A comparison of linear and nonlinear correlation factors for basis set limit Møller-Plesset second order binding energies and structures of He 2 , Be 2 , and Ne 2

David P. Tew and Wim Klopper

Articles you may be interested in

Comparison of some dispersion-corrected and traditional functionals with CCSD(T) and MP2 ab initio methods: Dispersion, induction, and basis set superposition error
J. Chem. Phys. **137**, 134109 (2012); 10.1063/1.4755990

Complete basis set limit second-order Møller–Plesset calculations for the fcc lattices of neon, argon, krypton, and xenon
J. Chem. Phys. **131**, 244508 (2009); 10.1063/1.3279303

Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set
J. Chem. Phys. **130**, 184103 (2009); 10.1063/1.3126249

Improved supermolecular second order Møller–Plesset intermolecular interaction energies using time-dependent density functional response theory
J. Chem. Phys. **128**, 144112 (2008); 10.1063/1.2905808

Extrapolating to the one-electron basis-set limit in electronic structure calculations
J. Chem. Phys. **126**, 244105 (2007); 10.1063/1.2741259

A comparison of linear and nonlinear correlation factors for basis set limit Møller-Plesset second order binding energies and structures of He₂, Be₂, and Ne₂

David P. Tew and Wim Klopper[a)]

Lehrstuhl für Theoretische Chemie, Institut für Physikalische Chemie, Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany

The basis set limit Møller-Plesset second-order equilibrium bond lengths of He₂, Be₂, and Ne₂, accurate to $0.01a_0$, are computed to be $5.785a_0$, $5.11a_0$, and $6.05a_0$. The corresponding binding energies are 22.4 ± 0.1, 2180 ± 20, and 86 ± 2 μE_h, respectively. An accuracy of 95% in the binding energy requires an aug-cc-pV6Z basis or larger for conventional Møller-Plesset theory. This accuracy is obtained using an aug-cc-pV5Z basis if geminal basis functions with a linear correlation factor are included and with an aug-cc-pVQZ basis if the linear correlation factor is replaced by $\exp(-\gamma r_{12})$ with $\gamma=1$. The correlation factor $r_{12}\exp(-\gamma r_{12})$ does not perform as well, describing the atom more efficiently than the dimer. The geminal functions supplement the orbital basis in the description of both the short-range correlation, at electron coalescence, and the long-range dispersion correlation and the values of γ that give the best binding energies are smaller than those that are optimum for the atom or the dimer. It is important to sufficiently reduce the error due to the resolution of the identity approximation for the three- and four-electron integrals and we recommend the complementary auxiliary basis set method. The effect of both orbital and geminal basis set superposition error must be considered to obtain accurate binding energies with small orbital basis sets. In this respect, we recommend using $\exp(-\gamma r_{12})$ with localized orbitals and the original orbital-variant formalism. © 2006 American Institute of Physics.
[DOI: 10.1063/1.2338037]

I. INTRODUCTION

Ab initio computation is now a reliable method for investigating thermochemistry. Using modern computing techniques quantum chemists are able to obtain highly accurate numerical solutions to the Schrödinger equation and the errors arising from the approximations in their methods are often orders of magnitude less than the values being computed.[1–10] The energetics of weak interactions such as van der Waals (vdW) attractions, however, pose a significant challenge to computational chemistry, since they can be as small as $10-1000$ μE_h, which is in the range of the errors of many conventional computational methods.

Most *ab initio* methods are based on the expansion of the electronic wave function in terms of antisymmetrized products of one-particle functions, Slater determinants, where the one-particle functions are expanded as linear combinations of basis functions and are chosen such that the Hartree-Fock (HF) energy is minimized. The number of these Slater determinants is systematically increased until convergence is obtained to within a desired threshold. A given one-particle basis defines a set of Slater determinants, and the hierarchy of coupled-cluster methods CCSD, CCSDT, CCSDTQ, etc., affords rapid convergence within this space, if the true wave function is well represented by a single Hartree-Fock refer-ence Slater determinant.[11] For the accurate reproduction of experimental binding energies of vdW complexes, it is considered that, in many cases, CCSD(T) is sufficient.[12,13]

However, the convergence with respect to the size of the one-particle basis is extremely slow, the error is observed to reduce as N^{-1} for an optimal one-particle basis of N basis functions characterized by a maximum principal quantum number n.[14] For a method with a computational cost that scales as N^4 with the size of the one-particle basis, each new decimal place in the energy requires a factor of 10 000 more computer time. Consequently, μE_h accuracy is in general unattainable by the conventional methods. Indeed, for accurate vdW binding energies it has been shown that at least augmented quintuple-zeta quality basis sets are required,[13,15–17] and most calculations are thus focused on small systems with rare gas dimers receiving the most attention.[12]

The basis set convergence problem may be partially alleviated through extrapolation techniques, which estimate the basis set limit using empirical formulas for the decay of the error in the energy with the cardinal number of the basis set.[14,16,17] Alternatively, special bond functions may be introduced to saturate the basis in the chemically important bonding regions, which accelerates the convergence of the binding energy.[18] The basis set requirement for both of these methods is nonetheless large. Among the most accurate values for the binding energy of He₂ is from a quantum Monte

[a)]Fax: 49-721-6083319. Electronic mail: klopper@chem-bio.uni-karlsruhe.de

Carlo calculation,[19] but surprisingly few calculations have been performed on larger vdW complexes using this method.

It is well known that the slow convergence of conventional methods is due to the inability of the smooth one-particle basis functions to accurately describe the cusp in the wave function at the point where two electronic coordinates coincide.[20–22] Explicitly correlated methods such as the Gaussian geminal and R12 methods introduce a set of geminal basis functions, explicitly dependent on the interelectronic distance r_{12}.[23,24] Since these basis functions are designed to reproduce the electronic cusp, the energy convergence is greatly accelerated and μE_h accuracy can be obtained in practical time. Both of these methods have been applied successfully to the computation of vdW binding energies[25,26] and have been used to assess various extrapolation formulas.[14] However, the application of these methods to larger complexes has been limited. Indeed, most of the applications of R12 methods have been devoted to benchmarking.[27–30]

Through recent advances in R12 methodology,[31] it has now become evident that R12 methods are capable of delivering near basis set limit accuracy using relatively small conventional basis sets and that they are now a competitive way of obtaining 95% of the correlation energy within a given post-HF method. The two most important advances in this respect are the introduction of an auxiliary basis set for the resolution-of-the-identity (RI) approximation in the evaluation of the three- and four-electron integrals that arise in R12 methods[32] and the introduction of geminal basis functions that depend on functions of r_{12} rather than on linear r_{12}.[33] We refer to the R12 methods with these new geminal basis functions as the F12 methods. In particular, the two correlation factors

$$f(r_{12}) = \exp(-\gamma r_{12}), \tag{1}$$

$$f(r_{12}) = r_{12} \exp(-\gamma r_{12}) \tag{2}$$

are very well suited to the description of the electronic cusp, with exponents γ in the ranges of 1–2 and 0.4–0.8, respectively, for valence shell electrons of neutral species.[34] For reference within this paper we denote the F12 methods using correlation factors (1) and (2) as F12(1) and F12(2), respectively.

The R12 methodologies are well suited to the study of weak interactions since highly accurate correlation energies are required to be confident of meaningful energy differences. It is the purpose of this paper to investigate the performance of the new geminal basis functions on van der Waals dimers. In this exploratory study we focus on the Møller-Plesset second order perturbation (MP2) energies, since if the F12 method is successful for MP2 it is then very likely to be equally successful for coupled-cluster methods. For our test cases, we chose He_2, Be_2, and Ne_2, which are held together by dispersion forces.

A. MP2-R12 and dispersion

A common approach to van der Waals interactions is to take as zeroth order a correlated treatment of the isolated monomers and include the interaction between the monomers as a perturbation. In this description, dispersion arises from the long-range correlation of the electronic motion within monomer A with that within monomer B, such that the fluctuating dipoles align, leading to attraction. In our calculations we take the mean-field, HF description of the dimer as our zeroth order reference, and include the instantaneous interactions between the electrons as a perturbation. It no easy task to quantitatively assign the terms in the MP2 energy of the dimer to terms belonging to the MP2 energy of the isolated monomers, and those due to the interaction between them. It is, however, clear that the long-range correlation of the electronic motion is absent in both zeroth order treatments and is included in our calculations in the MP1 wave function of the dimer. It enters the MP2 energy through the matrix elements $\langle ij|r_{12}^{-1}|ab\rangle$, where ij are occupied spin orbitals and ab are virtual orbitals.

Let us consider the simple example of He_2, with occupied HF orbitals $\sigma_g = (s_A + s_B)/\sqrt{2}$ and $\sigma_u = (s_A - s_B)/\sqrt{2}$ and virtual orbitals $\sigma'_g = (p_A - p_B)/\sqrt{2}$ and $\sigma'_u = (p_A + p_B)/\sqrt{2}$, where the overlap between basis functions on A and B is neglected. Orbital rotation reveals the equivalence of this HF determinant with that corresponding to the orbital occupation $s_A^2 s_B^2$. Similarly the determinant for the configuration $\sigma_g \sigma'_g \sigma_u \sigma'_u$ is equivalent to $s_A p_A s_B p_B$. It is easily verified that the orbital replacement $\sigma_g \sigma_u \rightarrow \sigma'_u \sigma'_g$ includes two kinds of correlation, corresponding to matrix elements $\langle s_A s_A | r_{12}^{-1} | p_A p_A \rangle$ and $\langle s_A s_B | r_{12}^{-1} | p_A p_B \rangle$. The first is associated with the motion of two electrons within monomer A and the second is associated with dispersion, correlating the motion of an electron near A with that of an electron near B, such that there is an attraction between the instantaneous dipoles on A and B.

In explicitly correlated R12 methods, geminal basis functions $w_{12}|ij\rangle$ are considered in addition to the usual orbital pairs. These geminals may be viewed as closed sums of the complete set of orbital pairs with the same symmetry as ij, which includes the orbitals represented by the finite one-particle basis and those outside this basis. The geminal functions are designed to efficiently span the space of high angular momentum functions, absent from a given finite basis, by reproducing the electron cusp.[22] However, since the correlation factors extend to larger r_{12}, the geminals are also seen to compensate for a lack of diffuse functions and contribute to long-range correlation.[34]

For our simple example of He_2 above, we expect that the geminal basis functions will contribute to both types of matrix elements, and therefore to both the short-range and the long-range dispersion correlation. For He_2, however, since the occupied σ orbitals are composed mainly of s orbitals, we do not expect the geminals to be very efficient at describing dispersion effects. In addition, there is no R12 contribution to the dispersion terms arising from the p orbitals orthogonal to the nuclear axis, since there are no π_u or π_g occupied orbitals and therefore no geminals with the correct symmetry. This is no longer true for Ne_2 and we expect the geminal basis functions to have a larger contribution to the vdW interactions. Using a similar analysis, Ten-no comments that the long-range behavior of the correlation factor should decay at least as fast as r_{12}^{-1} so that the r_{12}^{-6} behavior of the

vdW energy is not compromised.[35] For linear r_{12} this requirement is not fulfilled, but the correlation factors in Eqs. (1) and (2) decay exponentially and are expected to yield the correct asymptotic behavior.

