Correction of the basis set superposition error in SCF and MP2 interaction energies. The water dimer

M. M. Szczęśniak and Steve Scheiner[a)]
Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois 62901

There has been some discussion concerning whether basis set superposition error is more correctly evaluated using the full set of ghost orbitals of the partner molecule or some subset thereof. A formal treatment is presented, arguing that the full set is required at the Møller–Plesset level. Numerical support for this position is provided by calculation of the interaction energy between a pair of water molecules, using a series of moderate sized basis sets ranging from 6-31G** to the [432/21] contraction suggested by Clementi and Habitz. These energies, at both the SCF and MP2 levels, behave erratically with respect to changes in details of the basis set, e.g., H p-function exponent. On the other hand, after counterpoise correction using the full set of partner ghost orbitals, the interaction energies are rather insensitive to basis set and behave in a manner consistent with calculated monomer properties. For long intersystem separations, the contribution of correlation to the interaction is repulsive despite the attractive influence of dispersion. This effect is attributed to partial account of intrasystem correlation and can be approximated at long distances via electrostatic terms linear in MP2-induced changes in the monomer moments.

I. INTRODUCTION

Early *ab initio* studies of molecular interactions via the supermolecule treatment were plagued with a number of difficulties such as limitation to small basis sets and inability to adequately include electron correlation.[1–3] The partial successes of these early calculations were therefore due in large measure to a fortuitous cancellation between the various sources of error. Recent advances in computational techniques have made it possible to address some of the earlier problems and drastically lower their associated errors. For example, it is now possible to apply rather extended basis sets and thereby adequately describe the properties of the monomers.[4–6] Nevertheless, one problem which has persisted and which remains a source of significant uncertainty is basis set superposition error (BSSE). The mutual enlargement of the basis set of each monomer by the presence of the orbitals of its partner molecule results in an artificial energy lowering by the variation principle.

The magnitude of the BSSE at the SCF level has been the subject of a number of previous studies[7–13] and it is now generally agreed that this error can be brought down to almost negligible proportions by the use of very long atomic orbital expansions.[13] However, basis sets of this size can seldom be applied to systems of real chemical interest and it is therefore essential to be able to correct the errors that will inevitably occur with sets of modest size. The question of superposition errors at post-SCF levels is a much newer matter and one that has received only very limited attention to this point.[13–21] This neglect is due primarily to past scarcity of applications of correlated procedures to molecular interactions rather than to any supposition of small magnitude of this error. Indeed, recent work has demonstrated that the BSSE at the post-SCF level is comparable to its SCF analog and can, in fact, be even larger than the true correlation

contribution to the interaction itself.[13,18,21] Moreover, the correlated BSSE appears to remain quite large even in the face of very extended basis sets,[13] adding import to identification of a procedure which will accurately correct the associated error.

It is the purpose of the present work to investigate appropriate means of dealing with superposition error at the SCF and, in particular, the correlated level. We begin in the next section with a formal analysis of the source of this error and what we believe to be the most correct procedure for eliminating it. We then provide computational data to support our claims, focusing our efforts on the water dimer for which there exists the largest body of data. In particular, we examine a number of basis sets, all of moderate size, but which differ from one another in only minor respects. This approach has two advantages. First, it is possible to associate each change in basis set with a particular alteration of results and thereby obviate the complex interplay of trends which makes other types of comparisons difficult to interpret. Secondly, the similarity of basis sets implies that the results, after correction for BSSE, should approximately converge with one another, providing a numerical criterion for the veracity of our approach.

II. THEORETICAL FRAMEWORK

In response to previous calculations which had indicated a clearly spurious attractive force between rare gas atoms at the Hartree–Fock level,[22] Boys and Bernardi first proposed elimination of basis set superposition error by what they termed the functional counterpoise procedure.[23] In this approach, the energy of each monomer is calculated within the basis set of the entire dimer, including the "ghost orbitals" of the partner molecule. Soon thereafter, Johansson *et al.*[24] applied this technique to various dimers but found what appeared to be an overcorrection, i.e., after subtraction of

---

the counterpoise term, the interactions became much less attractive than the experimental cases. These results initiated a number of suggestions over the years that the counterpoise corrections represent an overestimate of the artificial lowering of the monomer energies due to superposition of orbitals.[25] For example, the Pauli exclusion principle was invoked in an explanation as to why the occupied orbitals of monomer B are not available to A and vice versa; hence only the virtual orbitals of each monomer should be used when calculating the counterpoise correction.[25] We discuss this issue first with regard to the SCF level before moving on to address the question of counterpoise corrections at correlated levels.

A. SCF level

Most of the arguments supporting the contention that only virtual ghost orbitals should be included in the SCF counterpoise correction rest largely on numerical comparison of data corrected by either the full (all orbitals) or partial (virtuals only) procedure. However, since the calculations were carried out primarily with small basis sets (e.g., STO-3G), ill equipped by their very nature to handle molecular interactions, these comparisons are of dubious validity. For example, as mentioned above, the very weak H-bonding interactions between monomers resulting from full BSSE correction of the STO-3G potentials provided the initial impetus to question the counterpoise technique.[24] However, a weakly attractive H-bond potential is *not* unreasonable for this interaction at the STO-3G level since repulsive forces are much better represented with this basis set than any attractive contributions, especially the first-order exchange energy calculated with a dimer-centered basis set.[26] The inappropriateness of basing BSSE arguments on STO-3G results is underscored by the work of Kolos who was able to obtain quite reasonable attractive potentials using the *full* counterpoise correction within the framework of a minimal basis set, provided that the latter is prepared via atomic SCF calculations rather than a least-squares fit to STOs.[11] Other investigators have further confirmed the necessity to include the full counterpoise correction in connection with much larger basis sets.[10]

