FEATURE ARTICLE

Exploration of Basis Set Issues for Calculation of Intermolecular Interactions

Elena Jakubikova, Anthony K. Rappé, and Elliot R. Bernstein*

Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523

Received: March 6, 2006; In Final Form: May 25, 2006

The ab initio calculation of intermolecular interactions requires a large basis set to describe systems with dominant dispersion interaction accurately. This paper focuses on calculation of intermolecular bonding energies of weakly bound systems within the supermolecular method and on issues related to the choice of a basis set for these calculations, in particular size of the basis set, efficiency of 2-electron integral codes, basis set superposition error (BSSE), and the linear dependence of basis functions. In an attempt to find more efficient basis sets for calculations of intermolecular interactions, standard basis sets (10s Huzinaga, 6-311G**, cc-pV6Z), or their parts, are extended (tessellated) by a set of off-centered, s or p functions, symmetrically placed around the nuclei. Standard basis sets (10s Huzinaga, 6-311G**, cc-pVXZ, aug-cc-pVXZ, X = D, T, Q, 5, 6) are also augmented by sets of atom-centered, higher angular momentum functions (p, d, f). The distance from the nucleus of tessellating functions and orbital exponents of tessellating and augmenting functions are optimized with respect to the BSSE-corrected bonding energy at the MP2 or UCCSD level of theory. The two approaches are tested on the model systems with dominant dispersion interactions $^3H_2$, $(CH_4)_2$, and Ne_2, and their efficiency is compared. Both tessellation and augmentation are successful in describing the intermolecular interactions of these model systems, with augmentation being more efficient. Our results draw attention to the linear dependence problems inevitably present in accurate calculations and confirm the need for underlying standard basis sets that provide good descriptions of core and valence electrons for the tessellation and augmentation approaches to be reliable.

I. Introduction

Intermolecular interactions play a crucial role in understanding a variety of phenomena involving solids, liquids and gases and their accurate calculation is one of the major challenges for today's computational chemistry.

The calculation of intermolecular interactions usually focuses on computing the properties of weakly bound systems, including geometry, vibrational modes, and bonding energy. The quality of such calculations is often measured by their ability to describe the bonding energy properly. At the ab initio level, bonding energy can be computed in two ways: (1) directly, as a sum of physically distinct contributions from at least first and second-order perturbation theory calculations;[1] (2) as a difference between the energy of monomers and the energy of the complex. The second approach is also known as the supermolecular method. Both approaches have their strengths and weaknesses and are complementary rather than competitive.[2] The advantage of perturbational methods lies in the fact that the interaction energy is calculated directly and not as a difference of two large, almost identical numbers. Also, perturbational methods are free from the basis set superposition error (BSSE), which is a major problem in the application of the supermolecular method. Despite this, the vast majority of calculations use the supermolecular method because it is very simple and straightforward and many standard quantum chemistry programs can be employed in the calculations using this method. The supermolecular method is, unlike perturbation approaches, also valid for any distance between the subsystems and higher order terms with respect to the interaction potential are implicitly taken into account.

The main focus of this paper is computation of total bonding energy for systems with dominant dispersion interaction (i.e., we are not interested in computing different components of bonding energy such as induction or dispersion, but in the bonding energy itself). Bonding energy is, unlike its components, a quantity of direct chemical interest as it is an experimentally accessible number. The supermolecular approach is used in all our calculations due to its simplicity as well as widespread use, and therefore further discussion will refer to the supermolecular approach only.

Once a decision is made on which method to use to compute the bonding energy (supermolecular or perturbational), one needs to determine the appropriate theory level and basis set. Some ab initio methods are more suitable for calculation of intermolecular interactions than others. For example, Hartree−Fock (HF) calculations completely miss the dispersion interaction, which involves correlation between the electrons on different molecules. Current density functional methods (DFT) also fail to account for dispersion.[3−5] HF and DFT methods can be more successfully used for computations on systems in which charge transfer or electrostatic interactions are dominant (i.e., hydrogen bonded systems). To describe systems for which dispersion plays an important role, methods that treat electron

10.1021/jp0680239 CCC: $33.50
Published on Web 07/19/2006

Elena Jakubikova received an M.S. degree in Physics from Comenius University (Bratislava, Slovakia) in 2000 and in Mathematics from Colorado State University in 2005. She is currently pursuing a Ph.D. in Chemistry at Colorado State University, working under the direction of Professors Bernstein and Rappé. Her research focuses on calculation of intermolecular interactions and metal oxide catalysis.

Anthony K. Rappé received his B.S. from the University of Puget Sound in 1974. He earned his doctorate from Caltech in 1981. Following postdoctoral work, Dr. Rappé joined Colorado State University. Dr. Rappé has used ab initio electronic structure and empirical force field technologies to study a number of organometallic reactions including nitrogen activation, deNOx catalysis, zeolite-based cracking, olefin metathesis and polymerization, as well as hydrocarbon oxidation. The Rappé group has also worked on the development of force field technology (functional forms and parameters) for the entire periodic table as well as for chemical reactivity studies. This work has resulted in the Universal Force Field (UFF), the Reaction Force Field (RFF), and the first generation Approximate Pair Theory (APT). Dr. Rappé has collaborated with chemists at the Big Three petrochemical companies and has served as a consultant for a number of other chemical firms.

Elliot R. Bernstein received an A.B. degree from Princeton University (1963) in Chemistry and M.S. and Ph.D. degrees from Caltech with a major in Chemistry and a minor in Physics (1967). From 1967 to 1969 he was a postdoctoral fellow at the Enrico Femi Institute and the Chemistry Department of the University of Chicago. He joined the faculty of Princeton University in 1969 and in 1975 moved to Colorado State University where he is currently a Professor. His research interests include intermolecular interaction in gas-phase clusters, spectroscopy of gas-phase radicals, reactions of model catalytic systems containing metal and metal oxide clusters, and ultrafast decomposition of energetic materials.

correlation at a higher level (e.g., configuration interaction, coupled-cluster, or perturbational methods) are needed.

Most of the calculations of intermolecular interactions suffer from the basis set superposition error (BSSE). This error was first noted in the calculation of interaction of two ground-state helium atoms.[6,7] The basis functions of each monomer in the supermolecular calculation are usually finite and far from complete; therefore they use basis functions of the other monomer to improve their energies. The lowering of monomer energy lowers the total energy of the dimer; however, it has nothing to do with the interaction energy one is trying to calculate—it is a mere mathematical artifact. This artificial lowering of the energy can be significantly large—of the order of the interaction energy itself.

