

# Electron Correlation Contribution to the Hydrogen Bond in (HF)<sub>2</sub>

Stephen C. Racine and Ernest R. Davidson\*

Chemistry Department, Indiana University, Bloomington, Indiana 47405

Received: February 1, 1993; In Final Form: March 29, 1993

Hydrogen bonding in the HF dimer is examined using a fixed basis set and geometry and many model wavefunctions. All of the correlated methods based on the Hartree-Fock reference function give essentially the same binding energy. These include MP2, MP4, LCC, CCSD(T), and SDCI (relative to the supermolecule). Multireference methods give an improved absolute energy but a worse binding energy. The basis set superposition error for the basis set developed here was negligible in the Hartree-Fock energy but was still significant in the correlation energy even though 85% of the total correlation energy was recovered. The best result at the optimized MP2 geometry recovered only about 30% of the empirical estimate of the correlation contribution to the binding energy.

## Introduction

The hydrogen fluoride dimer has been studied for many years by both experimentalists and theorists. The best current estimate<sup>1,2</sup> of the experimental dissociation energy  $D_0$  is  $4.87 \pm 0.02$  mE<sub>h</sub> (millihartrees). Theoretical estimates of the harmonic zero-point correction vary from 1.7 to 4.6 mE<sub>h</sub>, with the most reliable values centered about  $2.8 \pm 0.3$  mE<sub>h</sub>. This gives a dissociation energy,  $D_e$ , of  $7.6 \pm 0.3$  mE<sub>h</sub>. There is a further uncertainty in  $D_e$  due to the fact that the HF dimer is a fluxional molecule so the zero-point energy will not be given accurately by the harmonic oscillator approximation. Note that in order to avoid the customary mixture of units, we have chosen to use millihartrees consistently. One mE<sub>h</sub> is 0.627 kcal/mol or 2.625 kJ/mol or  $219.5 \text{ cm}^{-1}$ .

In principle, the only way to calculate the binding energy of a complex is to calculate the dimer and monomer energies separately to greater accuracy than the desired accuracy in the difference. Even for a dimer as small as (HF)<sub>2</sub>, this approach is not feasible with computers available today using expansions in the basis sets. The alternative is to try to truncate calculations for the dimer and monomer in such a way that the energy difference is much more accurate than either energy taken separately. Truncation of the one-electron basis set raises the possibility of basis set superposition error (BSSE) in which the basis functions of one monomer lower the energy of the other monomer in the dimer and destroy the desired cancellation of errors. Similarly, truncation of the many electron basis of configurations raises the possibility of size-inconsistency error (SIE) in which configurations omitted in the dimer are included in the monomers. This causes the calculated energy of the supermolecule (two monomers at infinite separation treated as a dimer) to differ from the sum of the energies of the monomers calculated separately. While these two sources of imbalance in a calculation have been identified and named, fixing them does not assure an accurate result. After all, a SCF calculation using as a basis just the exact occupied molecular orbitals of the monomers would have no BSSE and no SIE.

It is conventional to discuss the SCF and electron correlation contributions to the binding separately. The binding energy  $D_e$  (defined as a positive quantity) is given by

$$-D_e = E_{AB}(AB) - E_A(A) - E_B(B) \quad (1)$$

where the subscript denotes the species whose equilibrium geometry is used to define the monomer bond lengths. This can be split into one- and two-body parts as

$$-D_e' = E_{AB}(AB) - E_{AB}(A) - E_{AB}(B)$$

$$\Delta U_A = E_{AB}(A) - E_A(A)$$

$$\Delta U_B = E_{AB}(B) - E_B(B)$$

$$-D_e = -D_e' + \Delta U_A + \Delta U_B \quad (2)$$

Based on the experimental vibrational frequency,<sup>5</sup> for the small distortion of the monomers in the dimer of (HF)<sub>2</sub>,  $\Delta U_A$  and  $\Delta U_B$  are only about 0.01 mE<sub>h</sub> each and can be neglected. Thus,  $D_e$  and  $D_e'$  are equal to within the experimental uncertainty in  $D_e$ . When the interaction is split into its SCF and correlation parts, it is found that  $\Delta U_A + \Delta U_B$  consists of about +0.3 mE<sub>h</sub> of SCF energy loss and -0.3 mE<sub>h</sub> of correlation energy gain. This large SCF effect comes because the SCF energy minimum for the monomer has a bond length near  $1.695a_0$  which is very different from the experimental bond length of  $1.7328a_0$ . Consequently, the SCF energy has a linear change with bond length for displacements from the experimental bond length, while the total energy has only a quadratic dependence on displacement.

Conventionally, the SCF contribution to binding has been reported as the SCF contribution to  $D_e'$ . The values estimated for this have varied with the basis set. Selected literature values<sup>6</sup> are given in Table I. Many of these values are subject to large BSSE and have been corrected by the counterpoise correction (CPC) found by computing each monomer in the full dimer basis. The converged value for  $D_e'$  (SCF) seems to be  $6.1 \pm 0.1$  mE<sub>h</sub>. This means that the empirical electron correlation contribution to  $D_e'$  is  $1.5 \pm 0.4$  mE<sub>h</sub>. While it is feasible to saturate the basis set at the SCF level to obtain a converged BSSE-free result, this cannot be done for the correlation energy. The empirical correlation energy of the HF monomer is -388 mE<sub>h</sub>, so the correlation energy of the dimer is about -777 mE<sub>h</sub>. In order to obtain a theoretical estimate of the binding energy with an error of less than 0.1 mE<sub>h</sub>, it is necessary to calculate the change in correlation energy between the complex and the monomers to an accuracy of 0.01% of the total correlation energy. The accuracy of the correlation energy is limited by the basis set and the method of calculation. Better basis sets, like the one used here, can still only describe about 85% of the correlation energy, so the calculation requires cancellation of the remaining error to better than 0.1% of the error itself. Consequently, effects such as BSSE and SIE, which may lead to a slightly better description of one system than the other, can easily lead to errors greater than 100% in the correlation contribution to the binding. An attempt is made here to examine methods beyond SCF that will give

\* Author to whom correspondence should be addressed.