From a more rigorous point of view, it has been demonstrated recently that the occupied orbitals of the partner subunit, as well as their vacant counterparts, are required to minimize the so-called zero-exchange term in the Heitler–London energy.[10,26] This artifact represents the effect of incompleteness of the basis set of subsystems and enters into the first-order energy if not dealt with. In addition, Groen and van Duijneveldt have shown that the energy lowering of the proton acceptor monomer has an improper R dependence if only the virtual orbitals of the donor are included in the counterpoise correction.[10] Gutowski *et al.*[27] have advanced particularly convincing arguments which we paraphrase here concerning the necessity to include all orbitals. Let us first calculate the SCF energy of a complex AB in the complete basis set centered midway between A and B. The occupied orbitals of A are of course not available for occupancy by the electrons of B (and vice versa) by the Pauli

principle. However, in calculating the energies of monomers, it would clearly not be appropriate to exclude the occupied orbitals of the partners, since one would then spoil the completeness of the basis set. Gutowski *et al.*[27] clarify the role played by the Pauli principle in that it certainly contributes to the interaction energy via the presence of exchange terms but has nothing to do with the mutual improvement of the monomer basis sets in a supermolecule calculation. Exchange terms, which represent the energetic consequence of inaccessibility of certain regions of space to some electrons, will be better represented if these regions are accounted for by the use of dimer-centered basis sets.[26]

B. Correlated level

It would seem natural to apply the arguments put forth in Ref. 27 to any variational treatment involving correlation effects. In fact, Gutowski *et al.*[20] formally proved that the entire set of ghost orbitals is required for proper evaluation of interaction energies evaluated by the supermolecular CEPA(1) treatment. However, since the need to include the full basis set may not be so obvious with perturbational approaches such as the Møller–Plesset technique,[28] we include the following.

The correlation contribution to the interaction energy can be calculated as the difference between MP2 correlation energies of a dimer AB and the monomers:

$$E_{\text{int}}(\text{corr}) = E_{\text{AB}}^{\text{MP2}} - (E_{\text{A}}^{\text{MP2}} + E_{\text{B}}^{\text{MP2}}). \qquad (1)$$

Each term on the right-hand side of Eq. (1) may be expressed as a minimum of a Hylleraas-type functional[29]:

$$J[\phi] = \langle \phi | H_0 - E_0 | \phi \rangle + 2 \langle \psi^0 | W - E^{(1)} | \phi \rangle, \qquad (2)$$

where H_0 and W refer to a standard Møller–Plesset partitioning of the total Hamiltonian, ψ^0 describes the unperturbed state, and ϕ is a trial function. If ϕ happens to be equal to the first-order correction to the wave function, $J[\phi]$ becomes E^{MP2}, the second-order perturbed energy in MP theory. (An analogous treatment can be applied to any nth order of MP theory.)

Returning now to the arguments advanced by Gutowski *et al.*,[27] suppose that the calculations of the dimer and monomers are performed in a complete basis set, centered in the middle of the A–B bond. After optimization of the trial function ϕ, J represents the exact second-order MP energy in the basis set limit for each of these species. If, however, some orbitals are excluded in the monomer calculations (e.g., those occupied in the dimer), J can no longer refer to the basis set limit and must in fact be above E^{MP2} by the variation principle. In such a scheme, the supermolecule calculations are basis set inconsistent. Hence, contrary to previous suggestions,[17,19] a basis set consistent treatment of interaction energy at correlated levels must involve the full set of the partner molecule's ghost orbitals, as first applied by Newton and Kestner.[18]

Although the reasoning above applies to complete basis sets, Gutowski *et al.*[20] have demonstrated the validity of this approach for the opposite extreme of minimal basis sets (with no virtual space). There is no reason to believe that truncated basis sets of intermediate size, as generally applied

to molecular interactions, would behave any differently.

In addition to the primary BSSE described above and hopefully corrected by the counterpoise procedure, the interaction energies are subject to a more subtle superposition effect as well. As pointed out earlier by Karlström and Sadlej,[30] secondary effects are caused when the one-electron properties of each monomer are affected by the partner's ghost orbitals. This secondary effect is not removed by the counterpoise procedure. For example, the electrostatic interaction energy between two spherically symmetric closed-shell atoms is nonzero in the dimer basis set due to the spurious moments induced in each atom by the orbitals of its partner. Fowler and Buckingham[31] have recently suggested reevaluation of the multipole energies in terms of the dimer-centered moments and polarizabilities but their solution addresses only part of the problem because changes in penetration terms are ignored. The importance of the latter terms is underscored by the work of Gutowski et al.[27]

III. COMPUTATIONAL DETAILS

Interaction energies were calculated using the *ab initio* GAUSSIAN-80 package of computer codes.[32] MONSTER-GAUSS[33] was used to evaluate molecular polarizabilities. The basis set superposition error (BSSE) was computed using the counterpoise procedure of Boys and Bernardi.[23] The coordinates of the atoms were taken from the experimental geometry of the water dimer[34] and the interoxygen distance R varied. Specifically, $r(OH) = 0.957$ Å and $\theta(HOH) = 104.5°$ for each subunit. The O–H bond of the proton donor molecule was taken as coincident with the O–O axis while the angle between the latter axis and the HOH bisector of the proton acceptor was taken as 120°. The entire complex belongs to the C_s point group and the proton of the donor group not participating in the H bond is *trans* to the acceptor bisector.