There has been a considerable amount of discussion about the BSSE and the ways to correct it (see refs 8−12 and references therein). The most widely accepted is the counterpoise correction scheme,[13,14] which is also used in our calculations.

Another problem associated with calculations of intermolecular interactions is their remarkable dependence on the quality of the basis set employed. One can argue that to obtain reliable results, large basis sets of cc-pVTZ quality or better should be employed in the calculations.[15,16] Extensive polarization functions (i.e., d shell for the first row atoms and p shell for hydrogen) and diffuse functions must be included to describe weakly bonded systems with reasonable accuracy.[17−19] Many researchers also use specially tailored basis sets designed to reproduce monomer properties, such as polarizibilities,[2,20−25] relevant to intermolecular forces. Some also augment standard basis sets with functions optimized with respect to the bonding energies[26] or construct new interaction optimized basis sets.[27] In addition, "bond-centered" basis functions located near the midpoint of the van der Waals bond are also found effective in recovering most of the interaction energy, even in the absence of basis functions with excessively large angular momentum quantum numbers.[28−31]

In 2000, Rappé and Bernstein proposed to supplement standard basis sets with tessellated spherical Gaussians (TSG) optimized for the calculation of intermolecular interactions.[14] TSG functions consist of s basis functions placed at the vertexes, faces or edges of an octahedron centered at the atom's nucleus, thus offering a better description of electron density away from the nucleus as well as providing the higher angular momenta components to the basis sets important for the proper description of dispersion electron correlation events. Another reason behind introducing TSG was a general belief that 2-electron integrals over s basis functions are simpler and therefore faster to evaluate than integrals over higher angular momentum functions. Thereby, the extensive use of TSG should lead to substantial computational savings.

Use of basis functions not centered on nuclei but rather allowed to "float" in space for computation of molecular properties is an idea that dates back to 1950s, when Neumark and Kimball first introduced a "free-cloud" approximation to molecular orbital calculations.[32,33] Their model has been extended by Frost, who employed floating spherical Gaussian orbitals (FSGO) to represent a pair of electrons in a molecule, with the radius and position of FSGO optimized to achieve minimum energy.[34] FSGO were later used in a similar manner by Archibald et al.,[35] Spangler et al.,[36] and more recently by Pakiari.[37,38] A different approach was adopted by Whitten, who instead of explicitly using functions with higher angular momentum (p, d, f, ...), reconstructed them from linear combinations of off-centered Gaussian s-functions ("lobe-functions").[39,40] Bond functions mentioned earlier are another example of the use of the off-centered functions for the ab initio calculations.

The aim of this work is to explore the viability of the tessellation approach for the calculation of intermolecular interactions and compare it with a more standard approach, augmentation. Other issues pertaining to the calculation of intermolecular interactions, such as BSSE and linear dependence of the basis set, are also explored and discussed.

II. Method

Tessellated basis sets are composed of a standard, nucleus centered, valence basis set (or a part of the set) and a set of s or p functions centered away from the nuclei. To place these functions around the nuclei, we use three different tessellation patterns or "shells". The first shell is created by placing s or p functions at the vertexes of an octahedron centered at the atom's nucleus; we call this shell the "v" shell. The second shell of functions, the "f" shell, is created by placing functions at the centers of the eight triangular faces of the octahedron. The third, "e" shell, places twelve functions in the middle of the octahedron's edges. Moreover, there can be more than one function centered at the same point of space. See Figure 1 for a graphical description of the shells.

Each shell of functions is characterized by its radial displacement R from the nucleus and an orbital exponent χ. Both of these parameters are variationally hand-optimized to maximize the BSSE-corrected intermolecular bonding energy.

Augmented basis sets are created by adding sets of p, d, or f functions centered on nuclei to the standard basis sets. Their orbital exponents χ are hand-optimized with respect to the BSSE-corrected bonding energy.

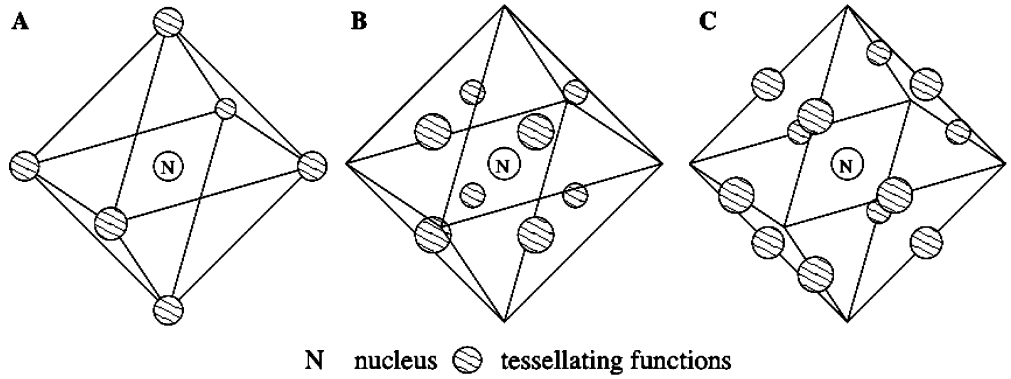

Figure 1. Tessellation shells: A, v-shell, functions placed in the vertexes of octahedron; B, f-shell, functions placed in the faces of octahedron; C, e-shell, functions placed in the middle of the octahedron's edges.

Geometries of dimers are kept fixed during the basis set optimization and are shown in Figure 2. Bond distance for $^3H_2$ is 4.15 Å. The distance between Ne atoms in the dimer is 3.15 Å. Geometry of $(CH_4)_2$ is optimized at the MP2/cc-pVTZ level of theory.