TABLE I: Summary of Some Previous Calculations<sup>a</sup>

$D_e(\text{SCF})$	$D_e(\text{Corr})^b$	$E(\text{SCF})^c$	basis	method	ref
9.4		-100 002.9	6-31G*		Dill 1975 <sup>6</sup>
6.1	6.5		EP	CEPA	Lischka 1979 <sup>7</sup>
8.7(5.5)			MODPOT		Sokalski 1983 <sup>8</sup>
	7.3	-100 060.9	TZP	ACCD	Michael 1984 <sup>9</sup>
10.4	10.6		DZ	ACCD	Michael 1984 <sup>9</sup>
7.5	9.1		DZP	ACCD	Michael 1984 <sup>9</sup>
10.4	10.5	-100 017.0	DZ	CISD	Gaw 1984 <sup>3</sup>
7.5	9.1	-100 044.2	DZP	CISD	Gaw 1984 <sup>3</sup>
6.3(6.1)		-100 067.2	E**/2L		Schwenke 1985 <sup>10</sup>
8.3(6.9)		-100 046.7	6-311G**		Schwenke 1985 <sup>10</sup>
6.5(6.2)		-100 065.2	E/2L		Schwenke 1985 <sup>10</sup>
	7.9		6-311++G(2d,2p)	MP2	Frisch 1986 <sup>4</sup>
	8.0		6-311++G(3df,3pd)	MP4	Frisch 1986 <sup>4</sup>
7.6(6.3)	9.2(6.5)		DZP	MP2	Hobza 1986 <sup>11</sup>
9.5(7.1)	11.7(7.2)		6-31G**	MP2	Hobza 1986 <sup>11</sup>
6.2			DZ		Amovilli 1986 <sup>12</sup>
5.0(5.0)		-100 067.3	WR3P		Loushin 1986 <sup>13</sup>
5.5(5.3)	6.7(5.7)		TZ'2P	ACCD	Loushin 1987 <sup>14</sup>
6.0	7.4		6-31G+(2d,2p)	MP4	DelBene 1987 <sup>15</sup>
5.9	7.5		6-311G+(2d,2p)	MP4	DelBene 1987 <sup>15</sup>
6.3(6.2)	7.0(6.2)	-100 030.3	6-31G**VP*(2d)*	MP2	Latajka 1987 <sup>16</sup>
6.5(6.2)	7.1(6.3)	-100 031.0	6-31G**VP*(2d) <sup>p</sup>	MP2	Latajka 1987 <sup>16</sup>
6.3(6.2)	6.9(5.9)	-100 062.8	EP	MP2	Latajka 1987 <sup>16</sup>
(5.8)	(7.0)			SAPT	Rybak 1991 <sup>17</sup>
	(6.4)			MBPT4	Rybak 1991 <sup>17</sup>
6.1(6.1)	6.9(6.5)	-100 070.6	[4s3p2d1f]3s2p1d]	CCSD(T)	present

<sup>a</sup> Binding energy in mE<sub>h</sub> as reported in original article. Some are  $D_e$  relative to monomers at experimental bond length, some are relative to monomers at dimer bond length, and some are relative to computed bond length.  $D_e$  values in parentheses are counterpoise corrected for BSSE. <sup>b</sup>  $D_e$  from calculations including electron correlation. <sup>c</sup> Monomer SCF energies in mE<sub>h</sub>.

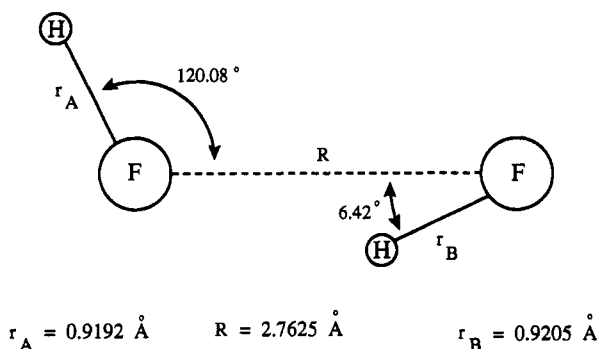


Figure 1. Assumed structure of HF dimer.

meaningful data. Among the methods examined are coupled cluster, many-body perturbation theory, variation-perturbation, and configuration interaction.

## Methods

**A. Geometry.** Michael et al.<sup>9</sup> obtained a theoretical structure by fitting a parametrized potential energy to theoretical energies. We used this as a starting point but decided to change the monomer bond lengths. We felt that a more accurate geometry could be obtained by starting with the experimental equilibrium HF bond lengths and adding to them the difference between the Michael et al.<sup>9</sup> surface fit bond lengths in the dimer and their corresponding bond length of the monomer. Thus, they found the proton donor bond length was stretched by 0.0037 Å and the proton acceptor was stretched by 0.0024 Å. We added these increments to the experimental bond length of 0.9168 Å to give the coordinates given in Table II. We preserved the hydrogen bond F...H distance and the angles found by Michael et al.<sup>9</sup> as shown in Figure 1. This structure agrees with the one determined experimentally by Dyke et al.<sup>18</sup>

**B. Basis Set.** The s and p basis functions for fluorine were based on the Partridge 18s and 13p optimized Gaussian function set.<sup>19</sup> One additional small exponent was added to each set in an even-tempered manner with respect to the last two exponents.