Basis sets examined were basically of polarized double-ζ type. The standard 6-31G** set[35] is denoted as A_1. The B basis sets contain a second set of d functions on O and are hence termed 6-31G($2d,p$). The exponents of the two sets of d functions were taken as 1.0 and 0.15, as is true also for the C and D basis sets below. The description of hydrogen is altered from four primitives in the B sets to six in the C basis

sets. The first s function is expressed as a contraction of four primitives and the second by two.[36] Our last basis set studied, D, is identical to C for hydrogen but differs from the previous sets by a [432] contraction of the O basis, as used previously by Clementi and Habitz.[37] In order to determine the effect on the results of varying degrees of diffuseness of the polarization functions, a range of different values were used for the p-function exponents on hydrogen. Subscript 1 denotes the basis with the largest exponent (1.1), 2 is associated with exponent 0.75, 3 with 0.4, and 4 with 0.15.

IV. RESULTS
A. Uncorrected potentials

Potential energy curves were calculated as a function of R, the interoxygen distance, for each basis set. All curves contain a minimum in the neighborhood of 3.0 Å at the SCF level; this minimum shifts to smaller R (by about 0.1 Å) when correlation effects are included via MP2. The interaction energies are presented in Table I for three key regions of the potential. $R = 3.0$ Å represents the approximate position of the minimum, 2.5 Å is on the repulsive part of the curve, and 4.5 Å lies in the intermediate region where the interaction energy is still sizable and where exchange effects are negligible.

It may be seen in Table I that the SCF interaction energies at $R = 3.0$ Å range from 3.8 kcal/mol with the D_3 basis set up to 6.1 with B_1. This energy is quite sensitive to specific details of the basis set. For example, lowering the p-orbital exponent of H from 1.1 to 0.15 within the context of the B_n basis sets reduces the energy from 6.1 to 4.9 kcal/mol. The short-range interaction (2.5 Å) is affected even more strongly, varying by several kcal/mol for the different basis sets. Another interesting feature is that the interaction potentials tend to cross one another in the 4–5 Å range. For example, although the B_1 curve is substantially more attractive than B_4 at 3 Å, the opposite is true when the two water molecules are 4.5 Å distant. Since the H bond is composed largely of an electrostatic attraction, it would be natural to expect some correlation between the dipole moment of the water monomer and the computed interaction energy either at the vdW minimum or at long range. However, comparison of the data

TABLE I. Interaction energies (kcal/mol) calculated with various basis sets.

ζ_p(H)	A_1 1.1	B_1 1.1	B_2 0.75	B_3 0.40	B_4 0.15	C_3 0.40	C_4 0.15	D_3 0.40
R(Å)				E_{int} (SCF)				
2.5	+0.59	−0.31	−0.08	+0.28	+1.45	+0.31	+1.42	+2.36
3.0	−5.58	−6.09	−5.97	−5.71	−4.86	−5.52	−4.91	−3.79
4.5	−1.62	−1.48	−1.47	−1.44	−1.77	−1.43	−1.77	−1.30
				E_{int} (corr)				
2.5	−2.55	−4.24	−4.60	−4.29	−2.55	−4.72	−2.56	−3.70
3.0	−1.19	−2.35	−2.42	−2.60	−1.58	−2.42	−1.48	−1.68
4.5	−0.03	−0.07	−0.07	−0.08	−0.28	−0.09	−0.24	−0.16
				E_{int} (MP2)				
2.5	−1.96	−4.56	−4.68	−4.58	−1.10	−4.41	−1.14	−1.35
3.0	−6.77	−8.44	−8.39	−8.31	−6.44	−7.94	−6.39	−5.48
4.5	−1.64	−1.55	−1.54	−1.52	−2.04	−1.52	−2.01	−1.46

TABLE II. Dipole moments[a] (D) and correlation corrections to the interaction energy (kcal/mol) for $R = 6$ Å.

ζ_p (H)	A_1 1.1	B_1 1.1	B_2 0.75	B_3 0.40	B_4 0.15	C_3 0.40	C_4 0.15	D_3 0.40
μ^{SCF}	2.185	2.007	1.998	1.984	2.006	1.968	2.009	1.998
μ^{MP2}	2.10	1.88	1.86	1.85	1.89	1.83	1.88	1.87
$\Delta\mu^{b}$	-0.09	-0.13	-0.13	-0.13	-0.12	-0.14	-0.13	-0.13
$E_{ES}^{(12)c} + E_{disp}{}^{d}$	0.020	0.031	0.033	0.033	0.026	0.033	0.030	0.021
E_{int} (corr)[e]	0.031	0.040	0.050	0.043	0.029	0.043	0.033	0.021

[a] Derived within context of dimer basis set.
[b] $\mu^{MP2} - \mu^{SCF}$.
[c] From Eq. (2).
[d] $E_{disp}(6) = -0.0129$ kcal/mol: C_6/R^6 term, C_6 from Ref. 29.
[e] Counterpoise corrected.

in Table I with the SCF dipole moments listed in the first row of Table II indicates no such correlation.