The optimization is done at the MP2 level of theory for Ne_2 and $(CH_4)_2$ and at CISD (UCCSD) level of theory for $^3H_2$. The GAUSSIAN 98 program[41] and MOLPRO 2000.1 and 2002.6[42] are used to perform all computations. Bonding energies are calculated using the supermolecular approach and the counterpoise method is applied to account for BSSE.[13]

The following naming scheme is introduced for the basis sets we developed. The main part of the name consists of the standard basis set used for description of core and valence (i.e., 6-311G**, Huz for a modified 10s Huzinaga basis set (see Table 1), and spd-cc-pV6Z for s, p, and d functions from cc-pV6Z basis set). The prefix stands for the type of functions added to the standard basis set: T for tessellated and A for augmented functions. The suffix describes the angular momentum of functions added: S, P, or D for adding a set of s, p or d functions. In the case of tessellated basis sets, the subscript further gives the tessellation pattern (v, f, or e shell) and the number of functions centered at each position. For example, T6-311G**$S_v^2{}_f$ describes a 6-311G** basis set tessellated with two sets of s functions centered at the vertexes of an octahedron (both sets centered at the same position) and one set of s functions centered in the faces of octahedron. AHuzP^2D is modified 10s Huzinaga basis set augmented with two sets of p functions and one set of d functions.

Finally, it is important to note that all calculations are done with basis sets consisting of spherical Gaussians (i.e., we use 5 functions for the d shell, 7 for the f shell, etc.), which is a default choice in the MOLPRO program.

III. Results

$^3H_2$. We have extended (i.e., tessellated and augmented) two standard basis sets for $^3H_2$: the modified 10s Huzinaga[43] and 6-311G** basis set.[44] The radial distance from the nucleus for each tessellation shell and orbital exponents for tessellating, as well as augmenting functions, are optimized at the UCCSD level of theory with respect to the BSSE-corrected bonding energy. Note that in the case of the modified 10s Huzinaga basis set BSSE is of no concern, because the energy of the hydrogen atom computed with this basis set is 0.499 999 3 hartree, which suggests a maximal BSSE of 1.4 μhartrees [1 μhartree = 2.7211 $\times$ 10^{-5} eV = 0.2195 cm^{-1} = 2.6255 $\times$ 10^{-3} kJ/mol = 6.2751 $\times$ 10^{-4} kcal/mol]. BSSE-corrected bonding energies and BSSE for each optimized basis set are shown in Figures 3−6.

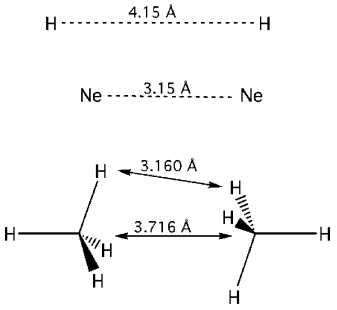

Figure 2. Geometries of dimers used in study: $^3H_2$, Ne_2, and $(CH_4)_2$.

TABLE 1: Exponents and Contraction Coefficients for the Modified 10s Huzinaga Basis Set Used for the Hydrogen Atom

exponents	contraction coefficients
1776.77556	0.44000×10^{-4}
254.017712	0.37200×10^{-3}
54.6980390	0.20940×10^{-2}
15.0183440	0.88630×10^{-2}
4.91507800	0.30540×10^{-1}
1.79492400	0.90342×10^{-1}
0.71071600	0.213239
0.30480200	0.352350
0.13804600	0.339657
0.06215700	0.107330
1.79492400	1.0
0.71071600	1.0
0.30480200	1.0
0.13804600	1.0
0.06215700	1.0
0.0310800	1.0
0.0155400	1.0
0.0077700	1.0

Both tessellated and augmented functions are able to recover substantial amounts of the accepted bonding energy for $^3H_2$, −19.52 μhartrees.[45] Tessellated 10s Huzinaga basis sets recover between 59.2 and 100.4% of the accepted bonding energy, whereas tessellated 6-311G** recover 74.5−105.7% of accepted bonding energy. $^3H_2$ bonding energy computed with augmented 10s Huzinaga and augmented 6-311G** basis sets lies between 16.5 and 79.1% and between 55 and 92.1% of accepted bonding energy, respectively. In the case of 10s Huzinaga and 6-311G** basis sets, higher numbers of functions optimized for bonding recover larger amounts of bonding energy irrespective of whether they are augmenting or tessellating the basis set. Interestingly, bonding energies computed with extended (tessellated or augmented) 6-311G** basis sets recover larger amounts of the bonding energy compared to the 10s Huzinaga basis set tessellated or augmented with the same type and number of functions.

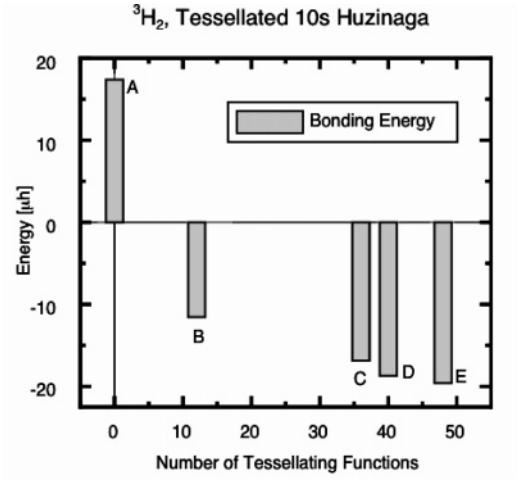

Figure 3. Bonding energy for $^3H_2$ computed with the tessellated modified 10s Huzinaga basis sets. Radial distance from nucleus and the orbital coefficients for tessellating functions are optimized with respect to the $^3H_2$ bonding energy. The letter labels indicate different basis sets: A, modified 10s Huzinaga basis set, no tessellation; B, $THuzS_v$; C, $THuzP_v$; D, $THuzS_v{}^2{}_f$; E, $THuz(SP)_v$.

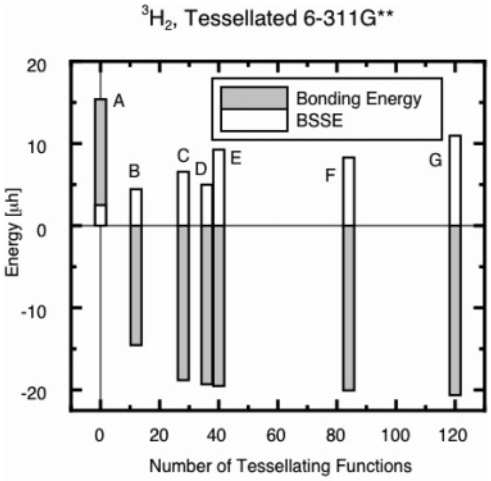

Figure 5. BSSE-corrected bonding energy and BSSE for $^3H_2$ computed with tessellated 6-311G** basis sets. Radial distance from nucleus and the orbital coefficients for tessellating functions are optimized with respect to $^3H_2$ BSSE-corrected bonding energy. The letter labels indicate different basis sets: A, 6-311G** basis set, no tessellation; B, T6-311G**S_v; C, T6-311G**S_{vf}; D, T6-311G**P_v; E, $THuzS_v{}^2{}_f$; F, T6-311G**P_{vf}; G, T6-311G**$P_v{}^2{}_f$.