We are interested in examining the performance of R12 and F12 for the three van der Waals dimers He$_2$, Be$_2$, and Ne$_2$. The paper is organized as follows: In Sec. II we briefly summarize the sources of deviation from the basis set limit in R12 and F12 methodologies. In Sec. III we present a summary of the computational details of the calculations performed in this paper. In Sec. IV we concern ourselves with the question of how large a basis is required for the MP2-R12 and MP2-F12 methods in order to provide accurate binding energies for our test complexes at a fixed bond length. We examine the sources of error discussed in Sec. II, and investigate the dependence of the MP2-F12 method on the exponent γ. In Sec. V we examine the dependence of the predicted equilibrium bond length on the basis set and the correlation factor. In Sec. VI we compare the orbital-invariant formalism[36] with the original orbital-variant formalism with regard to the linear and nonlinear correlation factors. In particular, we examine the effect of the geminal basis set superposition error and the effect of the long-range behavior of the correlation factor on the vdW binding energy. In Sec. VII we summarize the findings of our investigations.

II. THE ACCURACY OF THE R12 METHOD

The MP2-R12 energy deviates from the basis set limit in several respects. Firstly, since a finite orbital basis is used in the HF calculation, the HF contribution to the energy deviates from the true (basis set limit) energy and approximate HF orbitals are taken as the reference rather than the true orbitals. Secondly, the geminal functions do not exactly compensate for all the deficiencies of the one-particle basis, but merely provide an appropriate description of the electronic cusp, which is the feature most difficult to describe with one-particle functions. Therefore the MP2-R12 energy is at variance with the basis set limit to the extent that the one-particle basis plus geminals (projected onto the one-particle space) fails to span the entire one-particle space. As the one-particle basis increases, this error decreases: the one-particle basis takes care of the overall shape of the wave function and the geminal basis takes care of the cusp. The final source of deviation from the basis set limit is the approximate way in which the three- and four-electron integrals are computed. This deviation comprises of both the RI approximation and the assumption that the generalized Brillouin condition is satisfied for the HF orbitals.

May *et al.* have recently performed an extensive study of these deviations for the correlation energies of several atomic and molecular systems.[37] They conclude that by far the largest source of deviation is the incomplete representation of the one-particle space, showing that this error is greatly reduced when the new geminal functions are employed. Since we require μE_h accuracy for our vdW calculations, it is important for us, likewise, to assess the magnitude of all of the deviations from the basis set limit.

The HF contribution to the vdW binding energy is easily

computed and we report the convergence of these values with respect to the basis set for all the presented examples. The error in the correlation energy due to using an approximate, finite set of HF orbitals and eigenvalues is the intrinsic basis set error of the MP2-R12 method. The remaining deviations are independent of the one-particle basis set. The choice of geminal basis functions has a large effect on the deviation from the basis set limit, and we compare linear r_{12} with the two functions in Eqs. (1) and (2) with various values of γ. We do not present any estimate of the effect of the generalized Brillouin approximation, but we point to the work of May *et al.*, which reports that it is smaller than the RI error in their calculations. Finally, we assess the effect of the RI approximation by using the successively larger uncontracted aug-cc-pVXZ basis sets for the auxiliary basis until the RI error is effectively removed.

III. COMPUTATIONAL DETAILS

All MP2-R12 and MP2–F12 calculations are performed with ansatz 2. For our investigations of equilibrium binding energies and structures we use approximation B and for reasons of clarity we adopt approximation A for our discussion of the original orbital-variant method (see Ref. 32 for details of ansatz 2 and approximations A and B; we use the hybrid approach of Ref. 38). In ansatz 2, the explicitly correlated pair functions take the form

$$w_{12}|ij\rangle = (1 - \hat{O}_1)(1 - \hat{O}_2)f_{12}|ij\rangle, \tag{3}$$

where $\hat{O} = \Sigma_k |k\rangle\langle k|$ is the projector onto the space spanned by the occupied HF orbitals and where $|ij\rangle$ denotes the symmetrized product of two such HF orbitals φ_i and φ_j. The invocation of approximation B implies that no terms are neglected, whereas the contributions involving the exchange operator and the eigenvalue-weighted overlap of the geminals are neglected in approximation A. The auxiliary basis sets for the RI approximation are used within the complementary auxiliary basis set (CABS) approach of Valeev,[39] where the set of functions used for the RI approximation is the union of the orbital basis and the auxiliary basis.

We approximate the correlation factor f_{12} of Eq. (1) by a linear combination of six Gaussian-type geminals, and similarly, the correlation factor f_{12} of Eq. (2) by a linear combination of six products of r_{12} with Gaussian-type geminals, weighting the fit to bias small r_{12}. The expansion coefficients and exponents are those of Ref. 34. All MP2-R12 and MP2–F12 calculations were carried out with a local version of the DALTON program.[40]

It is important to include diffuse functions in the one-particle basis for MP2-R12 and MP2-F12 calculations,[34] and we therefore use the augmented Dunning correlation consistent basis sets. For He we use the aug-cc-pVXZ basis sets with $X = T$–6.[41] For Be we augment the cc-pVXZ basis sets $X = T$–5 (Ref. 42) with one extra set of diffuse functions for each angular momentum quantum number l represented in the basis. For each l, the exponent of the extra diffuse functions α_{n+1} is chosen such that the ratio $\alpha_{n+1}: \alpha_n = \alpha_n: \alpha_{n-1}$, where α_n is the smallest exponent of the n sets of functions with angular momentum l present in the cc-pVXZ basis set

TABLE I. Computed HF, MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ binding energies of He$_2$ at 5.79a_0, together with the removed BSSE and the RI error for an uncontracted aug-cc-pVYZ auxiliary basis where $Y=X+1$. All energies are in μE_h.

	HF		MP2		R12			F12(2) (γ=0.4)			F12(1) (γ=1.0)		
X	E_{bind}^X	E_{BSSE}^X	E_{bind}^X	E_{BSSE}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X
T	−18.2	−1.1	17.6	−4.6	19.6	−10.8	0.1	18.8	−3.2	−0.3	22.1	−2.2	1.0
Q	−18.3	−1.3	19.5	−2.9	20.9	−9.2	−0.3	20.5	−2.9	−0.1	22.1	−1.6	0.2
5	−18.3	−0.7	20.7	−1.7	22.2	−7.2	−0.8	21.3	−1.6	0.0	22.2	−0.9	0.1
6	−18.3	−0.1	21.6	−0.8	22.0	−4.8	⋯	22.0	−0.5	⋯	22.4	−0.1	⋯

and α_{n-1} is the next smallest. Since in the cc-pVXZ basis sets there is only one set of functions with the highest angular momentum, we arbitrarily choose the ratio to be equal to one-third. For Ne we use the aug-cc-pVXZ basis sets with X=T−6.[42,43] All our calculations involving Be and Ne are with a frozen core.

The auxiliary basis sets employed are the corresponding uncontracted aug-cc-pVXZ sets. It became necessary to use auxiliary basis sets larger than aug-cc-pV6Z to sufficiently reduce the error due to the RI approximation. We therefore construct an 18s10p6d5f4g3h2i basis for He, a 21s12p7d5f4g3h2i for Be, and a 21s14p8d7f5g4h3i2k for Ne from the s functions of the well tempered basis sets (WTBSs),[44,45] using the rule $\zeta^l = \zeta^0(l+3)/3$ to generate the exponents for functions with angular momentum l from the s function exponents. We extended each WTBS by one s function at the low exponent end and always chose the set of smallest s exponents to generate the set of l exponents.

IV. BINDING ENERGIES

Let us first concentrate on the prediction of the MP2 binding energy for a vdW complex with a given bond length. The main aim of this work is to compare the performance of MP2-F12 with MP2-R12 for vdW energies, and to examine the size of one-particle basis required to provide accurate vdW energies in both cases. An accurate value is obtained if the energies of both the dimer and the monomer are converged to well below the value for the binding energy. Since this is usually not the case in *ab initio* calculations and significant errors remain in both energies, an effort is made to reduce the error in the binding energy through the counterpoise correction.[46] The same orbital basis is thus used for the monomer as for the dimer in the hope that the error due to the incomplete basis is consistent for the monomer and dimer, and therefore cancels in the binding energy. For explicitly correlated methods there is also a basis set superpo-

sition error (BSSE) arising from the geminal functions. We do not attempt to remove this geminal BSSE from our calculations, but we discuss this issue further in Sec. VI A. In the following sections we report counterpoise corrected binding energies, together with the value of the removed orbital BSSE. The chosen bond lengths are 5.79a_0 for He$_2$, 5.11a_0 for Be$_2$, and 6.05a_0 for Ne$_2$, each of which is close to the basis set limit MP2 equilibrium bond length.

For each of the vdW dimers He$_2$, Be$_2$, and Ne$_2$ we perform MP2-R12 and MP2-F12 calculations on the dimers and the atoms using systematically larger orbital and auxiliary basis sets. For a given aug-cc-pVXZ orbital basis, we compute the counterpoise corrected binding energies E_{bind}^X, the BSSE E_{BSSE}^X, and the error due to the RI approximation E_{RI}^X when an uncontracted aug-cc-pV$(X+1)$Z auxiliary basis is used. These quantities are given by the following formulas:

$$E_{bind}^X = E_{bind}^{X,\infty}, \quad E_{bind}^{X,Y} = 2E_A^{X,Y}(A_2) - E_{A_2}^{X,Y}(A_2), \quad (4)$$

$$E_{BSSE}^X = E_{BSSE}^{X,\infty}, \quad E_{BSSE}^{X,Y} = 2E_A^{X,Y}(A_2) - 2E_A^{X,Y}(A), \quad (5)$$

$$E_{RI}^X = E_{RI}^{X,X+1}, \quad E_{RI}^{X,Y} = E_{bind}^{X,Y} - E_{bind}^{X,\infty}, \quad (6)$$

where $E_A^{X,Y}(A_2)$ denotes the energy of monomer A computed in the basis of the dimer A_2, using an orbital basis aug-cc-pVXZ and an uncontracted auxiliary basis aug-cc-pVYZ. $Y=\infty$ denotes our large auxiliary basis sets described in the previous section. Note that a bound dimer is characterized by a positive binding energy and that the BSSE correction always reduces the binding energy and is already included in E_{bind}^X. The RI error E_{RI}^X can have any sign, a positive value indicating that $E_{bind}^{X,X+1}$ is overestimated due to the RI error.

We present the results from MP2, MP2-R12, and MP2-F12 calculations of the binding energy in Tables I–III. We present MP2-F12 results for the two functions f_{12} in Eqs. (1) and (2), with exponents γ=1.0 and 0.4, respectively. In the following three sections we comment on these results and

TABLE II. Computed HF, MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ binding energies of Be$_2$ at 5.11a_0, together with the removed BSSE and the RI error for an uncontracted aug-cc-pVYZ auxiliary basis where $Y=X+1$. All energies are in μE_h.

