-92.47 <sup>f)</sup>			
CBS(full) apCXz [D-T] <sup>a,b)</sup>	-161.50	-104.35	-50.00	-36.23	
CBS(full) apvXz [D-T] <sup>a,c)</sup>	-178.41	-148.78	-55.77	-42.34	-48.49
CBS(Frozen) apvXz [D-T] <sup>a,c)</sup>	-155.52	-93.22	-50.79	-36.07	-37.74
$E_{BE}^{3SPT+Disp}(all)$ apvXz <sup>d,c)</sup>	-161.75	-103.65	-55.09	-30.97	-26.36
$\Delta E_{Disp}^{2LP-DPT}$ apvXz <sup>d,c)</sup>	-1.41	-1.60	-22.31	-23.95	-24.59
$\Delta E_{CT}^{3LP-SPT}$ apvXz <sup>d,c)</sup>	-8.31	-0.63	-3.70	-2.32	-2.37

a)[47], b)Geometry optimized with MP2/aug-cc-pCdz [47],c)Geometry optimized with MP2/aug-cc-pVdz [47], d)Present results. X=D for benzene- $\text{M}^+$  and X=T for triazine- $\text{X}^-$ , e)[48], f)[49]

In benzene -  $\text{M}^+$ , the dispersion energy is very small, and the CT from benzene to  $\text{Li}^+$  has a small contribution to the bonding, but it is a minor part of the total binding energy. In benzene -  $\text{M}^+$  the interaction is mostly the electronic polarization of the benzene molecule. On the other hand, in triazine -  $\text{X}^-$ , the dispersion energy is very large. Without the dispersion energy, the proper binding of a halogen anion with aromatic molecules cannot be described. This has an important implication for the computational study of biomolecular systems containing halogen anions.

## Further Discussion

In the above test calculation, the severe restriction on the configuration space in the 2nd order double excitations is imposed; only the dispersion type excitations, except for  $\text{He}_2$ , are included in the wave function. The intramolecular electronic correlation in these calculations is changed through molecular interaction. The occupied MOs are deformed through the polarization interaction, and the charge-transfer (or electron delocalization over the constituent molecules) changes the number of "effective" electrons in the constituent molecules). For more accurate computations, at least the change of the pair correlation through molecular interaction has to be evaluated. It is not a simple task to keep the consistency of the configuration space in the supermolecule type theory. Because the many-electron configurations are constructed from the molecular orbitals, the consistency of the configuration space of the cluster and its components is related to the consistency of the orbital space. In other words, CBI is coupled with OBI. Keeping in mind this difficulty, we have to look for an appropriate theoretical tool. The present rather simple procedure can be the first step.

## Acknowledgment

The work is partially supported by the Grants-in-Aid for Science Research (No.20550018) of JSPS. A part of computations was carried out at the RCCS, Okazaki Research Facilities, National Institutes of Natural Sciences (NINS). A technical help for a workstation by Mr. S. Shirai of Toyota Central R&D Laboratories, Inc. is acknowledged. We thank Prof. A. Frontera for sending us the geometric parameters.

## References

- [1] Hobza,P.; Müller-Dethlefs,K., "Non-covalent Interactions", RSC Publishing (2010)