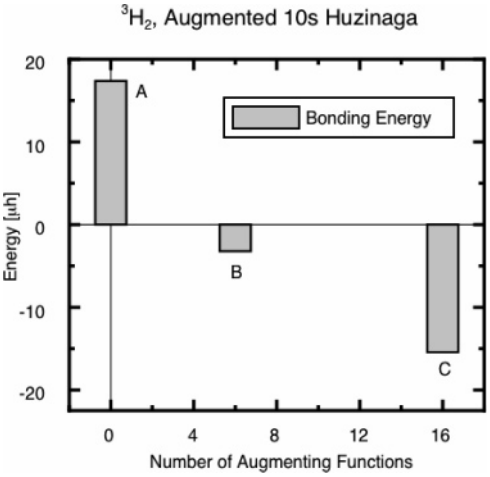

Figure 4. Bonding energy for $^3H_2$ computed with the augmented modified 10s Huzinaga basis sets. Orbital coefficients for augmenting functions are optimized with respect to the $^3H_2$ bonding energy. The letter labels indicate different basis sets: A, modified 10s Huzinaga basis set, no augmentation; B, AHuzP; C, AHuzPD.

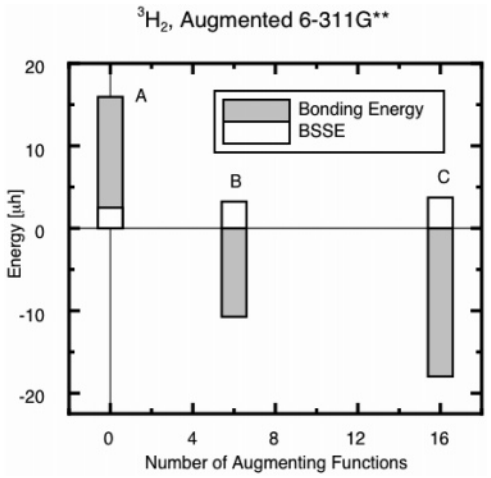

Figure 6. BSSE-corrected bonding energy and BSSE for $^3H_2$ computed with augmented 6-311G** basis sets. Orbital coefficients for augmenting functions are optimized with respect to $^3H_2$ BSSE-corrected bonding energy. The letter labels indicate different basis sets: A, 6-311G** basis set, no augmentation; B, A6-311G**P; C, A6-311G**PD.

Ne_2. We have extended (tessellated or augmented) s, p and d functions from the cc-pV6Z[46] and 6-311G** basis sets for Ne_2. All parameters are optimized at the MP2 level of theory with respect to the BSSE-corrected bonding energy. The BSSE-corrected bonding energy and BSSE for each optimized basis set are shown in Figures 7−10.

As in the case of $^3H_2$, both tessellated and augmented basis sets are able to recover substantial amounts of the Ne_2 accepted bonding energy (−134 μhartrees).[47] Tessellated spd-cc-pV6Z basis sets recover 38.9−61.1% of the accepted bonding energy, whereas tessellated 6-311G** basis sets recover 19.3−75.6% of the Ne_2 bonding energy. Augmented basis sets recover between 33.6% and 52.6% of the Ne_2 accepted bonding energy for spd-cc-pV6Z and between 51.4 and 72.44% for 6-311G** basis sets. Again, with one exception, the higher number of tessellating or augmenting functions recovers more bonding energy, irrespective of the type of the extending function or method employed. Tessellated and augmented 6-311G** basis sets have substantially larger BSSE and recover more bonding energy per shell than tessellated and augmented spd-cc-pV6Z basis sets.

Note that the best available calculation for Ne_2 at the MP2 level of theory employing the t-aug-cc-pVQZ′ basis set of Woon[16] recovers only 64% of the accepted bonding energy. To approach Ne_2 bonding energy more closely, one needs to use the MP4 level of theory. Indeed, the MP4 calculation with selected basis sets reveals further lowering of the bonding energy: MP4/Tspd-cc-pV6ZP_v lowers the bonding energy to −108.08 μhartrees from −72.05 μhartrees at the MP2 level of theory and MP4/T6-311G**P_v lowers the bonding energy from −98.36 to −120.29 μhartrees. Augmented basis sets behave in a similar manner with respect to the MP4 calculations. This shows the importance of using levels of theory higher than MP2 to obtain the correct description of the neon dimer interaction.

The BSSE-corrected bonding energy computed with 6-311G** basis sets extended by more than 20 primitives at the MP2 level of theory lies between −93 and −101 μhartrees (70−76.5% of

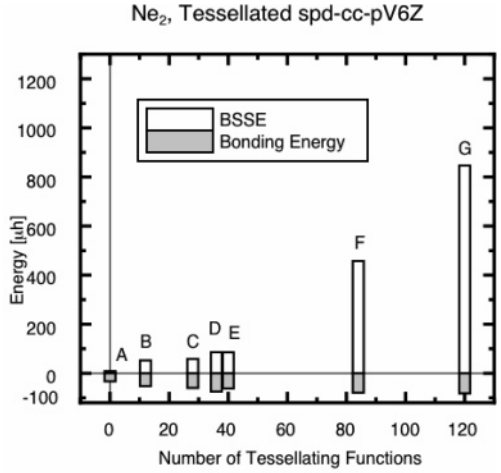

Figure 7. BSSE-corrected bonding energy and BSSE for Ne_2 computed with tessellated spd-cc-pV6Z basis sets. Radial distance from nucleus and the orbital coefficients for tessellating functions are optimized with respect to the Ne_2 BSSE-corrected bonding energy. The letter labels indicate different basis sets: A, spd-cc-pV6Z basis set, no tessellation; B, Tspd-cc-pV6ZS_v; C, Tspd-cc-pV6ZS_{vf}; D, Tspd-cc-pV6ZP_v; E, Tspd-cc-pV6Z$S_v^2{}_f$; F, Tspd-cc-pV6ZP_{vf}; G, Tspd-cc-pV6Z$P_v^2{}_f$.