	HF		MP2		R12			F12(2) (γ=0.4)			F12(1) (γ=1.0)		
X	E_{bind}^X	E_{BSSE}^X	E_{bind}^X	E_{BSSE}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X	E_{bind}^X	E_{BSSE}^X	E_{RI}^X
T	−7617	−4	1397	−351	1641	−104	−1.5	1996	−66	0.0	2076	−86	0.0
Q	−7584	−3	1825	−142	2063	−59	0.2	2153	−50	0.1	2159	−45	0.0
5	−7578	−2	2008	−93	2143	−58	⋯	2178	−20	⋯	2181	−23	⋯

TABLE III. Computed HF, MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ binding energies of Ne$_2$ at $6.05a_0$, together with the removed BSSE and the RI error for an uncontracted aug-cc-pVYZ auxiliary basis where $Y=X+1$. All energies are in μE_h.

	HF		MP2		R12			F12(2) ($\gamma=0.4$)			F12(1) ($\gamma=1.0$)		
X	E_bind^X	E_BSSE^X	E_bind^X	E_BSSE^X	E_bind^X	E_BSSE^X	E_RI^X	E_bind^X	E_BSSE^X	E_RI^X	E_bind^X	E_BSSE^X	E_RI^X
T	−56.7	−39.1	55.0	−76.1	87.4	−186.3	−8.4	65.5	−48.1	0.2	90.5	−33.3	16.8
Q	−55.3	−14.6	71.6	−42.3	83.2	−70.2	−3.3	78.7	−18.3	−0.8	92.7	−18.2	3.0
5	−56.3	−4.7	78.6	−20.3	82.6	−14.7	−1.2	82.7	−7.5	−0.5	86.7	−5.9	0.7
6	−56.3	−0.7	81.8	−8.9	84.4	−8.1	⋯	84.8	−2.3	⋯	86.4	−1.3	⋯

give an estimation of the total size of the deviation from the basis set limit for our best results and evaluate the size of the basis required to obtain 95% of the binding energy. We compare linear r_{12} with the two new correlation factors and comment upon the relative performance of each, paying particular attention to the dependence of the MP2-F12 results on the exponent γ.

A. He–He

We see from Table I that the HF contribution to the binding energy is converged to within 0.1 μE_h. However, for the atom, the difference in the computed HF energy between the aug-cc-pV5Z and aug-cc-pV6Z basis sets is 46 μE_h. The deviation from the basis set limit for the dimer is almost precisely twice that of the atom. It seems clear that the source of this error is the description of the nuclear cusp, which has a negligible effect on the vdW binding energy. The accuracy of the HF contribution, even at the aug-cc-pVTZ level, is more than sufficient for our purposes.

The computed binding energies for the MP2, MP2-R12, and MP2-F12 methods all increase in magnitude as the basis set increases. Since we observe that the correlation energy is always underestimated, it follows that the error in the dimer is dominant over that of the atom. It is apparent from Table I that MP2-F12(1) converges to the basis set limit much faster than MP2-R12. This is particularly evident from the BSSE corrections, since large BSSE values indicate that the energy of the atom is far from converged. Noting that the MP2 values contain the HF BSSE, at the aug-cc-pV6Z level the MP2-F12(1) BSSE is below 0.1 μE_h, compared to 0.5 and 4.7 μE_h for MP2-F12(2) and MP2-R12, respectively. Even though the MP2-F12(2) BSSE is an order of magnitude smaller than that of MP2-R12, the binding energies converge slower than those from MP2-R12 and are only a slight improvement over those from conventional MP2. We conclude that while both MP2-F12 methods yield much improved correlation energies over MP2-R12, MP2-F12(2) does not provide a sufficiently balanced treatment of the atom and the dimer, leading to less error cancellation in the binding energy, in contrast to MP2-R12 and MP2-F12(1).

The size of auxiliary basis required to sufficiently reduce the RI error is also of interest. In Table I we present E_RI^X, the RI error remaining when an uncontracted aug-cc-pV$(X+1)$Z auxiliary basis is used for a calculation with an aug-cc-pVXZ orbital basis. For MP2-F12 the three- and four-electron integrals are shorter range due to the exponentially damped r_{12} contribution; the integrals are more compact and

require less terms in the RI expansion for a given accuracy. This is a further clear advantage of MP2-F12 over MP2-R12.

It is important to assess the importance of the exponent γ in Eqs. (1) and (2) for the MP2-F12 vdW binding energies. In Fig. 1 we present the dependence of E_bind^X on γ for $X=$ T −6. The values for MP2-F12(1) and (2) recover the MP2 and MP2-R12 results, respectively, when γ is zero. The variation in the binding energy with γ originates from the relative accuracy in the description of the dimer and the atom. Consistent changes in the deviations from the basis set limit for all values of γ would give flat lines in Fig. 1. The values of E_bind^X all display the same behavior as γ increases from zero, increasing initially, reaching a maximum at a relatively small value of γ, then decreasing to a minimum at intermediate values of γ, and thereon increasing as γ increases. The range of binding energies obtained by varying γ shrinks as the basis set increases, primarily since the region around the cusp not captured by the orbital description shrinks and the value of γ becomes less important. MP2-F12(1) is much less dependent on γ than MP2-F12(2). Taking the aug-cc-pVTZ basis results as the worst case example, the difference in energy between the maxima and minima of the undulating binding energy is 3.7 μE_h for MP2-F12(2), compared to 0.8 μE_h for MP2-F12(1).

The observed dependence of E_bind^X on γ is unexpected and warrants special attention. It is known that it is important to consider separately the singlet and triplet pair contributions to the binding energy of He$_2$, since there is only a singlet pair contribution to the energy of the dissociated
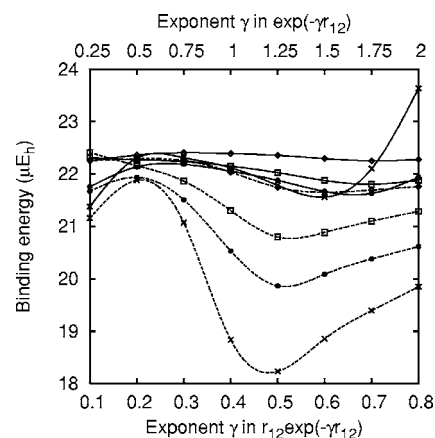

FIG. 1. The MP2-F12/2B binding energy of He$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) with various γ and basis sets: ($\times$) aug-cc-pVTZ, ($\bullet$) aug-cc-pVQZ, ($\square$) aug-cc-pV5Z, and ($\diamond$) aug-cc-pV6Z.

atoms.[17] The values of γ that give the maximum singlet pair energies for the monomer and the dimer are almost identical and are 0.9 and 0.6 for (1) and (2), respectively. However, the values that give the maximum triplet pair energies for the dimer are much smaller, 0.4 and 0.2 for (1) and (2), respectively, which is consistent with our study of the optimum correlation factor for ortho- and parahelium in Ref. 34. For MP2-F12(1) there is a balanced description of the singlet pair contributions to the atom and dimer and the interaction energy is fairly flat over the range of 0.6–0.9. The improvement of the description of the wave function near triplet pair coalescence is responsible for the increase in the vdW binding energy for smaller values of γ in Fig. 1. MP2-F12(2), however, gives a less balanced description of the atom and dimer over the range of 0.2–0.4, increasingly favoring the atom.

As γ decreases to zero, correlation factor (2) converges to linear r_{12}. Similarly, the structure of correlation factor (1) also diminishes and the term in linear r_{12} dominates over the chemically important range of r_{12}. Thus, for small values of γ, the correlation factors describe the immediate region of the electron coalescence, but do not capture intermediate or long-range correlation effects efficiently. At small γ there is therefore little bias of the dimer over the atom, or vice versa and, since the orbital basis is optimized for atom, the binding energy is underestimated at small γ in an identical manner to MP2-R12 and conventional MP2.

The behavior at large γ remains to be understood. It is both counterintuitive and somewhat alarming that the description of the atom deteriorates faster than that of the dimer as γ increases, resulting in greatly overestimated binding energies for MP2-F12(1) with $\gamma>2$ using the aug-cc-pVTZ basis. The reason for this failure is that, although the conventional orbital BSSE is removed from our calculations, the BSSE arising from the geminal functions is not (see Sec. VI A for a detailed discussion). Transforming to localized orbitals and examining the γ dependence of the dimer F12 pair energy contributions reveals that the excitations of the type $w_{12}|s_A s_A\rangle \rightarrow w_{12}|s_A s_B\rangle$ become increasingly important as γ increases. Since this excitation is only present in the dimer basis and not in the atom, the dimer is therefore stabilized relative to the atom, leading to overbinding. We argue that as γ increases, the efficiency of the geminals at describing long-range correlation decreases and the importance of the above excitation increases since it represents a long-range interaction. The geminal BSSE begins to be a problem for (1) for $\gamma>1.5$. Since the ratio between the values of γ for (1) and (2) that give the same values for $\langle f_{12}|r_{12}|f_{12}\rangle/\langle f_{12}|f_{12}\rangle$ is equal to 5/3, we argue that the problems for (2) will begin for $\gamma>0.9$. We show in Sec. VI A that this spurious behavior can be avoided by using local orbitals and reverting to the original orbital-variant formalism.

The basis set limit of the binding energy is estimated to be 22.4 μE_h and we attach an uncertainty of ± 0.1 μE_h. This is lower than many earlier computations of the basis set limit interaction energy of He$_2$ since they were performed at a bond length of $5.60a_0$.[18,47,48] This value is in very good agreement with the value of 22.3 μE_h computed by van Mourik *et al.* using conventional MP2 and a t-aug-cc-pV6Z
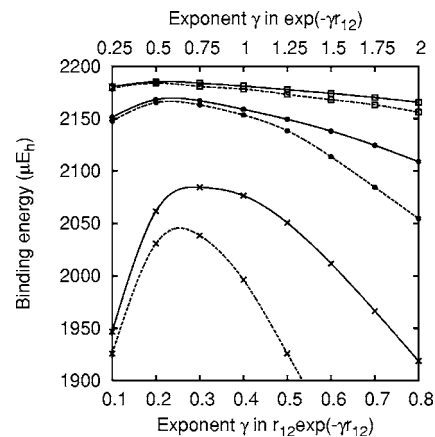

FIG. 2. The MP2-F12/2B binding energy of Be$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) with various γ and basis sets: ($\times$) aug-cc-pVTZ, ($\bullet$) aug-cc-pVQZ, and ($\square$) aug-cc-pV5Z.

basis.[49] Their estimation of the basis set limit, however, looks suspect, probably due to the use of aug-cc-pVDZ values in their extrapolation formula.[14] It is remarkable that, for the new MP2-F12 method with correlation factor $\exp(-r_{12})$, the vdW binding energy is within 95% of this value using the aug-cc-pVTZ basis. This level of accuracy requires the aug-cc-pV5Z basis set with the MP2-R12 method, and the aug-cc-pV6Z basis for conventional MP2. We highlight the importance of the CABS approach in removing the RI error.