TABLE II: Coordinates of Dimer<sup>a</sup>

	x	y
H	-0.870 599 532	1.503 072 006
F	0	0
H	3.491 808 494	-0.194 503 612
F	5.2204	0

<sup>a</sup> In Bohr units.

We developed the polarization functions for fluorine as 4d,3f sets. Both sets are even tempered with

$$\alpha_d = 0.4933, \beta_d = 2.8, \alpha_f = 0.67, \beta_f = 3.0 \quad (3)$$

$$\xi_k = \alpha \beta^k, \quad k = 0, 1, 2, \dots \quad (4)$$

The s functions for hydrogen were the Partridge 9s basis;<sup>19</sup> the p and d functions were taken from the Dunning 3p,2d,1f hydrogen basis set.<sup>20</sup>

The uncontracted set on the HF molecule (including the s components of Cartesian d and the p components of Cartesian f's) produced the values in Table III at  $R = R_e = 1.7328a_0$ . The SCF energy was within 0.2 mE<sub>h</sub> of the numerical Hartree-Fock value.<sup>21</sup> The total empirical correlation energy<sup>22</sup> of HF is -388 mE<sub>h</sub>. The estimated valence-only value is -320 to -330 mE<sub>h</sub>. If the MP4 contribution from the contracted set is added to the MP2 + MP3 result of the uncontracted basis set, the estimated correlation energy with the uncontracted basis set is 95% of the empirical valence correlation energy.

Most contraction schemes today are based on a set of atomic calculations for each atom in the molecule of interest. We decided to develop our contraction coefficients from natural orbital coefficients of the HF molecule. Since the interaction energy between monomers was the focus of the study, we chose the monomers as the appropriate subunit for coefficients instead of the atoms. The basis was rather tightly contracted to [4s,3p,2d,1f/3s,2p,1d], but the s component of the Cartesian d and the p component of the Cartesian f were retained for added flexibility.

First, a frozen-core single- and double-excitation configuration interaction (FC-SDCI) calculation was done for the monomer (at  $R = 1.732a_0$ ) in the uncontracted basis. The frozen natural orbitals from this wavefunction were then formed (i.e., diago-

TABLE III: Test Results for HF<sup>a</sup>

method	energy
numerical Hartree-Fock <sup>5</sup>	-100 070.8
empirical correlation energy	-388.
estimated valence-valence correlation energy	-325.
Hartree-Fock (uncon.)	-100 070.658
SDCI correlation (uncon.)	-340.993
FC-SDCI correlation (uncon.)	-296.325
MP2 (uncon.)	-352.933
MP3 (uncon.)	-338.263
FC-MP2 (uncon.)	-295.748
FC-MP3 (uncon.)	-269.831
Hartree-Fock (con.)	-100 070.577
FC-SDCI (con.)	-271.011
FC-MP2 (con.)	-278.495
FC-MP3 (con.)	-250.899
FC-MP4 (SDTQ) (con.)	-289.304
FC-CCSD(T) (con.)	-287.652
FC-SDCI(Q) (con.)	-281.120

<sup>a</sup> Energies in mE<sub>h</sub>,  $R = 1.7328a_0$ . SCF total energies and correlation energies by various methods. SDCI = single- and double-excitation CI, FC = frozen core, MP $n$  = perturbation energy through the  $n$ th order, CC = couple cluster, uncon. = uncontracted (19s,14p,4d,3f/9s,3p,2d), con. = contracted [4s,3p,2d,1f/3s,2p,1d].

nalization only of the virtual-virtual block of the density matrix). The fluorine s contracted basis was then chosen using the fluorine component of the 1 $\sigma$ , 2 $\sigma$ , 3 $\sigma$ , and 4 $\sigma$  molecular orbitals. The fluorine p basis was chosen from the 3 $\sigma$ , 1 $\pi$ , and 2 $\pi$  molecular orbitals, the d basis was taken from 3 $\sigma$  and 1 $\pi$ , and the f basis was taken from 3 $\sigma$ . Similarly, the hydrogen s contracted basis was taken as the hydrogen s part of 2 $\sigma$ , 3 $\sigma$ , and 4 $\sigma$ . The hydrogen p basis was taken from 3 $\sigma$  and 1 $\pi$ , and the d basis was taken from 1 $\pi$ . As shown in Table III, this lost only 0.1 mE<sub>h</sub> in the SCF energy but 19 mE<sub>h</sub> in the MP3 energy. The MP4 energy with the final basis is 88% of the estimated valence correlation energy. The SCF dipole moment with the contracted basis is 0.761 au as compared with 0.756 au for the numerical Hartree-Fock.