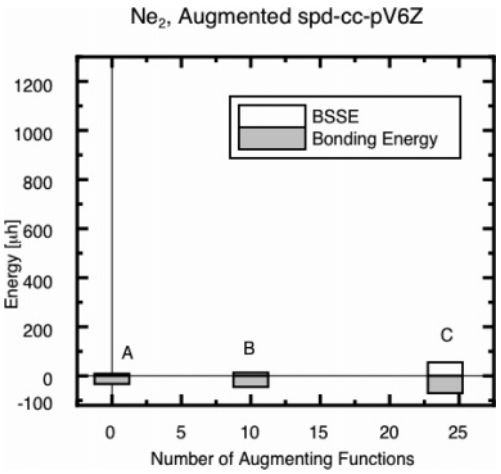

Figure 8. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the augmented spd-cc-pV6Z basis sets. Orbital coefficients for augmenting functions are optimized with respect to the Ne_2 bonding energy. The letter labels indicate different basis sets: A, spd-cc-pV6Z basis set, no augmentation; B, Aspd-cc-pV6ZP; C, Aspd-cc-pV6ZPD.

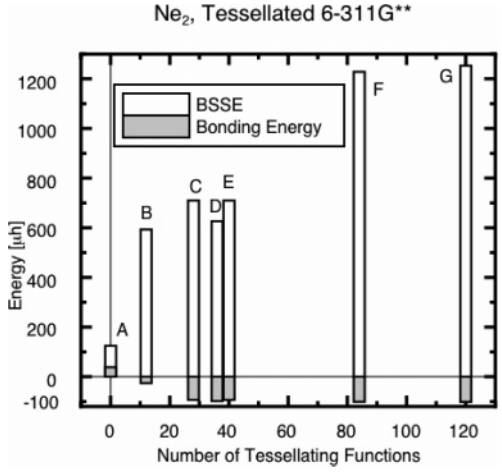

Figure 9. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the tessellated 6-311G** basis sets. Radial distance from nucleus and the orbital coefficients for tessellating functions are optimized with respect to Ne_2 BSSE-corrected bonding energy. The letter labels indicate different basis sets: A, 6-311G** basis set, no tessellation; B, T6-311G**S_v; C, T6-311G**S_{vf}; D, T6-311G**P_v; E, THuz$S_v^2{}_f$; F, T6-311G**P_{vf}; G, T6-311G**$P_v^2{}_f$.

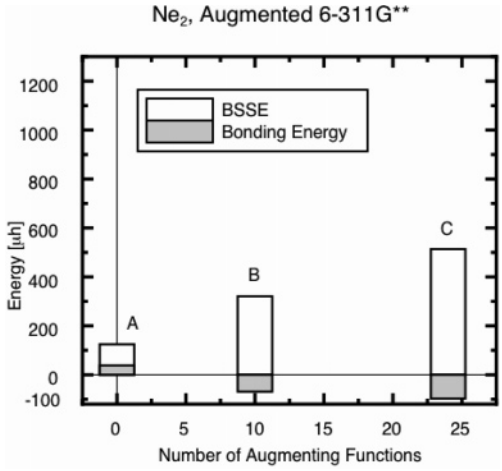

Figure 10. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the augmented 6-311G** basis sets. Orbital coefficients for augmenting functions are optimized with respect to Ne_2 BSSE-corrected bonding energy. The letter labels indicate different basis sets: A, 6-311G** basis set, no augmentation; B, A6-311G**P; C, A6-311G**PD.

the accepted bonding energy). This indicates possible overestimation of the bonding energy with extended 6-311G** basis sets. We attribute this to the approximate nature of the counterpoise model and an undercorrection of the large BSSE that arises with these basis sets.

$(CH_4)_2$. The first family of basis sets constructed for calculation of the bonding energy for $(CH_4)_2$ uses the 10s Huzinaga basis set on hydrogen and s, p, and d functions from the cc-pV6Z basis set on carbon. The second family of basis sets consists of the 6-311G** basis sets on both carbon and hydrogen atoms. Basis sets on carbon are extended with augmenting or tessellating functions optimized for BSSE-corrected bonding energy of the methane dimer at MP2 level of theory without the presence of additional functions on hydrogen. We also compute the bonding energy for the methane dimer using extended functions on hydrogen which are optimized for the bonding energy of $^3H_2$ and their combination with extended functions on carbon. BSSE-corrected bonding energies and BSSE for each of the basis sets are shown in Figures 11−14.

There are several experimental estimates of the methane dimer bonding energy based on spherically averaged potentials obtained from the fit of experimental data such as viscosity, virial coefficients, and methane-methane scattering.[48−50] The bonding energy given by these empirical potentials lies in the range of 574−797 μhartrees. All bonding energies computed with extended basis sets containing 32 or more augmenting or tessellating functions fall within this range. Extended basis sets with 12 augmenting or tessellating functions recover between 56 and 79.9% of the average bonding energy given by empirical potentials.

Interestingly, basis sets for $(CH_4)_2$ do not behave in the same manner as the basis sets for $^3H_2$ and Ne_2. First, a higher number of basis functions does not always generate a larger energy (although this general pattern is roughly followed in here too). Second, bonding energies and BSSE computed with extended 10sHuz/spd-cc-pV6Z basis sets and 6-311G** basis sets are of comparable size, with 10sHuz/spd-cc-pV6Z basis sets having noticeably larger BSSE in certain cases. Third, the contribution

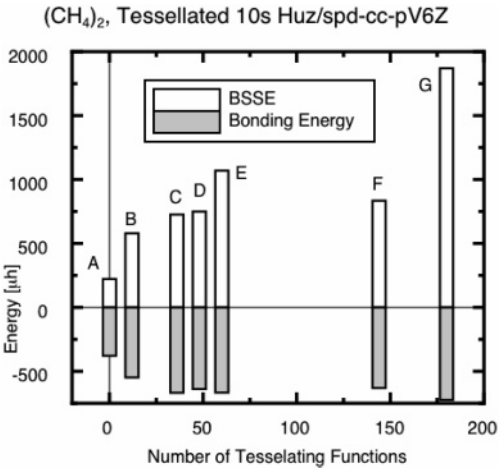

Figure 11. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with the tessellated modified 10s Huzinaga basis set on hydrogen and the spd-cc-pV6Z basis set on carbon. The radial distance from nucleus and the orbital coefficients for tessellating functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy. Parameters of the tessellated 10s Huzinaga basis set are optimized with respect to the $^3H_2$ bonding energy. The letter labels indicate different basis sets: A, 10s Huzinaga/spd-cc-pV6Z basis set, no tessellation; B, Huz/Tspd-cc-pV6ZS_v; C, Huz/Tspd-cc-pV6ZP_v; D, THuzS_v/spd-cc-pV6Z; E, THuzS_v/Tspd-cc-pV6ZS_v; F, THuzPv/spd-cc-pV6Z; G, THuzPv/Tspd-cc-pV6ZP_v.