B. Be–Be

We see from Table II that the energy difference between the HF binding energies computed with the aug-cc-pVQZ and the aug-cc-pV5Z is 6 μE_h, but that HF binding energy computed using the aug-cc-pVTZ is in error by over 39 μE_h, which corresponds to an error of 2% in the MP2 binding energy. The HF BSSE is negligible.

Convergence of the MP2 binding energies is from below in all cases and the basis set limit of the atom is reached more easily than that of the dimer. It is clear from Table II that the convergence of the conventional MP2 values is extremely slow, with a 10% error still remaining in the binding energy computed with an aug-cc-pVSZ basis. MP2-R12 is a great improvement over conventional MP2, particularly for the larger basis sets where, for example, the MP2-R12/aug-cc-pVQZ value is better than the MP2/aug-cc-pV5Z value. The MP2-R12/aug-cc-pVTZ binding energy, however, is in error by 25%. Both of the MP2-F12 methods improve significantly on the MP2-R12 values for all basis sets, but the improvement is most evident at the aug-cc-pVTZ level, where MP2-F12(1) recovers 95% of the binding energy, with 2% of the remaining error arising from the HF contribution. The MP2-F12(1) values are seen to be superior to MP2-F12(2) at all levels and we highlight the smaller E_{BSSE}^X values. The RI errors are negligible for all of the explicitly correlated methods and, in contrast to He$_2$, the size of the RI basis is not critical for 95% accuracy.

In Fig. 2 we present the dependence of the MP2-F12(1) and MP2-F12(2) binding energies on the exponent γ for the aug-cc-pVXZ basis sets $X=T-5$. We see that, for all three

basis sets, MP2-F12(1) is superior to MP2-F12(2) and that the binding energy computed using (1) is less dependent on γ than that using (2). In addition we report that the values of γ for which the maximum MP2 correlation energy is recovered are between 1.0 and 1.2, and 0.4 and 0.5 for (1) and (2), respectively, increasing over these ranges as the basis set increases. The optimum γ is a few percent smaller for the dimer than for the atom, which indicates that the orbital basis lacks sufficient diffuse functions that are important for the long-range correlation in the dimer.

The curves in Fig. 2 therefore indicate that the description of the dimer deteriorates faster than that of the atom as γ increases from 1.0 or 0.4 for (1) and (2), respectively, and the binding energy decreases. Due to the relatively strong interaction, long-range correlation is much more important in the dimer than in the atom and the energy of the dimer is more sensitive to contracting the geminals by increasing γ than that of the atom. This effect is much larger than the small increase in the binding energy due to the extra excitations between the geminals in the dimer and the geminal BSSE does not cause problems for the calculation of the binding energy of Be$_2$.

Conversely, as γ decreases from 1.0 to 0.5 for (1) and from 0.4 to 0.2 for (2), the description of the atom deteriorates faster than that of the dimer and E_{bind}^{X} increases slightly. Decreasing γ reduces the efficiency with which the electronic cusp is described, but increases the flexibility for describing longer-range correlation. For very small values of γ the lack of structure in the correlation factor removes this advantage, the description of the dimer becomes worse than that of the atom and E_{bind}^{X} decreases.

The basis set limit of the binding energy for Be$_2$ is estimated to be 2180 μE_{h} and we attach an uncertainty of ± 20 μE_{h}. It is remarkable that, for the new MP2-F12 method with correlation factor $\exp(-r_{12})$, the vdW binding energy is within 95% of this value using the aug-cc-pVTZ basis. This level of accuracy requires the aug-cc-pVQZ basis set with the MP2-R12 method, and is not reached using an aug-cc-pV5Z basis with conventional MP2.

C. Ne–Ne

We see from Table III that the HF binding energy oscillates slightly as the basis set increases from $X=$T to $X=$5 before converging. The HF binding energy for Ne$_2$ is thrice that of He$_2$, but the individual HF energies of the dimer and atom are 50 times larger. The convergence of the HF binding energy of Ne$_2$ is much more demanding than that of He$_2$. Indeed, the difference between the HF energies of the dimer computed with an aug-cc-pV5Z and aug-cc-pV6Z basis is 550 μE_{h}, but this is associated with the nuclear cusp and cancels in the binding energy, which is converged to less than 0.1 μE_{h}.

The MP2-R12 and MP2-F12(1) binding energies also oscillate as the basis set increases and the oscillations are larger than those due to the HF contribution. The conventional MP2 values converge slowly from below as one would expect and the MP2-F12(2) values are an improvement over those of conventional MP2, also converging from below, but
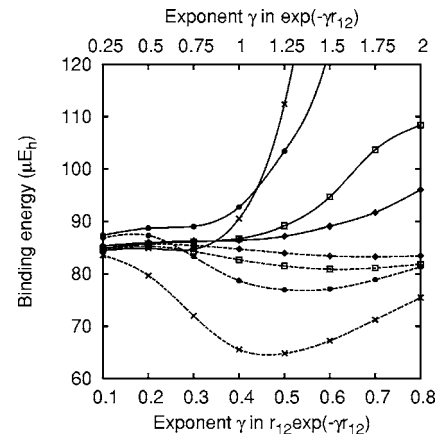

FIG. 3. The MP2-F12/2B binding energy of Ne$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) with various γ and basis sets: ($\times$) aug-cc-pVTZ, ($\bullet$) aug-cc-pVQZ, ($\square$) aug-cc-pV5Z, and ($\diamond$) aug-cc-pV6Z.

are much more slowly convergent than those of MP2-R12. Indeed, it is remarkable that, even though convergence is somewhat erratic, the MP2-R12 values are within 3% of the basis set limit for all basis sets considered. In this instance MP2-R12 gives better values for E_{bind}^{X} than both of the MP2F12 methods. However, the size of the BSSE removed from the binding energy is much larger for MP2-R12 than for either MP2-F12(1) or MP2-F12(2), indicating that the individual energies of the atom and the dimer are far from converged. We conclude that the MP2-R12 gives a more balanced treatment of the atom and dimer than either MP2-F12(1) or MP2-F12(2).

In Fig. 3 we present the dependence of the MP2-F12 binding energies on the exponent γ in Eqs. (1) and (2) for the aug-cc-pVXZ basis sets $X=$T-6. An undulating pattern is observed, very similar to that of He$_2$. In addition we report that the optimum values of γ for both the atom and the dimer are almost identical and are 1.7 and 0.7 for MP2-F12(1) and MP2-F12(2), respectively. We see that MP2-F12(2) gives a better description of the atom than the dimer close to this value of γ, which was also observed for He$_2$. For smaller values of γ the bias due to the incomplete basis shifts towards the dimer due to the importance of diffuse functions. For both correlation factors, the dependence of the binding energy on γ decreases as the basis set increases.

For MP2-F12(1), we see that as γ increases the binding energy increases exponentially, reaching 400 μE_{h} at $\gamma=2.0$ for the aug-cc-pVTZ basis, more than 300% overbound. This catastrophic failure is due to the unremoved BSSE arising from the larger geminal basis in the dimer than in the atom. The geminal BSSE is much larger than for He$_2$ both because there are many more geminal basis functions that contribute and because the individual energies are much further from convergence than in the case of He$_2$. The geminal BSSE does not cause problems for MP2-F12(2) over the computed range of $0 \leqslant \gamma \leqslant 0.8$, but is expected to cause difficulties for $\gamma > 0.8$. We show in Sec. VI A that the spurious behavior is removed if we use localized orbitals with the original orbital-variant formalism.

The RI error is also of interest, we report that the small

TABLE IV. Computed MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ equilibrium bond distances of He$_2$ in a_0.

X	MP2	R12	F12(2) γ			F12(1) γ		
			0.2	0.4	0.6	0.5	1.0	1.5
T	5.92	5.86	5.80	5.86	5.89	5.80	5.80	5.81
Q	5.86	5.83	5.80	5.81	5.86	5.80	5.80	5.80
5	5.82	5.81	5.80	5.80	5.83	5.80	5.79	5.79
6	5.80	5.80	5.785	5.79	5.80	5.785	5.785	5.785

values for E_{RI}^X are due to a large cancellation of the individual RI errors for the dimer and monomer. For the aug-cc-pVTZ basis, the differences in the correlation energy of the dimer computed with an uncontracted auxiliary aug-cc-pVQZ basis compared with that computed with the large auxiliary basis are some 500 μE_{h} for MP2-R12 and 300 μE_{h} for both MP2-F12 methods, an order of magnitude larger than those observed for He$_2$ and Be$_2$. If the errors due to the RI approximation are not properly controlled, large variations in the interaction energy may be introduced. We recommend the CABS approach.[39]

The basis set limit of the binding energy for Ne$_2$ is estimated to be 86 μE_{h} and we attach an uncertainty of ± 2 μE_{h}. This is in very good agreement with the value of 85.5 μE_{h} obtained by van Mourik *et al.* using conventional MP2 with a t-aug-cc-pV6Z basis.[49] Due to the large geminal BSSE for the MP2-F12(1) method, the binding energy is overestimated for small basis sets and 95% accuracy is not reached before the aug-cc-pV5Z level. For MP2-R12 on the other hand, this accuracy is reached with the aug-cc-pVTZ basis and we expect that once the issue of geminal BSSE is dealt with, MP2-F12(1) will perform at least as well as MP2-R12.

V. BOND LENGTHS

Accurate predictions of equilibrium bond distances require that the deviations from the basis set limit are sufficiently independent of the bond length, so as not to bias the potential curve, shifting the minimum. The size of energy bias B in, $E_{\mathrm{h}} a_0^{-1}$, required to shift the maximum in E_{bind}^X (minimum in the potential) by $\pm \Delta a_0$ is related to the force constant k,

$$B = \pm k\Delta. \qquad (7)$$

As before we are interested in comparing the performance of MP2-F12 with MP2-R12, and investigating the size of one-particle basis required to provide reliable equilibrium bond length predictions to within $0.02 a_0$. For the MP2-F12 method it is important to determine whether the choice of exponent γ in the correlation factor significantly affects equilibrium bond lengths.

For each of the vdW dimers He$_2$, Be$_2$, and Ne$_2$ we compute the minimum of the counterpoise corrected MP2-R12 and MP2-F12 potential curves using the available aug-cc-pVXZ orbital basis sets. We use the large $18s10p6d5f4g3h2i$, $21s12p7d5f4g3h2i$, and $21s14p8d7f5g4h3i2k$ auxiliary basis sets for He, Be, and Ne, respectively, which are sufficient to reduce the effect of the RI approximation on the bond length to below $0.005 a_0$ in

all cases. For MP2-F12(1) we use three exponents $\gamma = 0.5$, 1.0, and 1.5 and for MP2-F12(2) we use $\gamma = 0.2$, 0.4, and 0.6. We present the results from MP2, MP2-R12, and MP2-F12 calculations in Tables IV–VI.

In the following three sections we comment on these results, comparing the performance of MP2-R12 and of MP2-F12. We discuss the effect of each source of deviation from the basis set limit on the potential curve, paying particular attention to the dependence of the MP2-F12 results on the exponent γ.