As a further test of the adequacy of the contracted basis set and correlation methods, the potential curve for HF was computed. The SCF wavefunction gave an  $R_e$  of 1.695  $a_0$  with an energy of -100 071.069 mE<sub>h</sub>. The FC-SDCI minimum energy occurred for  $R_e = 1.722$  with  $E = -100 341.626$  mE<sub>h</sub>. The MP2  $R_e$  was 1.734 with  $E = -100 349.073$  mE<sub>h</sub>. The CCSD(T) (coupled cluster for single and double excitation and triple excitations treated by perturbation theory) gave  $R_e = 1.7328$ , in excellent agreement with experiment, and  $E = -100 358.229$  mE<sub>h</sub>. MP4 (SDTQ) gave 1.7357 and  $E = -100 359.883$  mE<sub>h</sub>. It is disturbing that even MP4 and CCSD(T) differ by 0.003  $a_0$  in the predicted bond length. Even at the fixed bond lengths used for Table III, the MP4 and CCSD(T) energies differ from each other by as much as the empirical correlation contribution to the bonding. Thus, none of these methods can be demonstrated to even reproduce the full CI energy for this modest basis set to within the tolerance needed to allow an unambiguous determination of the true binding energy for this basis.

At the geometry assumed in Table II, the SCF contribution to  $D_e'$  is 6.088 mE<sub>h</sub>. This is in good agreement with the results in Table I from other large basis sets, so the level of contraction is not losing too much flexibility. The counterpoise correction estimate<sup>23</sup> of the BSSE is only 0.012 mE<sub>h</sub>, so the corrected  $D_e'$  is 6.076 mE<sub>h</sub>. Thus, this contraction scheme starting from an accurate monomer basis has negligible BSSE in the SCF energy.

**C. Localized Molecular Orbitals.** In this study, we are chiefly concerned with the interaction energy between two HF monomers. This involves a careful comparison of the dimer and the monomers at each stage of the calculation. In order to compare a configuration in the monomer with the analogous configuration in the dimer, we need analogous MO's. To ensure this, we

localized the dimer orbitals to belong to one or the other monomer in a way that would go smoothly to monomer frozen natural orbitals for the large  $R$  supermolecule calculation.

Let  $U$  be the matrix of monomer MO vectors and  $V$  be the dimer MO's. The monomer vectors are chosen to be frozen natural orbitals from an FC-SDCI calculation in the full dimer basis. Each monomer virtual MO set was truncated to the dimension the virtual space would have been had we used only the basis functions for that monomer and the two nonorthogonal monomer sets were merged in  $U$ . In this way,  $U$  had the same dimension as  $V$ . The SCF occupied and virtual orbital subspaces were localized by applying the equations given below to each subspace separately.

Our first step was to project the  $U$  space onto the  $V$  space to give the MO set  $U'$ . If  $S$  is the overlap matrix of the basis functions, then the least-squares expansion of each MO in  $U$  in the  $V$  set is given by

$$U' = V(V^T S V)^{-1} (V^T S U) \quad (5)$$

These orbitals are not orthogonal to one another, so they must be transformed. The Löwdin orthonormalization<sup>24</sup> will transform the orbitals while maximizing their overlap with the original  $U'$  orbitals:

$$V' = U' Y \quad (6)$$

where  $Y = (U'^T S U')^{-1/2}$  and  $V'$  are the localized dimer orbitals. Note that  $V'$  spans the same vector space as  $V$ . We have used this method previously to localize the SCF occupied orbitals,<sup>25</sup> but a modification is required for the virtual space. When  $Y = (U'^T S U')^{-1/2}$ , the total overlap between  $V'$  and  $U'$  is maximized in an indiscriminate way. This was fine when this scheme was applied to the SCF occupied subspace, but it is not appropriate for the virtual subspace. We wanted the overlap of the most important virtual orbitals to be weighted more than the others. This can be achieved<sup>26</sup> if

$$Y = O(O U'^T S U' O)^{-1/2} \quad (7)$$

where  $O$  represents a diagonal matrix whose diagonal elements are the desired relative weightings on the MO's. We used a very "natural" weighting: the diagonal of the  $O$  matrix contained the occupation numbers from the FC-SDCI natural orbitals calculation and  $U$  contained monomer frozen natural orbitals.

**D. Size Consistency and Basis Set Superposition Error.** Size consistency is a very important consideration in a correlation energy study of this nature because any error introduced in this way can be larger than the true correlation energy contribution to binding. From a naive perspective, the interaction energy of a dimer can be computed from (2). When methods are used that have a SIE, it is usual to replace (2) by

$$-D_e' = E_{AB}(AB) - E_{AB}(AB_\infty) \quad (8)$$

where  $E_{AB}(AB_\infty)$  denotes a supermolecule calculation in which the monomers retain the bond length they have in the dimer. The size-inconsistency error is then given by

$$\text{SIE} = E_{AB}(AB_\infty) - E_{AB}(A) - E_{AB}(B) \quad (9)$$

Hartree-Fock, full CI, CC, and MP methods have no SIE, but truncated CI and variational perturbation methods usually have some error.

Correlated methods are also subject to a BSSE. The counterpoise correction can be symbolized as

$$\text{BSSE}_A = E_{AB}(A) - E_{AB}(Ab) \quad (10)$$

where  $b$  denotes the "ghost" basis functions of B. Ideally, SIE

and BSSE would be corrected simultaneously by

$$-D_e' = E_{AB}(AB) - E_{AB}(AbaB_\infty) \quad (11)$$

which requires a supermolecule calculation with ghost orbitals on both A and B. This is too expensive, so we will approximate this by assuming the errors are additive and

$$-D_e' = [E_{AB}(AB) - E_{AB}(A) - E_{AB}(B)] + BSSE_A + BSSE_B + SIE \quad (12)$$

Because the error in the correlation energy is so much larger than the error in the SCF energy, it is not surprising that ghost orbitals have a larger effect and the BSSE for the correlation energy is larger than desirable. Many workers have, in fact, rejected the counterpoise estimate of the BSSE for correlated calculations because it resulted in a physically unreasonable correlation contribution to  $D_e$ .