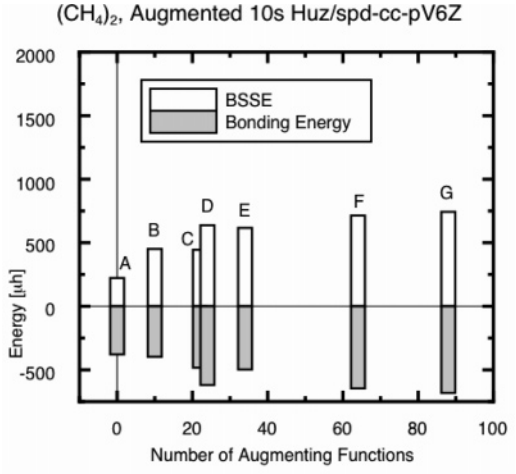

Figure 12. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with the augmented modified 10s Huzinaga basis set on hydrogen and the spd-cc-pV6Z basis set on carbon. Orbital coefficients of augmenting functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy. Parameters of augmenting functions on hydrogen are optimized with respect to the $^3H_2$ bonding energy. Basis sets C and D have the same number of tessellating functions (24). The letter labels indicate different basis sets: A, 10s Huzinaga/spd-cc-pV6Z basis set, no augmentation; B, Huz/Aspd-cc-pV6ZD; C, AHuzP/spd-cc-pV6Z; D, Huz/Aspd-cc-pV6ZDF; E, AHuzP/Tspd-cc-pV6ZD; F, AHuzPD/spd-cc-pV6Z; G, AHuzPD/Aspd-cc-pV6ZDF.

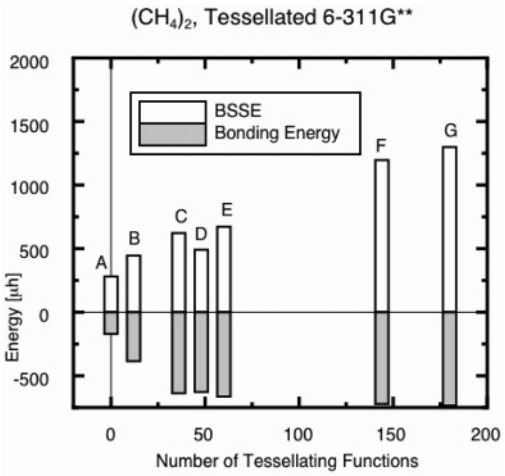

Figure 13. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with the tessellated 6-311G** basis set on hydrogen and carbon. The radial distance from the nucleus and the orbital coefficients for tessellating functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy. Parameters of tessellated 10s Huzinaga basis set are optimized with respect to the BSSE-corrected $^3H_2$ bonding energy. The letter labels indicate different basis sets: A, 6-311G**/6-311G** basis set, no tessellation; B, 6-311G**/T6-311G**S_v; C, 6-311G**/T6-311G**P_v; D, T6-311G**S_v/6-311G**; E, T6-311G**S_v/T6-311G**S_v; F, T6-311G**P_v/6-311G**; G, T6-311G**P_v/T6-311G**P_v.

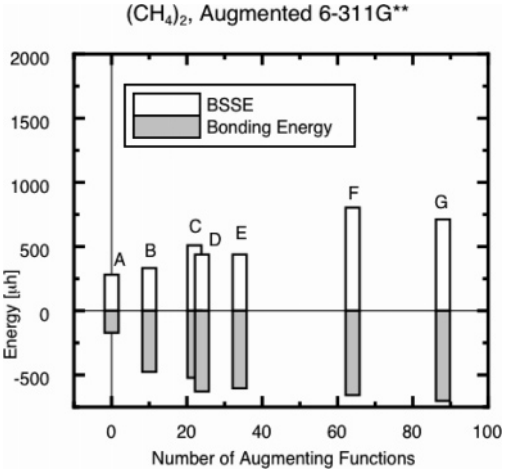

Figure 14. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with the augmented 6-311G** basis set on hydrogen and carbon. Orbital coefficients of augmenting functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy. Parameters of augmenting functions on hydrogen are optimized with respect to the BSSE-corrected $^3H_2$ bonding energy. Basis sets C and D have the same number of tessellating functions (24). The letter labels indicate different basis sets: A, 6-311G**/6-311G** basis set, no augmentation; B, 6-311G**/A6-311G**D; C, A6-311G**P/6-311G**; D, 6-311G**/A6-311G**DF; E, A6-311G**P/T6-311G**D; F, A6-311G**PD/6-311G**; G, A6-311G**PD/A6-311G**DF.

of MP4 to the bonding energy is not as significant as in the case of Ne_2. For example, the MP4 calculation with 6-311G** on hydrogen and T6-311G**P_v basis set on carbon gives -696.13 μhartrees, whereas the MP2 level of theory with the same basis set recovers -636.80 μhartrees for the methane bonding energy.

In all calculations for $^3H_2$, Ne_2, and $(CH_4)_2$ presented here, tessellated basis sets are more linearly dependent than augmented basis sets. The degree of linear dependence is usually measured by the smallest eigenvalue of the overlap matrix **S** (linear dependence is discussed in greater detail later in this paper). In a few cases the difference is quite significant: difference in the smallest eigenvalue between the tessellated and augmented basis set using approximately the same number of extending functions can be two to three orders of magnitude.