A. He–He

We see from Table IV that the relative rates of convergence of the computed MP2, MP2-R12, and MP2-F12 equilibrium bond lengths to the basis set limit follow a similar pattern to that of the binding energies in Table I. The convergence of conventional MP2 is very slow and the aug-cc-pV6Z value is still more than $0.01 a_0$ underbound. As expected, MP2-R12 converges significantly faster than conventional MP2, but the new MP2-F12(1) is a great improvement over MP2-R12. The basis set limit bond length is estimated to be $5.785 a_0$, with a vibrational frequency of approximately 28 cm^{-1}. This is in good agreement with the values of $5.791 a_0$ and 26.5 cm^{-1} obtained by van Mourik *et al.* using conventional MP2 and a t-aug-cc-pV6Z basis.[49]

We are interested in the minimum basis required to obtain an equilibrium bond length prediction to within $0.02 a_0$; we must therefore examine the various sources of deviation from the basis set limit and the bias that they introduce. The energy bias required to shift the minimum by $\pm 0.02 a_0$ is ± 0.2 $\mu E_{\mathrm{h}} a_0^{-1}$.

Let us first consider the HF contribution. Taking the aug-cc-pV6Z basis as a reference, the aug-cc-pVTZ values bias the potential curve by -0.2 $\mu E_{\mathrm{h}} a_0^{-1}$, which shortens the equilibrium bond length by $0.003 a_0$. The bias remaining when an aug-cc-pVQZ basis is used lengthens the equilibrium bond by $0.001 a_0$. Clearly the dominant contribution to the position of the minimum is the correlation energy.

In all cases the convergence to the basis set limit bond length is from above, which is consistent with the fact that the correlation contribution to the binding energy is underestimated near the equilibrium for these finite basis sets, as discussed in Sec. IV A. The HF contribution is repulsive and decreases in magnitude as the bond length r increases. At the equilibrium, the linear dependence of the correlation contribution on r exactly cancels the gradient of the HF contribution and introduces quadratic terms in r resulting in the minimum. If the same fraction of the correlation contribution is

TABLE V. Computed MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ equilibrium bond distances of Be$_2$ in a_0.

X	MP2	R12	F12(2) γ			F12(1) γ		
			0.2	0.4	0.6	0.5	1.0	1.5
T	5.33	5.22	5.14	5.17	5.17	5.14	5.15	5.15
Q	5.18	5.12	5.11	5.12	5.12	5.11	5.12	5.12
5	5.14	5.11	5.11	5.11	5.11	5.11	5.11	5.11

recovered for all separations, the HF contribution is not exactly canceled, introducing a bias that shifts the minimum to a longer bond length. This is the dominant factor in the trend in the computed equilibrium bond lengths. Treating the HF and correlation contributions as first and second orders in the bond length r, respectively, we may estimate the fraction α of the correlation energy required to obtain the equilibrium bond length correct to $\Delta = 0.02a_0$ as

$$\alpha = \frac{F}{F + k\Delta}, \tag{8}$$

where $F = -dE_{\text{bind}}^{\text{HF}}/dr$ is evaluated at the equilibrium r_0 and is found to be 45 $\mu E_{\text{h}} a_0^{-1}$. The required percentage of the correlation energy contribution to the binding energy is then estimated to be 97%, which corresponds to a binding energy of 21.2 μE_{h} at equilibrium. There is also an additional bias arising from the variation in the accuracy of the correlation energy with bond length, which can act either to partially compensate for or to accentuate the bias due to the HF repulsion. However, we would expect that such a bias would be small if 97% of the correlation energy is obtained close to equilibrium.

The trends displayed in Table IV are largely explained by the above considerations. From Fig. 1 we see that for all basis sets, the percentage of binding energy recovered using MP2-F12(1) and MP2-F12(2) decreases as γ increases from 0.5 to 1.5 and from 0.2 to 0.4, respectively. In Table IV we see that the computed minima lengthen as γ increases. For MP2-F12(1) more than 97% of the correlation contribution to the binding energy is recovered for $0.5 \leqslant \gamma \leqslant 1.5$ even with the aug-cc-pVTZ basis and the structures are universally of high quality. MP2-F12(2) performs much worse, especially at larger values of γ, as reflected in too long bond lengths. However, comparing the MP2-R12 and MP2-F12(2) values for E_{bind}^{X} in Table I, we see that even though MP2-R12 recovers a larger percentage of the binding energy, MP2-F12(2) with $\gamma = 0.4$ yields equilibrium bond lengths closer to the basis set limit. We observe that the percentage of the correlation contribution to the binding energy recovered with a finite basis decreases slightly for MP2-F12(2) with $\gamma = 0.4$ as the bond length increases, which acts to shorten the computed minimum. This is in contrast to MP2-R12, MP2-F12(1) for all γ, and MP2-F12(2) with $\gamma = 0.6$, where the percentage increases as the separation increases.

The bias that would be added to the potential curve if the RI error were not properly removed is also interesting. In all cases, the bias due to the RI error is only significant for the aug-cc-pVTZ basis. For MP2-F12(1) the RI bias is −0.3, −0.4, and −0.6 $\mu E_{\text{h}} a_0^{-1}$ for $\gamma = 0.5$, 1.0, and 1.5, respectively,

acting to shorten the bond length by almost $0.01a_0$. For MP2-F12(2) the bias is larger and changes sign, being −0.6, 0.3, and 0.8 $\mu E_{\text{h}} a_0^{-1}$ for $\gamma = 0.2$, 0.4, and 0.6, respectively. For MP2-R12 the RI bias is 0.8 $\mu E_{\text{h}} a_0^{-1}$ for the aug-cc-pVTZ basis and acts to lengthen the bond by $0.01a_0$.

We find that MP2-F12(1) is superior to MP2-R12 and MP2-F12(2) in several respects. High quality structures, within $0.02a_0$ of the basis set limit, are obtained with an aug-cc-pVTZ basis and the potential is relatively insensitive to the value of γ chosen. This is because more than 97% of the binding energy is obtained with this level for all values of γ in the computed range. This is in contrast to MP2-F12(2) where both the binding energies and the structures deteriorate significantly as γ increases, when the aug-cc-pVTZ basis is used.

B. Be–Be

We see from Table V that all of the explicitly correlated methods converge to give an equilibrium bond distance of $5.11a_0$. The vibrational frequency is approximately 138 cm^{-1} and if an accuracy of $0.02a_0$ is required, then the bias due to the deviations from the basis set limit must be smaller than 65 $\mu E_{\text{h}} a_0^{-1}$. We report that the aug-cc-pVTZ HF potential differs significantly from the aug-cc-pV5Z curve, introducing a bias of 53 $\mu E_{\text{h}} a_0^{-1}$, which lengthens the minimum by $0.02a_0$. The bias and its consequence are an order of magnitude smaller for the aug-cc-pVQZ basis. The gradient of the repulsive HF potential at $5.11a_0$ is 6590 $\mu E_{\text{h}} a_0^{-1}$.

Using Eq. (8) we estimate that 99% of the correlation contribution to the binding energy must be recovered in order to obtain the equilibrium structure to within $0.02a_0$. This corresponds to a binding energy of 2080 μE_{h} (assuming a HF contribution of −7580 μE_{h}). We see that this is achieved for MP2-R12 and both MP2-F12 methods over the range of γ presented, when an orbital basis of aug-cc-pVQZ quality or better is used. In fact, for the MP2-F12(1) method, this is achieved even with the aug-cc-pVTZ basis and the bond lengths are too long mainly due to the error in the HF contribution.

The trend in Table V is largely explained by the dependence of E_{bind}^{X} on γ displayed in Fig. 2. Since the percentage of the binding energy recovered for a given basis decreases as γ increases, the computed bond lengths increase. We observe that the MP2-F12(1) values increase less than those of MP2-F12(2). This is because for large γ, the computed dimer energy decreases too rapidly as the bond length increases due

TABLE VI. Computed MP2, MP2-R12, and MP2-F12/aug-cc-pVXZ equilibrium bond distances of Ne$_2$ in a_0.

X	MP2	R12	F12(2) γ			F12(1) γ		
			0.2	0.4	0.6	0.5	1.0	1.5
T	6.26	6.14	6.06	6.15	6.18	6.04	6.08	6.30
Q	6.14	6.06	6.04	6.07	6.11	6.03	6.05	$\cdots$
5	6.10	6.06	6.06	6.06	6.08	6.06	6.05	6.05
6	6.08	6.06	6.05	6.05	6.06	6.05	6.05	6.05

to the more contracted geminals capturing less long-range correlation. This adds a bias to the potential that acts to shorten the computed minima.

All three of the explicitly correlated methods provide bond lengths accurate to within $0.02a_0$ when an aug-cc-pVQZ basis is used. The errors arising with the aug-cc-pVTZ are dominated by the unconverged HF contribution. The bias that would be introduced if an uncontracted aug-cc-pVYZ auxiliary basis were used with $Y=X+1$ for an aug-cc-pVXZ orbital basis is completely negligible.

C. Ne–Ne

We see from Table VI that with the aug-cc-pV6Z basis the MP2-F12(1) and MP2-F12(2) methods converge to give an equilibrium bond length of $6.05a_0$ and the vibrational frequency is approximately 23 cm^{-1}. Our basis set limit results are in good agreement with the values of $6.066a_0$ and 22.5 cm^{-1} obtained by van Mourik *et al.* using conventional MP2 and a t-aug-cc-pV6Z basis.[49] The maximum bias due to basis set incompleteness must be below 1.9 $\mu E_h a_0^{-1}$ for structures accurate to within $0.02a_0$ and the HF contribution must be computed with an aug-cc-pV5Z basis or better in order to achieve this accuracy. Taking the aug-cc-V6Z curve as a reference, the incompleteness of the aug-cc-pVTZ and aug-cc-pVQZ basis sets introduces biases of -1.7 and -2.3 $\mu E_h a_0^{-1}$, respectively, which acts to shorten each computed minima by $0.01a_0$. The corresponding bias for the aug-cc-pV5Z basis is an order of magnitude smaller and acts to lengthen the minimum by $0.001a_0$.

The repulsive force of the HF potential is 138 $\mu E_h a_0^{-1}$ and therefore the correlation contribution must be accurate to within 3% to reduce the basis set incompleteness bias to below 3.9 $\mu E_h a_0^{-1}$, which corresponds to a binding energy between 81.2 and 89.8 μE_h. However, since this bias acts to lengthen the computed minima when the binding energy is underestimated, it partially cancels with the bias due to the HF contribution for the aug-cc-pVTZ and aug-cc-pVQZ basis sets. When the binding energy is overestimated, this bias acts to shorten the bond length.