Correlation contributions to the interaction energies are listed as E_{int} (corr) in Table I. We again note a high sensitivity to basis set and exponent choice. The diffuse H p orbital with exponent 0.15 (B_4 and C_4) leads to particularly small correlation interaction energies in the short range, being roughly half that obtained with slightly larger exponents; the same trend is seen for 3 Å. This is particularly surprising as the mean polarizabilities of the water monomer calculated with the latter sets are *larger* than those evaluated for the basis sets containing less diffuse H p orbitals. Assuming for the moment that E_{int} (corr) consists entirely of UCHF dispersion, the opposite trend would be expected. On the other hand, the diffuse p orbitals yield values of E_{int} (corr) at $R = 4.5$ Å several times higher than the other exponents.

The sums of the SCF interactions and E_{int} (corr) contributions, representing the total interaction energies at the MP2 level, are listed in the last section of Table I as E_{int} (MP2). Again, we see a high sensitivity of results to basis set, with this quantity varying between -5.5 and -8.4 kcal/mol at the minimum. It is particularly disturbing that this energy is grossly exaggerated, even by double-ζ basis sets with two sets of d functions on oxygen, when compared to the experimental value of -5.4 ± 0.5 kcal/mol.[38] A second point of concern is that the interactions remain attractive, i.e., negative E_{int}, even at the very short intermolecular distance of 2.5 Å. This short-range potential is fairly insensitive to H p-orbital exponent in the range 0.4–1.1 (B_1–B_3); however, lowering the exponent to 0.15 (B_4) diminishes the interaction energy precipitously by some 3 kcal/mol at $R = 2.5$ Å. The MP2 potentials tend to cross one another in the 4–5 Å range as did the SCF curves.

In summary, the results are extremely sensitive to basis set, behaving in an erratic fashion, especially considering the minor perturbations introduced by altering the exponent chosen for the H polarization functions. On the other hand, the results for the B_n and C_n basis sets are quite similar for each value of n, indicating the two descriptions of the H valence orbitals are pretty much equivalent. The [432] contraction for O leads to a less attractive interaction.

B. Corrected potentials

The computed counterpoise corrections (CC) for each basis set are listed in Table III at both the SCF and correlated levels. Let us concentrate first on the SCF values at 3 Å, the approximate position of the minimum, in the second row of the table. The correction is 1.0 kcal/mol for the A_1 basis set, but more than double this value for B_1–B_3, indicating enlargement of basis set is no insurance of BSSE reduction. The more diffuse p function used in B_4 leads to a much smaller CC; similar behavior is exhibited by the C basis sets. Note the very small value of the CC computed with the D_3 basis set. Very similar observations apply to the short-range (2.5 Å) corrections in the preceding row. The situation is somewhat different in the intermediate range (4.5 Å) where the diffuse p functions of B_4 and C_4 lead to *higher* corrections than do the other basis sets. Note that the D_3 SCF correction at this distance is quite negligible.

The second-order counterpoise corrections to E_{int} (corr), contained in the second half of Table III, exhibit some very interesting trends. For R equal to both 2.5 and 3.0 Å, the values CC^{corr} are surprisingly similar in magnitude to their SCF counterparts with one notable exception. In contrast to its very small SCF CC, the D_3 basis set leads to substantial second-order counterpoise corrections, comparable to the CC^{corr} values of the other basis sets and several times higher than the corresponding D_3 SCF corrections. As in the SCF case, CC^{corr} dies off more slowly for the B_4 and C_4 basis sets than for those with less diffuse p functions. In all cases

TABLE III. Counterpoise corrections (kcal/mol).

ζ_p (H)	A_1 1.1	B_1 1.1	B_2 0.75	B_3 0.40	B_4 0.15	C_3 0.40	C_4 0.15	D_3 0.40
R(Å)				$-CC^{SCF}$				
2.5	1.56	3.32	3.31	3.25	1.58	3.15	1.58	0.37
3.0	0.98	2.32	2.36	2.27	1.08	2.07	0.99	0.16
4.5	0.20	0.24	0.24	0.23	0.48	0.23	0.47	0.02
				$-CC^{corr}$				
2.5	1.65	3.33	3.52	3.69	1.69	3.44	1.68	2.28
3.0	0.87	2.05	2.08	2.15	1.11	1.92	1.03	1.09
4.5	0.07	0.13	0.13	0.12	0.27	0.15	0.25	0.18

(except perhaps D_3) the second-order correction approaches zero more quickly with increasing R than does CC^{SCF}.

In summary, second-order counterpoise corrections to the interaction energy are comparable to, and in some cases larger than, the SCF values. While the basis sets with diffuse p functions on H are associated with smaller corrections at short range, these corrections fall off more slowly with R and are hence appreciable even at fairly long intermolecular distances. In contrast to the quite small corrections of the D_3 basis set at the SCF level, the corresponding second-order corrections are no smaller than those computed for the other basis sets and, in fact, larger than those of the much smaller A_1 set.

Let us now focus our attention on the interaction potentials after the counterpoise corrections of Table III have been removed. The corrected interaction energies are displayed in Table IV for $R = 2.5$, 3.0, and 4.5 Å. The corrections have two major effects on the data of Table I: The sensitivity to basis set detail is drastically reduced and the potentials are markedly less attractive. For example, the SCF interaction energies at 3 Å for the group of B basis sets extend only over the narrow range between 3.4 and 3.8 kcal/mol, in contrast to the 1.2 kcal range in the uncorrected data. Moreover, the corrected interaction energies correlate in a linear fashion with the monomer dipole moments (see Table II), conforming to the known dominating role of electrostatics in H bonding. The high A_1 interaction energy in Table IV may thus be traced directly to the overestimated dipole moment. This predictable behavior contrasts with the erratic dependence of the uncorrected energies upon electric moments of the monomers. To reemphasize the uniformity of the corrected data, if the data for the A_1 basis set with its high monomer moment are excluded, E_{int}(SCF)–CC^{SCF} evaluated at 3 Å covers a range of only 0.4 kcal/mol, as compared to 2.3 for the uncorrected energies. This uniformity extends as well to other regions of the potential. The short-range corrected energies (2.5 Å) are all repulsive and rather close to one another while the variation among the long-range (4.5 Å) attractions has been greatly reduced as well. The crossing of the SCF potentials in the 4–5 Å range mentioned earlier is now avoided entirely or pushed back to much longer distances.