Augmented Basis Set Choices for Ne_2 and $(CH_4)_2$. Comparing the performance of tessellated and augmented basis sets leads us to the conclusion that, at present, augmented basis sets are more suited for calculation of intermolecular interactions. In this section we will focus on augmentation, in particular,

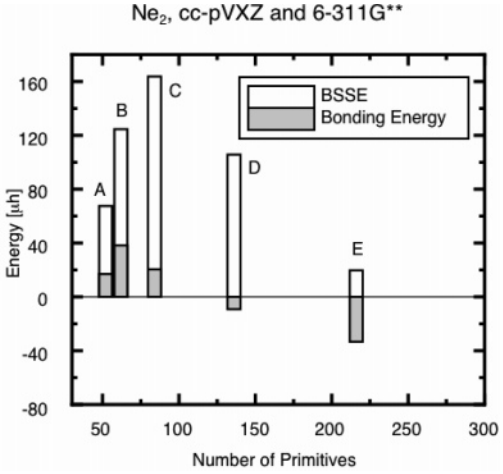

Figure 15. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the cc-pVXZ and 6-311G** basis sets at MP2 level of theory. The letter labels indicate different basis sets: A, cc-pVDZ; B, 6-311G**; C, cc-pVTZ; D, cc-pVQZ; E, cc-pV5Z.

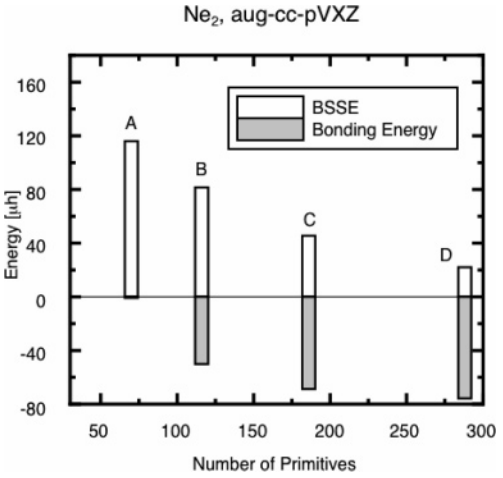

Figure 16. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the aug-cc-pVXZ basis sets at MP2 level of theory. The letter labels indicate different basis sets: A, aug-cc-pVDZ; B, aug-cc-pVTZ; C, aug-cc-pVQZ; D, aug-cc-pV5Z.

how to choose a basis set most suitable for augmentation. The two main criteria this basis set should satisfy are (1) it should provide a good description of the core and valence for the atoms present, so it will not be unbalanced after augmentation and (2) it should be as small as possible, so the calculation does not take an excessively long time and the basis set can be applied to larger systems (more than 3 or 4 first row atoms). Obviously, the choice of such a basis set will require some compromises. For example, the aug-cc-pV6Z basis set satisfies the first criterion, but it is too large to be efficient. The STO-3G basis set is small and fast; however, it does not provide a satisfactory description of the core and valence, and augmenting it results in an unbalanced basis set, leading to erroneous bonding energies even after the application of a BSSE-correction scheme. To illustrate this, we have augmented the STO-3G basis set with one p function optimized with respect to the $^3H_2$ bonding energy. The BSSE-corrected bonding energy obtained with this basis set at the UCCSD level of theory is -66.82 μhartrees, more than 3 times the size of the accepted bonding energy for this interaction, -19.52 μhartrees.[45] BSSE is 3631.60 μhartrees.

We have tested a series of Dunning's correlation consistent basis sets[51−54] as well as the 6-311G** basis set at the MP2 and LMP2 levels of theory for Ne_2. Figures 15 and 16 show

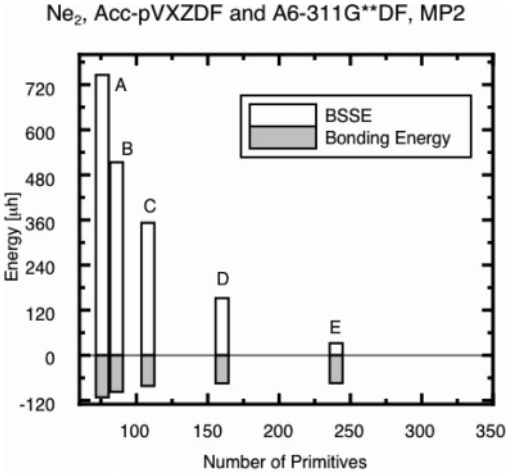

Figure 17. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the Acc-pVXZDF and A6-311G**DF basis sets at MP2 level of theory. Orbital coefficients of augmenting functions are optimized with respect to the Ne_2 BSSE-corrected bonding energy at MP2 level. The letter labels indicate different basis sets: A, Acc-pVDZDF; B, A6-311G**DF; C, Acc-pVTZDF; D, Acc-pVQZDF; E, Acc-pV5ZDF.

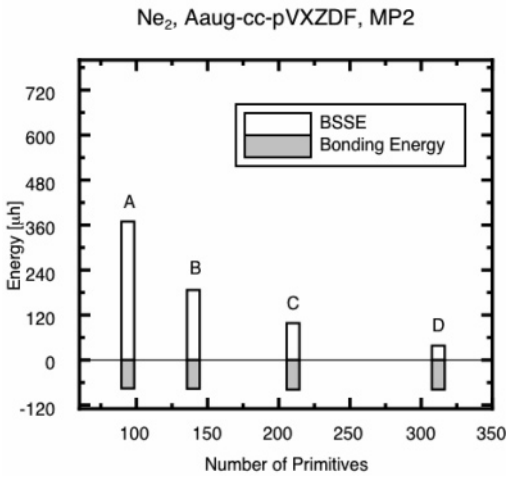

Figure 18. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the Aaug-cc-pVXZDF basis sets at MP2 level of theory. Orbital coefficients of augmenting functions are optimized with respect to the Ne_2 BSSE-corrected bonding energy at MP2 level. The letter labels indicate different basis sets: A, Aaug-cc-pVDZDF; B, Aaug-cc-pVTZDF; C, Aaug-cc-pVQZDF; D, Aaug-cc-pV5ZDF.