Truncated CI calculations are notorious for their large SIE. The difference between Hartree-Fock single- and double-excitation CI (HFSDCI) and full CI can be estimated within an error of  $\pm 10\%$  of the difference by various equations. Unfortunately, a 10% error in the correction exceeds the correlation contribution to  $D_e'$ .

Two methods have been employed here that correct the simple SDCI result to include an estimate of what the energy would be with higher excitations included for either the HFSDCI or a multireference SDCI. The first is the Davidson correction<sup>27</sup>

$$MR(Q)CI = E_{REF} + \frac{C}{2C-1}(E_{CI} - E_{REF}) \quad (13)$$

where  $C$  is the sum of squares of coefficients in the reference space. The second is our multireference extension of the Pople correction<sup>28</sup>

$$Pople(Q) = E_{REF} - \frac{p(2C-1)}{2(1-C)} \left( 1 - \left[ 1 + \frac{4C(1-C)}{p(2C-1)^2} \right]^{1/2} \right) \times (E_{CI} - E_{REF}) \quad (14)$$

where  $p$  is the number of correlated electron pairs and  $C$  is defined the same as in the Davidson correction. Notice that (13) is the limit of (14) as  $p$  goes to infinity. Traditionally, (14) is evaluated using  $p = N/2$  rather than  $\binom{N}{2}$ , where  $N$  is the number of electrons. The best estimate should lie between these two results.

Unfortunately, multireference single and double CI (MRSDCI) expansions generate enormous numbers of configurations and have to be limited further by perturbation selection. Let

$$\psi_0^A = \sum_i c_i^A \phi_i^A \quad \psi_0^B = \sum_j c_j^B \phi_j^B \quad (15)$$

where  $\psi_0^A$  and  $\psi_0^B$  are the reference space part of the monomer wavefunctions and  $c_i^A$  and  $c_j^B$  are the coefficients in the expansion. In order to be consistent, the reference part of the dimer wavefunction should be

$$\psi_0^{AB} = \sum_{ij} c_i^A c_j^B A[\phi_i^A \phi_j^B] \quad (16)$$

where the orbitals used in the configurations are localized and go smoothly to the free monomer orbitals when the monomers are separated.

In perturbation selection of configurations outside the reference space, the second-order perturbation theory energy contribution is calculated for each of the configurations generated. If the energy is less in magnitude than some threshold value, the configuration is discarded and is not used in the expansion. The energy contribution of a configuration  $g_i^A$  outside the reference

space in the free monomer is given by

$$E^{(2)} = \frac{\langle \psi_0^A | H^A | g_i^A \rangle^2}{E_0^A - E_i^A} \quad (17)$$

The final energy is corrected to extrapolate for what the final energy would be if no perturbation theory selection was done. This correction was performed with all methods involving PT selection of configurations

$$(\text{method})_{ex} = \frac{(E_{KEPT} + E_{DISC})(E_{(\text{method})} - E_0)}{E_{KEPT}} + E_0 \quad (18)$$

where  $E_{KEPT}$  and  $E_{DISC}$  refer to the second-order PT energy total values of all configurations kept and discarded.  $E_{(\text{method})}$  is the converged answer involving various methods, and  $E_0$  is the CI energy of the reference space.

Of course, second-order PT is limited in accuracy, but in addition to its accuracy limitations, size extensivity becomes intractable. In the supermolecule dimer, the perturbation estimate of the effect of the configurations  $\phi_j^B g_i^A$  for a fixed  $g_i^A$  is

$$E_{AB}^{(2)} = \sum_j \frac{\langle \psi_0^{AB} | H^{AB} | g_i^A \phi_j^B \rangle^2}{E_0^A + E_0^B - E_j^B - E_i^A} \quad (19)$$

The numerator of this expression can be simplified for the supermolecule to

$$E^{(2)} = \sum_j \frac{\langle \psi_0^A | H^A | g_i^A \rangle^2 (c_j^B)^2}{E_0^A + E_0^B - E_j^B - E_i^A} \quad (20)$$

In order for the values of  $E^{(2)}$  in (17) and (20) to be equal, we need to choose the denominators so that

$$(E_0^A - E_i^A)^{-1} = \sum_j (E_0^A + E_0^B - E_j^B - E_i^A)^{-1} (c_j^B)^2 \quad (21)$$

The only solution to this is to adopt a quasi-degenerate approach and assign each of the reference space configurations  $\phi_j^B$  the same zeroth-order energy  $E_0^B$ .

The usual Epstein-Nesbet approach leads to

$$E_j^B = \langle \phi_j^B | H^B | \phi_j^B \rangle$$

$$E_i^A = \langle g_i^A | H^A | g_i^A \rangle \quad (22)$$

with  $E_0^A$  and  $E_0^B$  being the eigenvalues of the reference space CI. This perturbation theory is not size-consistent in the multireference case.