After counterpoise correction, the second-order contributions to the interaction energies, E_{int}(corr), are substantially reduced. For example, at the approximate minima of the potentials, this term is now in the neighborhood of -0.5 kcal/mol, as opposed to the uncorrected values in excess of -2. Note again the insensitivity to basis set of the corrected energies at 3 Å as well as at the other distances. The counterpoise correction has shifted to smaller R the zero of E_{int}(corr), repulsive at long distances.

Like the SCF results, the corrected MP2 potentials in the last section of Table IV are repulsive at 2.5 Å. The high level of agreement among the B and C basis sets is particularly striking. The values of the total interaction energy at 3 Å are also quite insensitive to basis set, lying in the range of 3.9 to 4.4 kcal/mol. The notably higher value obtained with the A_1 basis set may be traced back to the high SCF interaction energy caused primarily by the exaggerated monomer dipole moment. The various basis sets converge nicely towards one another in the intermediate region of 4.5 Å.

In an attempt to efficiently simulate the contribution of the vacant ghost orbitals to the counterpoise correction, Schwenke and Truhlar[39] used only the polarization functions of the ghost molecule, their choice being motivated in part by the high participation of these orbitals in the vacant MOs. These workers found that such substitution did not significantly improve their equilibrium results at the SCF level. We have tested this idea at the correlated level here and the results are collected in Table V. Comparison of the polarization second-order counterpoise corrections (pol CC^{corr} in Table V) with the full CC^{corr} in Table III reveals that the polarization functions allow one to obtain only a small fraction of the full correction. Perhaps more important, the trends in Table III are not reproduced by the pol CC^{corr} data. For example, the sharp reduction in the B_4 CC^{corr} relative to the other B_n values at $R = 2.5$ and 3.0 Å is absent; a similar observation applies to C_3 and C_4. Moreover, the CC of the A_1 basis set is grossly underestimated. Although the pol CC^{corr} fails in the short and equilibrium regions, it does appear to more correctly reproduce at least the qualitative trends in the intermediate region around 4.5 Å even if still much underestimated.

As an alternative, we examined the counterpoise correction arising from use of the nonpolarization functions,

TABLE IV. Interaction energies (kcal/mol) corrected by counterpoise procedure.

ζ_p(H)	A_1 1.1	B_1 1.1	B_2 0.75	B_3 0.40	B_4 0.15	C_3 0.40	C_4 0.15	D_3 0.40
R(Å)				E_{int}(SCF)–CC^{SCF}				
2.5	+2.15	+3.01	+3.23	+3.53	+3.03	+3.46	+3.00	+2.73
3.0	−4.60	−3.77	−3.61	−3.44	−3.78	−3.44	−3.92	−3.64
4.5	−1.42	−1.24	−1.29	−1.22	−1.29	−1.20	−1.31	−1.27
				E_{int}(corr)–CC^{corr}				
2.5	−0.90	−0.92	−1.07	−1.17	−0.86	−1.28	−0.88	−1.42
3.0	−0.32	−0.30	−0.33	−0.45	−0.47	−0.50	−0.45	−0.59
4.5	+0.05	+0.06	+0.13	+0.04	−0.00	+0.03	+0.01	+0.12
				E_{int}(MP2)–CC^{MP2}				
2.5	+1.26	+2.09	+2.16	+2.36	+2.17	+2.18	+2.12	+1.31
3.0	−4.92	−4.07	−3.94	−3.89	−4.25	−3.94	−4.37	−4.22
4.5	−1.37	−1.18	−1.16	−1.17	−1.30	−1.14	−1.30	−1.26

TABLE V. Counterpoise corrections (kcal/mol) computed using subsets of full set of ghost orbitals.

ζ_p (H)	A_1 1.1	B_1 1.1	B_2 0.75	B_3 0.40	B_4 0.15	C_3 0.40	C_4 0.15	D_3 0.40
R(Å)				$-$ pol CC^{corr}				
2.5	0.35	1.17	1.27	1.48	1.04	1.55	1.03	1.19
3.0	0.07	0.70	0.71	0.91	0.75	0.94	0.74	0.54
4.5	0.00	0.05	0.04	0.04	0.18	0.05	0.17	0.10
				$-$ atomic CC^{corr}				
2.5	1.37	2.39	2.44	2.37	0.67	1.91	0.66	0.88
3.0	0.81	1.62	1.62	1.56	0.43	1.27	0.38	0.65
4.5	0.06	0.09	0.09	0.08	0.07	0.06	0.05	0.10

termed "atomic" by Schwenke and Truhlar.[39] It should be noted in this regard that Groen and van Duijneveldt have recently shown that the intersystem overlap integral of (HF)$_2$ in its equilibrium geometry is well represented by a basis set which contains no polarization functions but is instead saturated with sp functions.[10] We therefore calculated CC^{corr} using the unpolarized segments of our basis sets and the results are presented in the second half of Table V. While these corrections are substantially smaller than the full CC^{corr} in Table III, they are more successful at reproducing the short-range trends than are the contributions from the polarization functions; on the other hand, the intermediate-range behavior of the BSSE is misrepresented. In summary, the borrowing from the polarization subset of partner orbitals is largely responsible for intermediate range superposition effects whereas the behavior in the short range may be better attributed to the atomic or nonpolarization functions.