In Table VI we see that for MP2-F12(2) the trend in the computed minima with γ is consistent with the percentage of binding energy recovered, as displayed in Fig. 3. As γ increases, the percentage decreases and the minimum lengthens due to the incompletely removed HF repulsion. For MP2-F12(1) with the aug-cc-pV5Z basis, the opposite trend is observed, because the binding energy is overestimated. For the aug-cc-pVTZ and aug-cc-pVQZ basis sets the geminal BSSE has a large effect on the bond length predictions for

large γ. The percentage of the correlation contribution to the binding energy increases as the bond length increases, which adds a bias to the potential that acts to lengthen the minimum. The value for the aug-cc-pVQZ basis with $\gamma=1.5$ is missing from Table VI because the potential had a discontinuity since an eigenvalue of the **B** matrix that dropped below zero was eliminated.

All three of the explicitly correlated methods provide bond lengths accurate to within $0.02a_0$ when an aug-cc-pV5Z basis is used. The errors arising with the aug-cc-pVTZ and aug-cc-pVQZ basis sets have contributions from the HF and correlation terms, which leads to fortuitously–good structures when they partially cancel. The geminal BSSE causes severe problems for the MP2-F12(1) method with large γ leading to much too long equilibrium bond lengths.

VI. THE ORIGINAL ORBITAL-VARIANT FORMALISM

All of the results presented in the previous sections are computed using ansatz 2 and approximation B in the usual orbital-invariant formalism.[36] From the point of view of pair clusters, the MP2-R12 wave function is represented by the substitution

$$|ij\rangle \rightarrow |ij\rangle + \sum_{a<b} d_{ij}^{ab}|ab\rangle + \sum_{k<l} c_{ij}^{kl}w_{12}|kl\rangle, \qquad (9)$$

where i,j,k,l denote occupied orbitals, a,b virtual orbitals, and w_{12} is defined in Eq. (3). In other words, the adjustment of the orbital pair $|ij\rangle$ due to the electron interaction is represented by the additional sets of orbital pair functions $|ab\rangle$ and $w_{12}|kl\rangle$. In the original MP2-R12 method only the dominant R12 contributions $w_{12}|ij\rangle$ are included, and the formalism is therefore not invariant to unitary transformations of the occupied orbitals.

In most current applications, the orbital-variant formalism has been rejected in favor of the orbital-invariant formalism. There are two main reasons for this: Firstly, when there are degeneracies between the molecular orbitals, as in methane, for example, it is undesirable that the MP2-R12 energy should depend on the specific representation of the degenerate orbitals; secondly, the approximate way in which matrix elements such as $\langle \chi_A \chi_B | w_{12}^\dagger \hat{f}_{12} w_{12} | \chi_A \chi_B \rangle$ contribute to the pair energies leads to erroneous long-range interactions between monomer fragments A and B.

In our MP2-F12 calculations on He$_2$ and Ne$_2$ using the orbital-invariant method we encounter difficulties arising from geminal BSSE. In Sec. VI A we examine the geminal BSSE in detail and show that the primary deficiencies are

removed if we revert to the original orbital-variant formalism and use localized orbitals. In Sec. VI B we examine the effect of the long-range behavior of linear and nonlinear correlation factors by comparing the orbital-variant and invariant methods. In order to simplify our discussion in the following sections we restrict our calculations to approximation A, where the $\mathbf{B}^{(ij)}$ matrix is independent of the orbital pair (ij), but the results are indicative of both approximations A and B.

A. Removing geminal BSSE

All R12 and F12 methods utilize geminal functions constructed from a correlation factor multiplied by a pair of occupied orbitals. Since, for a monomer with n occupied orbitals, there are $n(n+1)/2$ singlet and $n(n-1)/2$ triplet geminals in the monomer basis and $n(2n+1)$ singlet and $n(2n-1)$ triplet geminals in the dimer basis, geminal BSSE is always present in our binding energies. The method of counterpoise correction, used to reduce the orbital BSSE, involves increasing the number of basis functions in the calculation of the monomer energy to match that of the dimer. Although this is possible in R12 methods, it would involve using the full set of dimer geminals in the monomer calculation and is not an attractive solution. The alternative is to reduce the basis of the dimer to match the monomer. Localizing the HF orbitals to monomer centers A and B, we can remove the interactions between the geminals centered on A with those centered on B. Likewise, we can remove the interactions between the mixed AB geminals with those centered on A and B, but we should not remove the interactions of the set of mixed geminals with themselves since they may be considered to belong exclusively to the dimer. The cleanest way to achieve this is to use the original orbital-variant method with localized orbitals.

Now let us examine the effect of geminal BSSE on the binding energy of our three vdW dimers. We transform to local orbitals by symmetry, simply taking plus and minus combinations of the bonding and antibonding pairs of canonical orbitals.[50] Since the overlap between orbitals on A with orbitals on B is not zero, the orbital localized on A
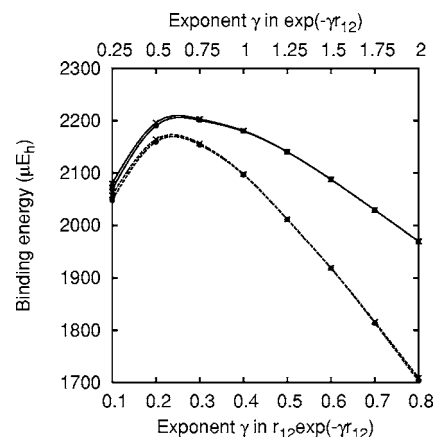

FIG. 5. The orbital-invariant ($\times$) and local-orbital-variant ($\bullet$) MP2-F12/2A binding energies of Be$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) as a function of γ, with the aug-cc-pVTZ basis.

contains a tail on B and vice versa. We distinguish the localized orbitals from atomic orbitals with a tilde. For both He$_2$ and Be$_2$ (frozen core) the transformation to local orbitals only affects the singlet pair interactions and removing the interactions of geminals $w_{12}|\tilde{s}_A\tilde{s}_A\rangle$ with $w_{12}|\tilde{s}_B\tilde{s}_B\rangle$ and $w_{12}|\tilde{s}_A\tilde{s}_B\rangle$ with $w_{12}|\tilde{s}_A\tilde{s}_A\rangle$ and $w_{12}|\tilde{s}_B\tilde{s}_B\rangle$ is entirely equivalent to the orbital-variant method in the basis of localized orbitals. For the binding energy of Ne$_2$ the contributions to the dimer from both the singlet and triplet geminal interactions contributes to BSSE. For Ne$_2$ the orbital-variant method using localized orbitals not only removes the terms contributing to BSSE but also many geminal interactions that act to improve the correlated description of both the atom and the dimer. In Figs. 4–6 we present the results of MP2-F12/2A calculations for the localized orbital-invariant and orbital-variant methods for both correlation factors (1) and (2), using the aug-cc-pVTZ basis sets for various values of the exponent γ. The effect of geminal BSSE is naturally much smaller if larger orbital basis sets are employed.

From Fig. 4 we see that the geminal BSSE has essentially zero effect on the MP2-F12 binding energy of He$_2$ for values of γ below 1.0 or 0.4 for correlation factors (1) and
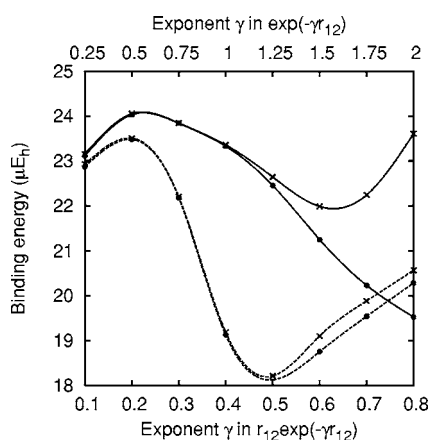

FIG. 4. The orbital-invariant ($\times$) and local-orbital-variant ($\bullet$) MP2-F12/2A binding energies of He$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) as a function of γ, with the aug-cc-pVTZ basis.
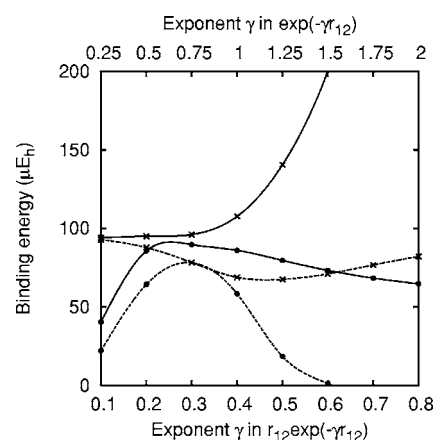

FIG. 6. The orbital-invariant ($\times$) and local-orbital-variant ($\bullet$) MP2-F12/2A binding energies of Ne$_2$ computed using correlation factors $\exp(-\gamma r_{12})$ (full line) and $r_{12}\exp(-\gamma r_{12})$ (dashed line) as a function of γ, with the aug-cc-pVTZ basis.

(2), respectively. For the orbital-invariant MP2-F12(1)/2A method, the binding energy increases for larger values of γ in the same way as observed for approximation 2B in Sec. IV A. Removing the interactions between the dimer geminal basis functions that contribute to BSSE removes this spurious behavior and the energy of the dimer deteriorates faster than that of the atom as γ increases, as one would expect. An analysis of the two kinds of interactions that are removed shows that the dominant contributions arise from excitations of the kind $w_{12}|\tilde{s}_A\tilde{s}_A\rangle \rightarrow w_{12}|\tilde{s}_A\tilde{s}_B\rangle$, indicating a deficiency in the orbital basis for describing long-range correlation. The fact that this is relatively unimportant for MP2-F12(2) and for small values of γ for MP2-F12(1) indicates that in these cases the orbital-variant geminal basis sufficiently captures the long-range effects. The geminal BSSE for the MP2-R12 method is 0.3 μE_h.

From Fig. 5 we see that the effect of geminal BSSE on the binding energy of Be_2 is negligible compared to the greater sensitivity of the energy of the dimer to γ over that of the atom. In contrast to He_2 the geminal BSSE decreases as γ increases and has equal contributions from both kinds of mixed center excitations. This is consistent with the observation that the aug-cc-pVTZ orbital basis is far from complete in all respects, not just in the space of very diffuse functions. We argue that smaller values of γ admit larger couplings between the geminal functions on different centers, contributing to the short-range correlation of the dimer. This prognosis is supported by the fact that the geminal BSSE is 10 μE_h for MP2-F12(1) and MP2-F12(2) with $\gamma=0.25$ and $\gamma=0.10$, respectively, and that the value of 44 μE_h for the MP2-R12 method is much larger.

From Fig. 6 we see that the local-orbital-variant values for the binding energy of Ne_2 are significantly different from the orbital-invariant energies, for both large and small values of γ. The reduction of the geminal BSSE removes the gross overbinding of the MP2-F12(1) method for large values of γ, caused by similar interactions to those described for He_2. However, since only the interactions of each geminal basis function with itself are taken into account, the descriptions of both the atom and the dimer deteriorate. The atom is described more accurately than the dimer, as one would expect, leading to underestimated binding energies. MP2-F12(1) then performs much better than MP2-F12(2), consistent with the studies of He_2 and Be_2. In sec. IV C we observe that MP2-R12 performs much better than either MP2-F12 methods using the orbital-invariant formalism. This situation is reversed for the local-orbital-variant method and MP2-R12 does not predict binding, when the aug-cc-pVTZ basis is employed.