**E. QDVARPT.** Quasi-degenerate variational perturbation theory, QDVARPT, is a method developed in this laboratory<sup>29,30</sup> as an extension of Hyleraas variation perturbation theory. For the single-reference case, it is identical to linearized coupled cluster, but the multireference version differs slightly from the usual MRLCC method. In the QDVARPT approach, we partition the full CI matrix into a reference space,  $P$ , and a space of single and double excitations,  $Q$ . The QDVARPT matrix equation is then

$$\begin{pmatrix} PHP - E & PHQ \\ QHP & QHQ - E_0 \end{pmatrix} \begin{pmatrix} |P\rangle \\ |Q\rangle \end{pmatrix} = 0 \quad (23)$$

Solving for  $|Q\rangle$  gives

$$Q = -(QHQ - E_0)^{-1} QHP |P\rangle \quad (24)$$

which can be substituted into (31) to give

$$[PHP - PHQ(QHQ - E_0)^{-1} QHP] |P\rangle = E |P\rangle \quad (25)$$

This is an effective Hamiltonian with the dimension of the

TABLE IV: Single-Reference Calculations<sup>a</sup>

	<i>E</i> (dimer)	<i>D</i> <sub>c</sub> '	corr contribution
SCF	-200 146.944	6.088(6.076)	0
MP2	-200 704.986	6.852(6.428)	0.764(0.352)
MP3	-200 704.471	6.968(6.588)	0.880(0.512)
MP4	-200 726.609	6.847(6.410)	0.759(0.334)
CCSD(T)	-200 723.378	6.940(6.487)	0.852(0.411)
SDCI	-200 667.332	-15.748(-16.114)	-21.836(-22.190)
		[6.955(6.589)]	[0.867(0.513)]
SDCI + Q(D)	-200 716.118	5.313(4.915)	-0.775(-1.161)
		[6.852(6.454)]	[0.764(0.378)]
SDCI + Q(P)	-200 709.004	5.657(5.268)	-0.431(-0.808)
		[6.871(6.482)]	[0.783(0.406)]
LCC	-200 718.465	6.822(6.417)	0.734(0.341)
LCC + APC	-200 710.615	6.849(6.458)	0.761(0.382)

<sup>a</sup> Energies in mE<sub>h</sub>. Numbers in parentheses are counterpoise corrected for BSSE. Numbers in square brackets are supermolecule corrected for SIE.

reference space. Methods for solving this for ground and excited states have recently been published.<sup>30</sup> The energy *E*<sub>0</sub> may be chosen as the eigenvalue *E*<sub>0</sub><sup>0</sup> of PHP or as the expectation value

$$E_0 = \langle P|PHP|P \rangle \quad (26)$$

Gdanitz and Ahlrichs recently proposed using an average pair correction in their CEP methods. We have modified their suggestion and incorporated it into a QDVARPT with the APC method.<sup>31</sup> This is accomplished by replacing *E*<sub>0</sub> by

$$E_0 = \langle P|PHP|P \rangle + (E - \langle P|PHP|P \rangle)/p \quad (27)$$

where *p* is the number of correlated electron pairs.

QDVARPT will be size consistent under certain conditions. It is necessary that the reference space of the supermolecule reduces to the direct product of the monomer reference spaces using corresponding localized orbitals and that the selected configurations for the *Q* space are *g*<sup>A</sup> ⊗ *φ*<sup>B</sup> and *φ*<sup>A</sup> ⊗ *g*<sup>B</sup> with the same *g*<sup>A</sup> and *g*<sup>B</sup> as used for the separate monomer *Q*<sup>A</sup> and *Q*<sup>B</sup> spaces. It is also necessary that *E*<sub>0</sub> be chosen to be within a additive constant of the energy eigenvalue *E*<sub>0</sub><sup>0</sup> within the reference space. The alternative choice ⟨*P*|*PHP*|*P*⟩ is not size consistent. The APC extension

$$E_0 = E_0^0 + (E - E_0^0)/p \quad (28)$$

will be consistent provided *p* for the dimer is twice *p* for the monomer and each monomer is identical. This APC extension will not be size consistent for inequivalent monomers such as the HF-H<sub>2</sub>O dimer. Alternatively, APC could be made consistent by adopting some fixed negative number *ε* for the average pair correlation and using

$$E_0 = E_0^0 + \epsilon \quad (29)$$

for all calculations on dimers and monomers.

In the present work, we have proceeded with a size-inconsistent method and used (26) for the QDVARPT calculations. Further, the perturbation selection did not select the full set *g*<sup>A</sup> ⊗ *φ*<sup>B</sup> since some of these configurations are more important than others.

## Results

Table IV summarizes calculations based on corrections to a single function. The MP<sub>*n*</sub> and LCC results are remarkably consistent within ±0.05 mE<sub>h</sub> in the estimate of the correlation contribution to the binding with this basis set at these coordinates. The BSSE correction estimated by the full counterpoise method is nearly the same for all methods and is about 0.38 ± 0.05 mE<sub>h</sub>. As expected, the SDCI method with separate calculations for the dimer and monomers is hopeless. When the SDCI is corrected by using the supermolecule to get the separated monomer energy, then SDCI produces results in the same range as the other

TABLE V: Multireference Calculations<sup>a</sup>

	<i>E</i> (dimer)	<i>D</i> <sub>c</sub> '	corr contribution
reference	-200 312.33	2.38(2.55)	-3.71(-3.53)
MRSDCI	-200 699.14	-6.91(-7.29)	-13.00(-13.37)
		[5.45(5.07)]	[-0.64(-1.01)]
MRSDCI + Q(D)	-200 723.63	4.22(3.84)	-1.87 -2.24
		[5.70(5.32)]	[-0.39 -0.76]
MRSDCI + Q(P)	-200 720.22	4.31(3.93)	-1.78(-2.15)
		[5.65(5.27)]	[-0.44(-0.81)]
QDVARPT	-200 724.39	4.74(4.36)	-1.35(-1.73)
		[5.68(5.30)]	[-0.41(-0.79)]
QDVARPT + APC	-200 720.77	4.74(4.36)	-1.35(-1.72)
		[5.65(5.27)]	[-0.44(-0.81)]
LCC	-200 724.26	5.11(4.71)	-0.98(-1.37)
		[5.69(5.29)]	[-0.40(-0.79)]
LCC + APC	-200 720.64	4.73(4.35)	-1.36(-1.73)
		[5.66(5.28)]	[-0.43(-0.80)]

<sup>a</sup> Energies in mE<sub>h</sub>. Numbers in parentheses are counterpoise corrected for BSSE. Numbers in square brackets are supermolecule corrected for SIE.

methods. Improving the SDCI energy by either (13) or (14) with *p* = *N*/2 produces no appreciable change in the calculated dissociation energy compared to MP4 or CCSD(T). While these results are consistent, the best result is still uncertain by ±15%. Also these results are still far from the empirical estimate of 1.5 ± 0.4 mE<sub>h</sub>, and the BSSE is still half of the uncorrected result.