C. Long-range behavior

For interoxygen distances longer than 4.5 Å, the SCF and MP2 potentials remain attractive and approach zero asymptotically; however, the second-order contribution to the interaction energy E_{int} (corr) becomes positive for all basis sets. Although these values are rather small, ranging between 0.02 and 0.04 kcal/mol at $R = 6$ Å, their sign does warrant some investigation.

At these long distances, the only contributions to the interaction which are nonnegligible are electrostatic and dispersion (the induction term is vanishingly small[29]), both of which should be attractive in this case. The former force is expected to be the predominant component of the SCF interaction at 6 Å while dispersion will be contained within E_{int} (corr). These two terms were computed independently via long-range perturbation theory and the multipole expansion. The electrostatic energy, evaluated up through R^{-5} in terms of calculated moments, is equal to -0.45 kcal/mol at $R = 6$ Å. This value agrees quite well with the SCF data which range between -0.44 and -0.47 [with the exception of basis set A_1 for which E_{int} (SCF) $= -0.52$ kcal/mol]. The long-range dispersion energy, evaluated through the R^{-6} term,[29] is equal to -0.013 kcal/mol. Why then is E_{int} (corr), containing the dispersion interaction, positive?

The answer lies in the content of the correlation interaction energy in the second-order MP energy[40]

$$E_{int} (corr) = N_0 \langle A'B_0 + A_0B'|V\mathscr{A}|A'B_0 + A_0B' \rangle$$
$$+ E_{disp}^{(20)} + E_{exch-disp}^{(20)}, \qquad (3)$$

where $\mathscr{A}$ represents the intermolecular antisymmetrizer and V denotes the intermolecular interaction operator. A_0 and B_0 refer to the Hartree–Fock wave functions of the interacting subsystems A and B whereas the prime denotes first-order corrections due to the intrasystem correlation potential. In the language of symmetry-adapted perturbation theory,[41] the first term describes the correction to the first-order interaction energy (electrostatic + exchange) due to the second-order intrasystem correlation energy; it is hence denoted as $E^{(12)}$. The second and third terms, respectively, represent the (UCHF) dispersion interaction between uncorrelated fragments and its exchange counterpart. If we use the multipole expansion for V and neglect exchange effects, the leading (dipole–dipole) term in the expansion for $E^{(12)}$ becomes for a homodimer:[40]

$$E_{ES}^{(12)} \approx -(2/R^3)\{ \mu_x^{SCF} \cdot \Delta\mu_x^{corr} + \mu_y^{SCF} \cdot \Delta\mu_y^{corr}$$
$$+ 2\mu_z^{SCF} \cdot \Delta\mu_z^{corr} \}, \qquad (4)$$

where $\Delta\mu^{corr}$ represents the MP2 correlation correction to the SCF dipole moment μ^{SCF}; the z axis is defined to be collinear with the line connecting the subunit centers of interaction. Since MP2 generally lowers the dipole moment of a given molecule, $\Delta\mu^{corr}$ is negative and the full expression of Eq. (4) is positive, i.e., repulsive. [Equation (4) represents only the partial influence of intrasystem correlation upon first-order energies; terms quadratic in $\Delta\mu^{corr}$ will appear in higher orders of MP theory but are expected to be small.[40]]

Let us now test the above analysis by incorporating our data directly into the above expressions. Table II reports the SCF and MP2 dipole moments calculated for each basis set, along with the correction induced by correlation. The next row contains $E_{ES}^{(12)}$ evaluated by Eq. (4) for $R = 6$ Å, plus the dispersion energy calculated through the R^{-6} term. The resulting quantity represents an estimate of the total effect of correlation upon the interaction energy which may be seen to compare quite favorably with the values of E_{int} (corr) calculated quantum mechanically and listed in the last row of the table. We thus conclude that the correlation-induced reduction of the monomer dipole moments is primarily responsible for the repulsive long-range interaction correlation energies. This effect would be missed entirely by approaches which evaluate the total interaction energy as a sum of the SCF potential and a term corresponding to dispersion. For example, it is well known that uncorrelated SCF theory predicts the wrong sign of the dipole moment of CO. Consequently, CO--HF is predicted to be more stable than the experimentally observed OC--HF at the SCF level. Any treatment involving only the addition of dispersion to E_{int} (SCF) would probably duplicate this error. On the other hand, due to partial account of intrasystem correlation, MP2 theory correctly predicts OC--HF as the more stable configuration.[42]

V. SUMMARY AND DISCUSSION

As pointed out in Sec. II B, a basis set consistent treatment of interaction energy at the MP level should involve the full set of partner orbitals. Whereas the uncorrected potentials are extremely sensitive to small changes in the basis sets, subtracting the full Boys and Bernardi counterpoise corrections at both the SCF and MP2 levels leads to much greater uniformity of results. Moreover, the corrected potentials correlate well with calculated monomer properties while the behavior of the uncorrected data is much more erratic. As noted previously, increasing the size of the basis set does not necessarily lead to a reduction in the counterpoise correction at either the SCF or correlated levels; indeed, the opposite trend may be observed.