B. Long-range vdW interactions

The fact that linear r_{12} correlation factor increases to infinity with the interelectronic distance has caused considerable concern over the computation of long-range interactions using R12 methodologies. The original orbital-variant MP2-R12 method was observed not to predict binding for He_2 and the potential energy curve for Be_2 gave a repulsion of 400 μE_h at large bond lengths.[36] These failures are asso-

ciated with spuriously large contributions to the **B** matrix and are rectified if we use either the orbital-invariant method[36] or the orbital-variant method with local orbitals.[50] The new correlation factors (1) and (2) decay to zero at long range and these problems are not expected to arise for the original orbital-variant method.

Let us first review the performance of the MP2-R12 method for the equivalent situations of He_2 and Be_2 (frozen core), comparing the orbital-invariant method with the orbital-variant method with local and canonical orbitals. The energy of the monomer is the same for each method since there is only one geminal function. The triplet pair contribution to the energy of the dimer is similarly unaffected by the change in method and we need only consider the singlet pair energy contributions to the dimer. There are three singlet geminal functions and we have the following relations between the canonical and local representations:

$$w_{12}|\sigma_g\sigma_g\rangle = \tfrac{1}{2}w_{12}(|\tilde{s}_A\tilde{s}_A\rangle + \sqrt{2}|\tilde{s}_A\tilde{s}_B\rangle + |\tilde{s}_B\tilde{s}_B\rangle), \tag{10}$$

$$w_{12}|\sigma_g\sigma_u\rangle = \tfrac{1}{2}w_{12}(\sqrt{2}|\tilde{s}_A\tilde{s}_A\rangle - \sqrt{2}|\tilde{s}_B\tilde{s}_B\rangle), \tag{11}$$

$$w_{12}|\sigma_u\sigma_u\rangle = \tfrac{1}{2}w_{12}(|\tilde{s}_A\tilde{s}_A\rangle - \sqrt{2}|\tilde{s}_A\tilde{s}_B\rangle + |\tilde{s}_B\tilde{s}_B\rangle). \tag{12}$$

We remind the reader that $|\tilde{s}_A\tilde{s}_B\rangle$ denotes the symmetrized pair function. The R12/2A contribution to the MP2 energy is given by

$$E_{R12}^{(2)} = -\operatorname{Tr}(\mathbf{V}^T\mathbf{B}^{-1}\mathbf{V}) \tag{13}$$

where the matrices **B** and **V** are defined by

$$B_{kl,mn} = \delta_k^m\delta_l^n - \frac{1}{2}\sum_{iq'}(r_{kl}^{iq'}t_{iq'}^{mn} + r_{lk}^{iq'}t_{iq'}^{nm} + r_{mn}^{iq'}t_{iq'}^{kl} + r_{nm}^{iq'}t_{iq'}^{lk})$$

$$+ \frac{1}{2}\sum_{ij}(r_{kl}^{ij}g_{ij}^{mn} + r_{mn}^{ij}g_{ij}^{kl}), \tag{14}$$

$$V_{kl,ij} = \delta_k^i\delta_l^j - \sum_{mq'}(r_{kl}^{mq'}g_{mq'}^{ij} + r_{lk}^{mq'}g_{mq'}^{ji}) + \sum_{mn}r_{kl}^{mn}g_{mn}^{ij}, \tag{15}$$

where i, j, k, l, m, n are occupied orbitals and q' runs over the auxiliary basis set and

$$r_{pq}^{rs} = \langle\varphi_p(1)\varphi_p(2)|r_{12}|\varphi_r(1)\varphi_s(2)\rangle, \tag{16}$$

$$g_{pq}^{rs} = \langle\varphi_p(1)\varphi_p(2)|r_{12}^{-1}|\varphi_r(1)\varphi_s(2)\rangle, \tag{17}$$

$$t_{pq}^{rs} = \langle\varphi_p(1)\varphi_p(2)|[\hat{T}_1 + \hat{T}_2, r_{12}]|\varphi_r(1)\varphi_s(2)\rangle. \tag{18}$$

The original orbital-variant formalism is equivalent to setting the off-diagonal elements of the **B** and **V** matrices to zero. At large separations we may neglect the overlap between functions on A with functions on B and therefore drop the tilde and assume that $B_{AA,AA}=B_{BB,BB}$ and $V_{AA,AA}=V_{BB,BB}$. In the local-orbital representation, the **V** and **B** matrices are then diagonal, containing only the interactions of each of $w_{12}|s_As_A\rangle$, $w_{12}|s_As_B\rangle$ and $w_{12}|s_Bs_B\rangle$ with themselves. Within this approximation, the local-orbital-variant and orbital-invariant methods give identical results,
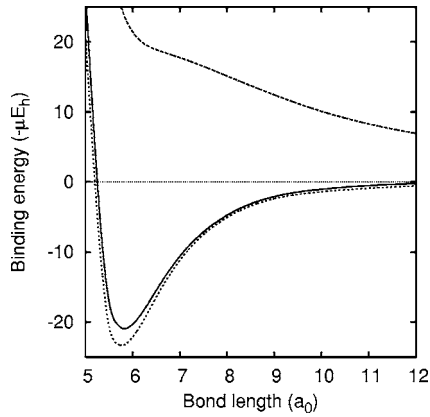
FIG. 7. The orbital-invariant and local-orbital-variant MP2-R12/2A (dashed line), canonical-orbital-variant MP2-R12/2A (solid line), and orbital-invariant and local-, and canonical-orbital-variant MP2-F12(1)/2A $\gamma=1.0$ (dotted line) potential curves for He₂ using the aug-cc-pVTZ basis.
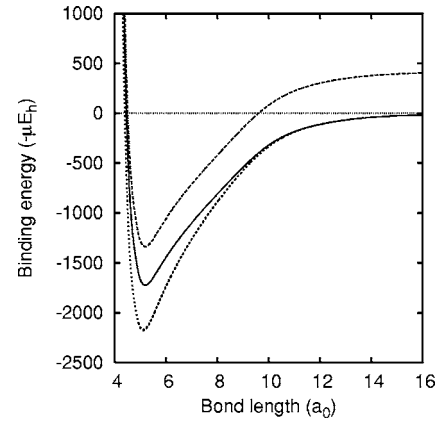

FIG. 8. The orbital-invariant and local- and canonical-orbital-variant MP2-R12/2A (solid line), canonical-orbital-variant MP2-R12/A (dashed line), and orbital-invariant and local- and canonical-orbital-variant MP2-F12(1)/2A $\gamma=1.0$ (dotted line) potential curves for Be₂ using the aug-cc-pVTZ basis.

$$E_{\mathrm{R12}}^{(2)} = -2\frac{V_{AA,AA}^2}{B_{AA,AA}} - \frac{V_{AB,AB}^2}{B_{AB,AB}}. \tag{19}$$

The canonical-orbital-variant method, on the other hand, gives

$$E_{\mathrm{R12}}^{(2)} = -\frac{V_{AA,AA}^2}{B_{AA,AA}} - \frac{(V_{AA,AA}+V_{AB,AB})^2}{B_{AA,AA}+B_{AB,AB}}. \tag{20}$$

When the linear-r_{12} correlation factor is used, the matrix elements r_{pq}^{rs} give spuriously large values for $B_{AB,AB}$ due to the unphysical long-range nature connecting centers A and B. For the orbital-invariant formalism this is largely canceled by small values for $V_{AB,AB}$ due to the vanishing g_{pq}^{rs}. For the original canonical-orbital-variant formalism this is no longer true and $E_{\mathrm{R12}}^{(2)}$ is underestimated, leading to underbinding or even repulsion.

In Fig. 7 we present the potential energy curves for the MP2-R12/2A canonical- and local-orbital-variant methods and the orbital-invariant method using the aug-cc-pVTZ basis with the large $18s10p6d5f4g3h2i$ auxiliary basis. In addition we plot the three corresponding MP2-F12(1)/2A curves using $\gamma=1.0$. We see that for MP2-R12, the orbital-invariant and local-orbital-variant methods indeed give the same curves, but that the canonical-orbital-variant method does not predict binding. It is remarkable that all three curves for the MP2-F12(1) method lie on top of each other. This is because the matrix elements $B_{AB,AB}$ and $V_{AB,AB}$ are at least three orders of magnitude smaller than $B_{AA,AA}$ and $V_{AA,AA}$, respectively, for all bond lengths. This is not the case for smaller values of γ, or for MP2-F12(2), and the canonical-orbital-variant method predicts slightly less binding than the local-orbital-variant method, but they agree at long range, indicating that these matrix elements are slightly overestimated at short bond lengths when too diffuse correlation factors are used. The $B_{AB,AB}$ matrix element computed using MP2-R12/2A, however, is two orders of magnitude too large. For large values of γ the orbital-invariant and local-orbital-variant MP2-F12 methods differ due to geminal BSSE as discussed in the previous section.

In Fig. 8 we present the MP2-R12/2A and MP2-F12(1)/2A ($\gamma=1.0$) orbital-invariant and canonical- and

local-orbital-variant potential energy curves for Be₂ using the aug-cc-pVTZ basis with the large $21s12p7d5f4g3h2i$ auxiliary basis set. We see that for Be₂ all three MP2-R12/2A methods lie on the same curve and that the potential from the original canonical-orbital-variant method is not repulsive at large separations, in contradiction to expectation based on the work of one of us in Ref. 36, where a repulsion of $400\ \mu E_{\mathrm{h}}$ is observed at a separation of $20a_0$. In this early work the MP2-R12/A method is used, where the orbital basis is used for the auxiliary basis and we include the curve for this method using the aug-cc-pVTZ basis in Fig. 8 for reference. The only difference between the MP2-R12/A and MP2-R12/2A methods is the removal of the RI error. We report that the matrix element $B_{AB,AB}$ is two orders of magnitude larger for MP2-R12/A than for MP2-R12/2A, due to the RI error, and this is the cause of the long-range repulsion predicted by the canonical-orbital-variant method. This failure is unimportant for the orbital-invariant method since the $V_{AB,AB}$ are very small and even though the AB contribution is too large, it is has a negligible effect on the potential energy. In contrast to He₂, the long-range nature of the linear r_{12} correlation factor does not result in the incorrect behavior of the MP2-R12 potential for Be₂. It is therefore no surprise that all three MP2-F12(1)/2A methods give the same curves for Be₂. This is true for both correlation factors and all values of γ.