The total energies of the dimer differ from each other by much more. Here the spread is over 20 mE<sub>h</sub> in total energy among methods producing the same net correlation contribution with a spread of only 0.1 mE<sub>h</sub>. So the errors in these methods are canceling between the monomer and dimer to better than 0.5% of the likely error in each method. The best of these methods, MP4 and CCSD(T), still differ from each other by 3.2 mE<sub>h</sub>, so it is not possible to determine an accurate estimate of the full CI result of the basis set with any of these methods to the desired ±0.1-mE<sub>h</sub> uncertainty.

The approximate corrections for size consistency given by (13) or (14) actually underestimate the correction to the SDCI energy needed to get the full CI energy (assumed to be near -200 725 ± 5 mE<sub>h</sub>). On the other hand, these results would have appeared to be an overestimate if compared with the MP2 result. This is not necessarily a general result, however, since most of the difference between MP4 and MP2 is due to triple excitations which certainly are not considered in the derivation of the approximate corrections to SDCI.

Table V gives the results for calculations using a multireference approach. There is no near degeneracy in this system and the multireference approach would not seem to be required. Nevertheless, the dimer total energy is improved by all of the multireference methods. QDVARPT and LCC produce energies in good agreement with the MP4 and CCSD(T) results in Table IV. All the results except the simple MRSDCI results are within the range of plausible estimates of the full CI energy. Even the multireference extensions of (13) and (14) using the sum of the squares of the coefficients in the reference space for *C* are very good even though they lack rigorous justification. On the other hand, the high-precision cancellation of error required to obtain the correlation contribution to the binding appears to have been lost. When corrected for BSSE and SIE, these methods all give a net repulsive contribution from electron correlation of 0.8 mE<sub>h</sub>. It is remarkable that these methods all produce this same result with a spread of only 0.05 mE<sub>h</sub>.

Part of the difficulty would appear to be due to the fact that the definition of virtual molecular orbitals for the dimer leads to a reference space that is quite repulsive. The orbitals for the monomers are natural orbitals from the SDCI, but for the dimer, they are defined as corresponding orbitals by (5)-(7). Thus, the monomer orbitals are more optimal than the dimer orbitals. Improving the dimer orbitals, however, did not lead to improved results.

The configurations in the reference space of the monomer were the SCF function, plus the six configurations corresponding to the  $2p^6 \rightarrow 2p^4 3p^2$  configurations of the united atom and the three configurations corresponding to  $2p \rightarrow 3p$ . For the dimer, the direct products of these 2 sets of 10 configurations were used with localized (corresponding) orbitals. The SCF and the 12 double excitations in this list are the 13 most important configurations for the dimer.

The LCC calculations should be exactly size consistent. They are off by 0.6 mE<sub>h</sub> because of perturbation selection of configurations (followed by extrapolation) and by a limitation of our CI program. At present, our CI program can only handle configurations with no more than eight singly occupied orbitals. This is sufficient for quadruple excitations from the closed-shell SCF and hence for MRSDCI when the reference space contains only double excitations from the SCF. In the present case, the reference space of the dimer contains quadruple excitations, and double excitations from them include many configurations with more than eight singly occupied orbitals. It would seem from the results that these configurations must contribute 0.6 mE<sub>h</sub> to the energy to make the LCC result size consistent. As these configurations are also missing from the supermolecule, the supermolecule energy forms a better reference point.

## Conclusions

None of the methods examined are able to produce a total energy that can be demonstrated to have an accuracy of better than  $\pm 5$  mE<sub>h</sub>. All of the single-reference methods agree on the correlation contribution to the binding of 0.4 mE<sub>h</sub> with this basis set. All of the multireference methods gave similar net repulsive contributions from electron correlation that are nevertheless probably not the correct full CI result.

While the geometry assumed for these calculations was somewhat arbitrary, it was nonetheless close to optimal. A careful optimization of the MP2 total energy without BSSE correction gave  $r_A = 0.9205$  Å,  $r_B = 0.9233$  Å, and  $R = 2.7637$  Å. The optimum angles were 112.1 and 6.30°. At this point, the MP2 dissociation energy was improved to  $D_e' = 6.973$  mE<sub>h</sub> (or 6.542 mE<sub>h</sub> after BSSE correction). The SCF contribution to the dissociation energy at this geometry was 6.081(6.069) mE<sub>h</sub>, and the correlation contribution was 0.892(0.473) mE<sub>h</sub>. This is an improvement of only 0.1 mE<sub>h</sub> over the geometry assumed in all the other calculations in this paper. Compared to the optimized monomer in the same MP2 approximation, the calculated  $D_e$  is 6.930(6.498) mE<sub>h</sub>, which is still about  $0.9 \pm 0.3$  mE<sub>h</sub> short of the empirical value.