The magnitude of the counterpoise correction is quite sensitive to the exponents of the polarization functions. Large exponents of the H p functions, which are known to maximize intramolecular correlation energy, lead to sizable superposition error in the van der Waals minimum of the dimer as well as shorter distances. On the other hand, the largest counterpoise errors in the intermediate region are associated with more diffuse functions which are better suited for describing dispersion effects.

With regard to attempts to partition the total BSSE into contributions from polarization and atomic orbitals, neither subset is capable of correctly reproducing the trends over the entire range of intermolecular distance. Hence, complete saturation of the nonpolarization or "radial" part of the wave function, e.g., by long systematic sequences of even-tempered basis sets, will not prevent a substantial BSSE from occurring when polarization functions are later added.

In addition to UCHF dispersion and exchange–dispersion interactions, the correlation interaction energy within second-order MP theory contains a correlation correction to the first-order interaction energy (second order with respect to intrasystem correlation).[40] In the long intersystem regime, exchange effects are negligible and only the electrostatic and dispersion forces remain. The effect of intrasystem correlation upon the former may be estimated reasonably accurately by the multipole expansion incorporating the MP2 correction to the monomer dipole moment. For those systems in which this moment is reduced by correlation, the resulting repulsive contribution to the electrostatic energy is likely to overcome the attractive force of dispersion; hence, the net effect of correlation may be repulsive at long distances, as has in fact been observed here for $(H_2O)_2$ and for $(HF)_2$ in Ref. 40. Of course, the latter principle would be obscured and the interaction appear attractive if the basis set superposition error is not removed from E_{int} (corr). Since the correlation interaction energy is composed of two primary factors, each with a very different dependence on polarization function exponents, optimization of the latter with respect to E_{int} (corr) would not generally maximize the dispersion energy. In fact, such a procedure involving uncorrected E_{int} (corr) would effectively maximize the BSSE instead.

The current experimental estimate of the interaction energy of the water dimer is -5.4 ± 0.5 kcal/mol.[38] It is likely that this value is overestimated since the Clementi and Habitz potential, with a minimum of -5.5 kcal/mol, nevertheless leads to a second virial coefficient twice the experimental value.[37] van Lenthe et al.[19] suggest that a scaling of the potential, including a reduction of the interaction energy to -4.7 kcal/mol, would produce the correct virial coefficient. The interaction energy calculated here with our best basis set, D_3, is -4.2 kcal/mol after full counterpoise correction. We expect our low value to be due not to an overcorrection by the counterpoise procedure but rather to a number of other factors. First, the dispersion energy is likely too small with our basis set since dispersion is known from previous work to require for saturation orbitals of angular quantum number $l > 3$.[43] For example, even with g orbitals included, the UCHF dispersion energy of He_2 was underestimated by 11%.[43] As a second point, even at full saturation of the basis set, truncation at second-order MP neglects some correlation effects and one would not necessarily expect good agreement with experiment.

Despite these limitations of basis set and correlation which make comparisons with experiment misleading, it is just such comparisons which have prompted previous workers to suggest the full counterpoise term to be an overcorrection of the true BSSE. For example, while their SDCI treatment of Ar--HCl was not attractive enough with full counterpoise correction, van Lenthe et al. obtained improved agreement with experiment after removal of the occupied MOs.[19] However, this apparent superiority of the virtual-only correction is true only for some intermediate-sized basis sets (albeit rather large by current standards). As demonstrated[20] for He_2, failure to include the occupied MOs in the counterpoise correction for a basis set containing g and h polarization functions leads to a potential well twice as deep as observed experimentally; the attraction is not inflated by the full counterpoise correction.

Finally, we would like to comment on a very recent work by Collins and Gallup[44] which would appear to favor the virtual-only correction. In their decomposition of the SCF interaction energy, the first-order term was evaluated as the difference in energy between the subunits in the monomer basis set on one hand and the first SCF iteration of the supermolecule calculation on the other hand. The remaining difference with the fully converged energy involves mainly the attractive second-order terms and is hence expected to be negative. When the full counterpoise correction was subtracted from the latter term, it became positive for selected orientations of He_2, HeH_2, and $(H_2)_2$, whereas the virtual-only correction led to only negative second-order terms. The authors concluded that the full correction can lead to overestimates in certain cases.

The problem with this analysis lies in the definition of the first-order interaction energy which Collins and Gallup assume to be completely free of superposition error. In fact, it has been demonstrated previously[7,10,12,26,27] that the dimer basis set is required for proper evaluation of this term so as to improve the exchange repulsion[26,27] and minimize the zero-exchange term.[10,26] Application of the dimer basis set to the individual monomers would produce modifications of their MOs which would raise the energy of the first SCF iteration.

It is therefore less likely that this energy would be lower than the fully converged dimer energy, even after inclusion of the full counterpoise correction. In light of these observations, we feel that the results do not warrant restriction of the counterpoise correction to only a partial set of orbitals.

In conclusion, there are no signs of overcorrection of our data by use of the full set of ghost orbitals on the partner molecule. Indeed, the evidence gathered to date leads us to believe that this prescription provides a basis set consistent manner of describing interaction energies at the correlated (as well as SCF) level.

ACKNOWLEDGMENTS

[1] P. Schuster, in The Hydrogen Bond—Recent Developments in Theory and Experiments, edited by P. Schuster, G. Zundel, and C. Sandorfy (North-Holland, Amsterdam, 1976).

[2] P. A. Kollman and L. C. Allen, Chem. Rev. 72, 282 (1972).

[3] P. A. Kollman, in Modern Theoretical Chemistry, Vol. 4. Applications of Electronic Structure Theory, edited by H. F. Schaefer III (Plenum, New York, 1977), pp. 109–152.