The situation for Ne₂ is more complicated since there are many more geminal basis functions. The orbital-invariant potential energy curves suffer from geminal BSSE, particularly the MP2-F12(1) curves for large γ, and the local-orbital-variant method should be taken as the reference. In the same way as for He₂ and Be₂, the AB contributions to the Ne₂ binding energy are correctly included in the local-orbital-variant method, but incorrectly included in the canonical-orbital-variant method. In Fig. 9 we present the orbital-invariant and local-orbital-variant MP2-R12/2A and the local- and canonical-orbital-variant MP2-F12(1)/2A ($\gamma=1.0$) potential curves for Ne₂ using the aug-cc-pVTZ orbital basis and the large $21s14p8d7f5g4h3i2k$ auxiliary basis. We see that the orbital-invariant MP2-R12/2A curve gives a good description of the vdW interaction, but the local-orbital-variant method does not predict binding. This is
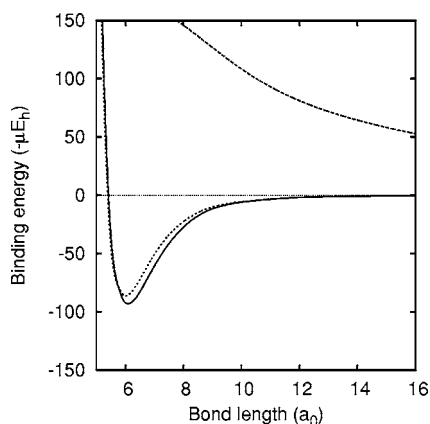
FIG. 9. The orbital-invariant (solid line), local-orbital-variant MP2-R12/2A (dashed line), and local- and canonical-orbital-variant MP2-F12(1)/2A $\gamma=1.0$ (dotted line) potential curves for Ne_2 using the aug-cc-pVTZ basis.

because many of the geminal-geminal interactions that stabilize the dimer, but do not contribute to BSSE, are neglected and the diagonal terms do not sufficiently describe the correlated wave function. It is therefore likely that the geminal BSSE is small for MP2-R12/2A. The canonical-orbital-variant results are not plotted since this method predicts a repulsion of over 1000 μE_h at the equilibrium bond length, due to the spuriously large **B** matrix elements. For MP2-F12(1)/2A, the local- and canonical-orbital-variant curves lie on top of each other, for all values of γ investigated, indicating that, as for He_2 the damped nature of the correlation factor results in a correct description of the vdW interactions since geminals of the type $w_{12}|\chi_A\chi_B\rangle$ do not lead to spuriously large interactions. This is not true for correlation factor (2), where the canonical-orbital-variant method predicts less binding than the local-orbital-variant method.

VII. CONCLUSION

If the conventional MP2 method is used, the computation of the binding energies of the vdW dimers He_2, Be_2, or Ne_2 to within 95% of the basis set limit requires an orbital basis set of aug-cc-pV6Z quality or larger. Even at this level, the error in the correlation energy of He_2 is 1 mE_h and there is a large cancellation of errors to give the accurate value of only 0.8 μE_h below the basis set limit binding energy, 22.4 μE_h. The use of geminal basis functions with an explicit dependence on the interelectronic distance r_{12} greatly reduces the individual errors in the correlation energies of the monomer and dimer. For He_2, the deviation from the basis set limit of the MP2-R12 correlation energy is approximately 50 μE_h when the aug-cc-pV6Z basis is used. The convergence to the basis set limit binding energy is much faster and 95% accuracy is obtained with an aug-cc-pV5Z basis for all three dimers.

The use of geminals with the nonlinear correlation factors $\exp(-\gamma r_{12})$ and $r_{12}\exp(-\gamma r_{12})$ improves the individual descriptions of the dimer and monomer even further, particularly for small basis sets. For Be_2, this leads to greatly improved binding energies over a wide range of the exponent γ. More than 95% of the correlation contribution is obtained with the aug-cc-pVTZ basis, although the aug-cc-pVQZ

level HF contribution is required for an overall 95% accuracy. For the very weakly bound dimers He_2 and Ne_2, the improvement of the nonlinear correlation factors over linear r_{12} is much more dependent on γ. The correlation factor $\exp(-\gamma r_{12})$ generally performs better than $r_{12}\exp(-\gamma r_{12})$ and 95% accuracy can be obtained using an aug-cc-pVQZ basis if an appropriate value of γ is chosen.

The values of γ that give the best binding energies are smaller than the values that yield the maximum correlation energy of the atom and the dimer. Due to the nonlinear structure of the correlation factors, the geminal basis functions act both to improve the description of the short-range correlation, at electron coalescence, and to improve the description of long-range correlation. The geminal basis functions are designed to efficiently describe the region of electron coalescence and, for He_2 and Ne_2, the accurate description of the long-range vdW correlation becomes critical for the binding energies, since the small errors remaining in the short-range correlation largely cancel.

The dual function of the correlation factors leads to complicated γ dependence. In particular, for larger values of γ, the long-range correlation is not well described by a single geminal, but is instead described through interactions between geminals on different centers. This geminal BSSE results in greatly overestimated binding energies if small orbital basis sets with insufficient diffuse functions are used. The cleanest way to remove geminal BSSE is to transform to local orbitals and to use the original orbital-variant formalism, where only the interaction of each geminal with itself is included. Although in general this results in a loss of accuracy for all γ, the dependence of the binding energy on γ is much more stable and using correlation factor $\exp(-r_{12})$ and an aug-cc-pVQZ basis recovers 95% of the basis set limit binding energies of both He_2 and Ne_2, which represents the remarkable accuracies of 1 and 5 μE_h, respectively. In the absence of a more sophisticated method of removing geminal BSSE, we favor the local-orbital-variant method over the orbital-invariant formalism for use with small basis, although the issues concerning degenerate orbitals remain.

The improvement of the nonlinear correlation factors over linear r_{12} for the binding energies is mirrored in the computation of equilibrium bond lengths. With the exception of the effect of geminal BSSE, the percentage of correlation that the geminals recover does not depend critically on bond length and the faster convergence of the correlation energy leads to a more rapid convergence of the minimum in the potential energy curve. The size of basis set required to obtain bond lengths to a given precision depends on the size of the HF repulsive force.

The new nonlinear correlation factors have two further advantages over linear r_{12}. The first is that the size of the auxiliary basis required for the RI approximation is significantly smaller, due to the more compact nature of the integrals being approximated. Indeed, if the CABS approach were not used, the RI error would be the limiting factor for the accuracy of the MP2-R12 calculations. The final advantage is that the nonlinear correlation factors decay to zero at long range and do not violate the asymptotic r_{12}^{-6} behavior of the vdW interactions. Although, in our MP2-R12 culations,

these unphysical terms are small, their approximate inclusion via the canonical-orbital-variant method leads to severe underbinding for He$_2$ and Ne$_2$. This is avoided if these terms are included properly through either using local orbitals or employing the orbital-invariant method. The new correlation factors afford similar difficulties if they are too diffuse, but the correlation factor $\exp(-r_{12})$ yields excellent results for all methods.

The accurate reproduction of rotation-vibration spectra and differential scattering cross sections of vdW complexes requires that the potential energy surface be computed at the CCSD(T) level of correlation treatment, or higher. The fact that basis sets larger than aug-cc-pV5Z are required for accurate energies has severely restricted the range of systems that may be treated. We have demonstrated that, for MP2, the R12 methodology utilizing the new correlation factors is capable of recovering 95% of the binding energy with an aug-cc-pVQZ basis. We expect that a similar behavior will be observed for CCSD(T)-F12, which would greatly extend the range of vdW systems that may be treated accurately. Work is already underway in this direction.

ACKNOWLEDGMENTS

[1] *Quantum Mechanical Prediction of Thermochemical Data*, edited by J. Cioslowski (Kluwer, Dordrecht, The Netherlands, 2001).

[2] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989).

[3] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[4] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).

[5] J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).

[6] A. D. Boese, M. Oren, O. Atasolu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[7] G. A. Petersson and M. Braunstein, J. Chem. Phys. **83**, 5129 (1985).

[8] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery, Jr., J. Chem. Phys. **104**, 2598 (1996).

[9] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. **112**, 6532 (2000).

[10] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[11] M. Musiał, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **116**, 4382 (2002).

[12] G. Chałasiński and M. M. Szcześniak, Chem. Rev. (Washington, D.C.) **100**, 4227 (2000).

[13] E. D. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **109**, 2233 (1998).

[14] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[15] T. van Mourik and T. H. Dunning, Jr., J. Chem. Phys. **111**, 9248 (1999).

[16] R. J. Gdanitz, J. Chem. Phys. **113**, 5145 (2000).

[17] W. Klopper, J. Chem. Phys. **115**, 761 (2001).

[18] T. van Mourik, R. J. Vos, J. H. van Lenthe, and F. B. van Duijneveldt, Int. J. Quantum Chem. **63**, 805 (1997).

[19] J. B. Anderson, J. Chem. Phys. **115**, 4546 (2001).

[20] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).

[21] R. T Pack and W. Byers Brown, J. Chem. Phys. **45**, 556 (1966).

[22] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).

[23] N. C. Handy, Mol. Phys. **23**, 1 (1972).

[24] W. Klopper and W. Kutzelnigg, Chem. Phys. Lett. **134**, 17 (1987).

[25] R. Bukowski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **104**, 3306 (1996).

[26] A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen, and P. R. Taylor, J. Chem. Phys. **111**, 9157 (1999).

[27] W. Klopper, K. L. Bak, P. Jørgensen, J. Olsen, and T. Helgaker, J. Phys. B **32**, R103 (1999).

[28] W. Klopper, Mol. Phys. **99**, 481 (2001).

[29] J. Noga, P. Valiron, and W. Klopper, J. Chem. Phys. **115**, 2022 (2001).

[30] W. Klopper and J. Noga, ChemPhysChem **4**, 32 (2003).

[31] W. Klopper, F. R. Manby, S. Ten-no, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).

[32] W. Klopper and C. C. M. Samson, J. Chem. Phys. **116**, 6297 (2002).

[33] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).

[34] D. P. Tew and W. Klopper, J. Chem. Phys. **123**, 74101 (2005).

[35] S. Ten-no, J. Chem. Phys. **121**, 117 (2004).

[36] W. Klopper, Chem. Phys. Lett. **186**, 583 (1991).

[37] A. J. May, E. Valeev, R. Polly, and F. R. Manby, Phys. Chem. Chem. Phys. **7**, 2710 (2005).

[38] W. Klopper, J. Chem. Phys. **120**, 10890 (2004).

[39] E. F. Valeev, J. Chem. Phys. **395**, 190 (2004).

[40] C. Angeli, K. L. Bak, V. Bakken *et al.*, DALTON, a molecular electronic structure program, Release 2.0, 2005 (http://www.kjemi.uio.no/software/dalton/dalton.html).

[41] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **100**, 2974 (1994).

[42] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[43] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[44] S. Huzinaga and B. Miguel, Chem. Phys. Lett. **175**, 289 (1990).

[45] S. Huzinaga and M. Klobukowski, Chem. Phys. Lett. **212**, 260 (1993).

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

[47] W. Klopper, J. Chem. Phys. **102**, 6168 (1995).

[48] M. Jeziorska, R. Bukowski, W. Cencek, M. Janunski, B. Jezioraki, and K. Szalewicz, Collect. Czech. Chem. Commun. **68**, 463 (2003).

[49] T. van Mourik, A. K. Wilson, and T. H. Dunning, Jr., Mol. Phys. **96**, 529 (1999).

[50] W. Klopper and W. Kutzelnigg, J. Chem. Phys. **94**, 2020 (1991).