The optimized angle between the monomer axes is 105.8°. This is remarkably close to the tetrahedral angle but is not very close to the best angle for dipole-dipole interaction. Nevertheless, the electrostatic interaction between the free monomer charge distributions (at the dimer values of  $r_A$  and  $r_B$ ) was 10.48 mE<sub>h</sub>, or 172% of the SCF part of  $D_e'$ . In the Morokuma definition, the steric repulsion contribution was -7.56 mE<sub>h</sub> and the contribution from orbital relaxation (polarization and charge transfer) was 3.16 mE<sub>h</sub>.

The MP2-optimized coordinates also agree reasonably well with two other recent estimates. Hancock et al.<sup>32</sup> have added some additional data points to Michael et al.<sup>9</sup> and refit the surface. This gave  $R = 2.764$ ,  $r_A = 0.9242$ ,  $r_B = 0.9257$ , and angles of 117.0 and 7.1°. Independently, Bunker et al.<sup>33</sup> fit a coupled pair functional surface<sup>34</sup> and obtained  $r_A = 0.922$ ,  $r_B = 0.924$ ,  $R = 2.752$ , and angles of 117.9 and 7.5°. Bunker et al. reported a bond length of 0.919 Å for the free monomer compared to the experimental 0.9168-Å value. The energy at the minimum of their PES was -200 614.6 mE<sub>h</sub>, which is rather far from the energies in Table IV but was slightly better than Hancock et al.<sup>32</sup> who obtained -200 570.0 mE<sub>h</sub>.

**Acknowledgment.** This work was supported by Grant No. PHS R01 34081 from the National Institutes of Health.

## References and Notes

- (1) Dayton, D. C.; Jucks, K. W.; Miller, R. E. *J. Chem. Phys.* **1989**, *90*, 2631.
- (2) Pine, A. S.; Howard, B. J. *J. Chem. Phys.* **1986**, *84*, 590.
- (3) Gaw, J. F.; Yamaguchi, Y.; Vincent, M. A.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1984**, *106*, 3133.
- (4) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F., III. *J. Chem. Phys.* **1986**, *84*, 2279.
- (5) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*; Van Nostrand Reinhold: New York, 1979.
- (6) Dill, J. D.; Allen, L. C.; Topp, W. C.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 7220.
- (7) Lischka, H. *Chem. Phys. Lett.* **1979**, *66*, 108.
- (8) Sokalski, W. A.; Hariharan, P. C.; Kaufman, J. J. *J. Phys. Chem.* **1983**, *87*, 2803.
- (9) Michael, D. W.; Dykstra, C. E.; Lisy, J. M. *J. Chem. Phys.* **1984**, *81*, 5998.
- (10) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.
- (11) Hobza, P.; Schneider, B.; Carsky, P.; Zahradnik, R. *THEOCHEM* **1986**, *138*, 377.
- (12) Amovilli, C.; McWeeny, R. *Chem. Phys. Lett.* **1986**, *128*, 11.
- (13) Loushin, S. K.; Liu, S.; Dykstra, C. E. *J. Chem. Phys.* **1986**, *84*, 2720.
- (14) Loushin, S. K.; Dykstra, C. E. *J. Comput. Chem.* **1987**, *8*, 81.
- (15) Del Bene, J. E. *J. Chem. Phys.* **1987**, *86*, 2110.
- (16) Latajka, Z.; Scheiner, S. *J. Comput. Chem.* **1987**, *8*, 663.
- (17) Ribak, S.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **1991**, *95*, 6576.
- (18) Dyke, T. R.; Howard, B. J.; Klempner, W. J. *J. Chem. Phys.* **1972**, *56*, 2442.
- (19) Partridge, H. NASA Technical Memorandum 101044, Jan 1989.
- (20) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
- (21) Sundholm, D.; Pyyko, P.; Laaksonen, L. *Molec. Phys.* **1985**, *56*, 1411.
- (22) Davidson, E. R.; Hagstrom, S. A.; Chakravorty, S. J.; Umar, V. M.; Fischer, C. F. *Phys. Rev. A* **1991**, *44*, 7071.
- (23) Boys, S. F.; Bernardi, F. *Molec. Phys.* **1970**, *19*, 553.
- (24) Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.
- (25) White, J. C.; Davidson, E. R. *J. Chem. Phys.* **1990**, *93*, 8029.
- (26) Carlson, B. C.; Keller, J. M. *Phys. Rev.* **1957**, *105*, 102.
- (27) Langhoff, S. R.; Davidson, E. R. *Int. J. Quant. Chem.* **1974**, *8*, 61.
- (28) Pople, J. A.; Seeger, R.; Krishan, R. *Int. J. Quant. Chem. Symp.* **1977**, *11*, 149.
- (29) Cave, R. J.; Davidson, E. R. *J. Chem. Phys.* **1988**, *89*, 6798.
- (30) Murray, C. W.; Racine, S. C.; Davidson, E. R. *Int. J. Quant. Chem.* **1992**, *42*, 273.
- (31) Gdanitz, J.; Ahlrichs, R. *Chem. Phys. Lett.* **1988**, *143*, 413.
- (32) Hancock, G. C.; Truhlar, D. G.; Dykstra, C. E. *J. Chem. Phys.* **1988**, *88*, 1786.
- (33) Bunker, P. R.; Jensen, P.; Karpfen, A.; Kofranek, M.; Lischka, H. *J. Chem. Phys.* **1990**, *92*, 7432.
- (34) Kofranek, M.; Lischka, H.; Karpfen, A. *Chem. Phys.* **1988**, *121*, 137.