[4] W. Meyer, P. C. Hariharan, and W. Kutzelnigg, J. Chem. Phys. 73, 1880 (1980).

[5] H.-J. Werner and W. Meyer, Mol. Phys. 31, 855 (1976).

[6] P. G. Burton and U. E. Senff, J. Chem. Phys. 76, 6073 (1982).

[7] M. Urban and P. Hobza, Theor. Chim. Acta (Berlin) 36, 207, 215 (1975).

[8] N. S. Ostlund and D. L. Merrifield, Chem. Phys. Lett. 39, 612 (1976).

[9] M. Bulski and G. Chałasiński, Theor. Chim. Acta (Berlin) 44, 399 (1977).

[10] T. P. Groen and F. B. van Duijneveldt (to be published).

[11] W. Kołos, Theor. Chim. Acta (Berlin) 54, 187 (1980); 51, 219 (1979).

[12] W. A. Sokalski, S. Roszak, P. C. Hariharan, and J. J. Kaufman, Int. J. Quantum Chem. 23, 847 (1983).

[13] B. H. Wells and S. Wilson, Mol. Phys. 50, 1295 (1985).

[14] G. H. F. Diercksen, V. Kellö, and A. J. Sadlej, Chem. Phys. 96, 59 (1985).

[15] S. L. Price and A. J. Stone, Chem. Phys. Lett. 65, 127 (1979).

[16] G. Chałasiński and M. Gutowski, Mol. Phys. 54, 1173 (1985).

[17] J. P. Daudey, P. Claverie, and J. P. Malrieu, Int. J. Quantum Chem. 8, 1 (1974).

[18] M. D. Newton and N. R. Kestner, Chem. Phys. Lett. 94, 198 (1983).

[19] J. H. van Lenthe, T. van Dam, F. B. van Duijneveldt, and L. M. J. Kroon-Batenburg, Faraday Symp. 19, 125 (1984).

[20] M. Gutowski, J. H. van Lenthe, J. Verbeek, F. B. van Duijneveldt, and G. Chałasiński, Chem. Phys. Lett. 124, 370 (1986).

[21] Z. Latajka and S. Scheiner, Chem. Phys. 98, 59 (1985); Chem. Phys. Lett. 105, 435 (1984).

[22] N. R. Kestner, J. Chem. Phys. 48, 652 (1968).

[23] S. F. Boys and F. Bernardi, Mol. Phys. 19, 553 (1970).

[24] A. Johansson, P. Kollman, and S. Rothenberg, Theor. Chim. Acta (Berlin) 29, 167 (1973).

[25] K. Morokuma and K. Kitaura, in Chemical Applications of Atomic and Molecular Electrostatic Potentials, edited by P. Politzer and D. Truhlar (Plenum, New York, 1981), pp. 215–242.

[26] M. Gutowski, G. Chałasiński, and J. van Duijneveldt-van de Rijdt, Int. J. Quantum Chem. 24, 971 (1984).

[27] M. Gutowski, F. B. van Duijneveldt, G. Chałasiński, and L. Piela, J. Chem. Phys. (submitted).

[28] C. Møller and M. S. Plesset, Phys. Rev. 46, 618 (1934); J. S. Binkley and J. A. Pople, Int. J. Quantum Chem. 9, 229 (1975).

[29] B. Jeziorski and M. van Hemert, Mol. Phys. 31, 713 (1976).

[30] G. Karlström and A. J. Sadlej, Theor. Chim. Acta (Berlin) 61, 1 (1982).

[31] P. W. Fowler and A. D. Buckingham, Mol. Phys. 50, 1349 (1983).

[32] J. S. Binkley, R. A. Whiteside, R. Krishnan, R. Seeger, D. J. DeFrees, H. B. Schlegel, S. Topiol, L. R. Kahn, and J. A. Pople, QCPE 1981, 406.

[33] M. R. Peterson and R. A. Poirier, MONSTERGAUSS, Department of Chemistry, University of Toronto, Toronto, Canada.

[34] T. R. Dyke and J. S. Muenter, J. Chem. Phys. 60, 2929 (1973).

[35] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta (Berlin) 28, 213 (1973).

[36] H. Popkie, H. Kistenmacher, and E. Clementi, J. Chem. Phys. 59, 1325 (1973).

[37] E. Clementi and P. Habitz, J. Phys. Chem. 87, 2815 (1983).

[38] Based on gas-phase enthalpy of formation [L. A. Curtiss, D. J. Frurip, and M. Blander, J. Chem. Phys. 71, 2703 (1979)] combined with temperature and zero-point vibrational corrections from frequencies calculated by Z. Slanina, Coll. Czech. Chem. Commun. 45, 3417 (1980).

[39] D. W. Schwenke and D. G. Truhlar, J. Chem. Phys. 82, 2418 (1985).

[40] G. Chałasiński and M. M. Szczęśniak (to be published).

[41] B. Jeziorski and W. Kołos, in Molecular Interactions Vol. 3, edited by H. Ratajczak and W. J. Orville-Thomas (Wiley, New York, 1982), pp. 1–46.

[42] L. A. Curtiss, D. J. Pochatko, A. E. Reed, and F. Weinhold, J. Chem. Phys. 82, 2679 (1985).

[43] M. Gutowski, J. Verbeek, J. H. van Lenthe, and G. Chałasiński (to be published).

[44] J. R. Collins, G. A. Gallup, Chem. Phys. Lett. 123, 56 (1986).