- 
- [2] Liu,B.; McLean,A.D., *J. Chem. Phys.* 91 (1989) 2348
- [3] Nooijen, M.; Shamasundar,K.R.; Mukherjee,D., *Mol. Phys.* 103 (2005) 2278
- [4] Jansen,H.B; Ros,P., *Chem. Phys. Letters* 3 (1969) 140
- [5] Boys,S.F.; Bernardi,F., *Mol. Phys.* 19 (1970) 553
- [6] White,J.C.; Davidson,E.R., *J. Chem. Phys.* 93 (1990) 8029
- [7] Valiron,P.; Mayer,I., *Chem. Phys. Letters* 275 (1997) 46
- [8] Varandas,A.J.C., *J.Phys.Chem.A* 114 (2010) xxxxx
- [9] Jurečka,P.; Šponer, J.; Černý, J.; Hobza,P., *Phys.Chem.Chem.Phys.* 8 (2006) 1985
- [10] Pitoňák,M.; Janowski,T.; Neogrady,P.; Pulay,P.; Hobza,P., *J.Chem.Theor,Comp.* 5 (2009) 1761
- [11] Jeziorsky,B.; Szalewicz,K., *Encyclopedia Comput.Chem.* 2 (1998) 1376
- [12] Pulay,P., *Chem. Phys. Letters* 100 (1983) 151
- [13] Mata,R.A.; Werner,H.-J., *J. Chem. Phys.* 125 (2006) 184110
- [14] Nagata,T.; Takahashi,O.; Saito,K.; Iwata,S., *J. Chem. Phys.* 115 (2001) 3553
- [15] Nagata,T.; Iwata,S., *J. Chem. Phys.* 120 (2004) 3555
- [16] Iwata,S., *J. Theor. Comp. Chem.* 5 (2006) 819
- [17] Iwata,S.; Nagata,T., *Theor. Chem. Acc.* 116 (2006) 137
- [18] Iwata,S., *Chem.Phys.Letters*, 431 (2006) 204
- [19] Gianinetti,E.; Raimondi, M.; Tornaghi,E., *Int. J. Quantum Chem.* 60 (1996) 157
- [20] Gianinetti,E.; Vandoni,I.; Famulari,A.; Raimondi,M., *Adv. Quantum Chem.* 31 (1999) 251
- [21] Iwata,S., *J. Phys. Chem. B*, 112 (2008) 16104
- [22] Iwata,S., *J. Phys. Chem. A* 114 (2010) 8697
- [23] Stone,A.J., *Science* 321 (2008) 787
- [24] Podeszwa,R.; Szalewicz,K., *Chem. Phys.Letters* 412 (2005) 488
- [25] Kamiya,M.; Tsuneda,T.; Hirao,K., *J. Chem. Phys.* 117 (2002) 6010
- [26] Sherrill,C.D., *J.Chem.Phys.* 132 (2010) 110902
- [27] Khaliullin,R.Z.; Head-Gordon,M.; Bell,A.T. *J.Chem.Phys.* 124 (2006) 204105
- [28] Khaliullin,R.Z.;Cobar,E.A.; Lochan,R.C.; Bell,A.T.; Head-Gordon,M., *J. Phys. Chem.* 111 (2007) 8753
- [29] Grimme,S., *J. Chem. Phys.* 118 (2003) 9095
- [30] 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.J.; Windus,T.L.; Dupuis,M; Montgomery,J.A., *J.Comput.Chem.* 14 (1993) 1347

- [31] M. J. Frisch *et al.* GAUSSIAN 03, Gaussiab Inc., Wallingford, CT (2004)
- [32] Valiron,P.; Mayer,I., *Chem.Phys.Letters* 275 (1997) 46
- [33] Kamiya,M.; Hirata,S.; Valiev,M., *J. Chem. Phys.* 128 (2008) 074103
- [34] Wells,B.H.; Wilson, S., *Chem. Phys. Letters*, 101 (1983) 429
- [35] Salvador,P.; Szcześniak,M.M., *J. Chem. Phys.* 118 (2003) 537
- [36] Paizs,B.; Salvador,P.; Császár,A.G.; Duran,M.; Suhai,S., *J. Comp. Chem.* 22 (2001) 196
- [37] Mayer,I.; Valiron,P., *J. Chem. Phys.* 109 (1998) 3360
- [38] Anderson,J.B.; Traynor,C.A.; Boghsian,B.M., *J. Chem. Phys.* 99 (1993) 345
- [39] Cencek,W.; Szalewicz,K., *Int. J. Quantum Chem.* 108 (2008) 2191
- [40] Halkier,A.; Helgaker,T.; Jorgensen,P.; Klopper,W.; Koch,H.; Olsen,J.; Wilson,A. K., *Chem. Phys. Letters* 286 (1998) 243
- [41] Bytautas,L.; Ruedenberg,K., *J. Chem. Phys.* 128 (2008) 214308
- [42] Halkier,A.; Klopper,W.; Helgaker,T.; Jorgensen,P.; Taylor,P.R., *J. Chem. Phys.* 111 (1999) 9157
- [43] Fernández,B.; Koch,H., *J.Chem. Phys.* 109 (1998) 10255
- [44] Patkowski,K.; Murdachaew,G.; CHENG-MING Fou,C.-M.; Szalewicz,K., *Mol. Phys.* 103 (2005) 2031
- [45] Patkowski,K.; Szalewicz,K., *J.Chem. Phys.* 133 (2010) 094304
- [46] Noro, T.; Sekiya, M.; Osanai, Y.; Miyoshi, E.; Koga, T.; Tatewaki, H., "Segmented Gaussian Basis Set", <http://setani.sci.hokudai.ac.jp/sapporo/Welcome.do>
- [47] Estarellas,C.; Lucas,X.; Frontera,A.; Quiñonero,D.; Deyà,P.M., *Chem. Phys. Letters* 489 (2010) 254
- [48] Taft,R.W. et al, *Pure Appl. Chem.* 62 (1990) 17
- [49] Amicangelo,J.C.; Armentout,P.B., *J. Phys. Chem. A* 104 (2000) 11420