bonding energies and BSSE computed with cc-pVXZ, aug-cc-pVXZ (X = D, T, Q, 5) and 6-311G** basis sets. Clearly, the 6-311G** basis set as well as the cc-pVXZ series are not efficient in describing the interaction of Ne_2. The aug-cc-pVXZ family performs better, although even here larger basis sets are required to describe the interaction with satisfactory accuracy. Figures 17 and 18 show the results of MP2 calculations performed with the same basis sets but with each basis set augmented with a set of d and f functions optimized for the Ne_2 bonding energy. All of the augmented basis sets are able to describe the Ne_2 interaction, with the smallest basis sets (Acc-pVDZDF, A6-311G**DF) overestimating the bonding energy. Bonding energies and BSSE of these augmented basis sets behave smoothly with increasing size of the basis set. In both cases, the BSSE decreases with an increase in number of functions. Bonding energies calculated with Acc-pVXZDF and A6-311G**DF basis sets converge to the Ne_2 bonding energy from below and vary from -111.64 to -73.82 μhartrees. It is

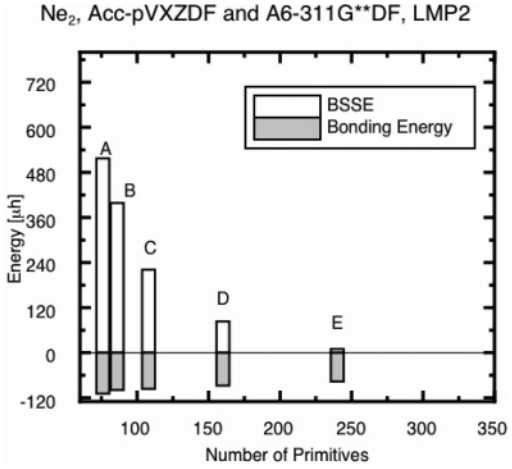

Figure 19. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the Acc-pVXZDF and A6-311G**DF basis sets at LMP2 level of theory. Orbital coefficients of augmenting functions are optimized with respect to the Ne_2 BSSE-corrected bonding energy at MP2 level. The letter labels indicate different basis sets: A, Acc-pVDZDF; B, A6-311G**DF; C, Acc-pVTZDF; D, Acc-pVQZDF; E, Acc-pV5ZDF.

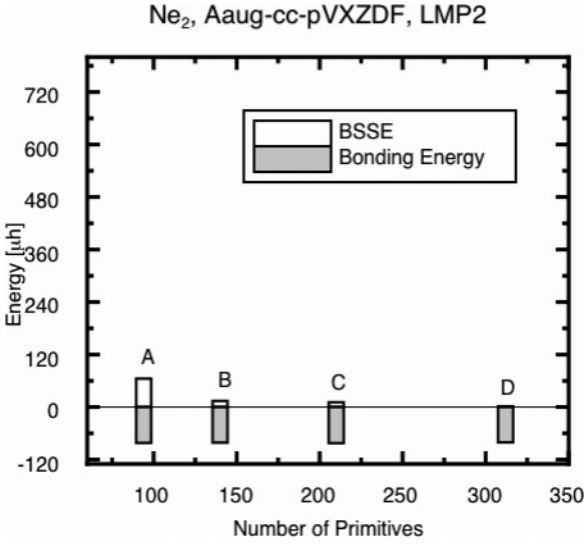

Figure 20. BSSE-corrected bonding energy and BSSE for Ne_2 computed with the Aaug-cc-pVXZDF basis sets at LMP2 level of theory. Orbital coefficients of augmenting functions are optimized with respect to the Ne_2 BSSE-corrected bonding energy at MP2 level. The letter labels indicate different basis sets: A, Aaug-cc-pVDZDF; B, Aaug-cc-pVTZDF; C, Aaug-cc-pVQZDF; D, Aaug-cc-pV5ZDF.

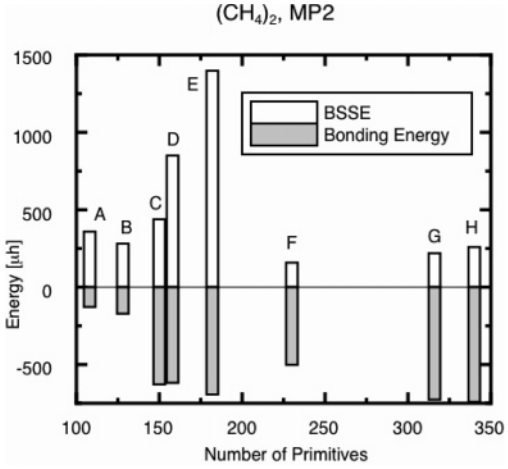

Figure 21. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with a mixture of standard and augmented basis sets at MP2 level of theory. Orbital coefficients of augmenting functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy at MP2 level, hydrogen basis set is not augmented. The letter labels indicate different basis sets: A − cc-pVDZ; B − 6-311G**; C − A6-311G**DF; D − aug-cc-pVDZ; E − Aaug-cc-pVDZDF; F − cc-pVTZ; G − aug-cc-pVTZ; and H − Aaug-cc-pVTZDF.

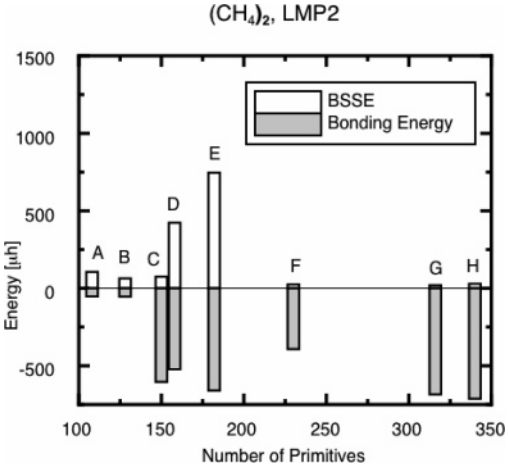

Figure 22. BSSE-corrected bonding energy and BSSE for $(CH_4)_2$ computed with a mixture of standard and augmented basis sets at LMP2 level of theory. Orbital coefficients of augmenting functions on carbon are optimized with respect to the $(CH_4)_2$ BSSE-corrected bonding energy at MP2 level, hydrogen basis set is not augmented. The letter labels indicate different basis sets: A, cc-pVDZ; B, 6-311G**; C, A6-311G**DF; D, aug-cc-pVDZ; E, Aaug-cc-pVDZDF; F, cc-pVTZ; G, aug-cc-pVTZ; H, Aaug-cc-pVTZDF.

far more common for systematic basis set enhancement to converge to an answer from above. Convergence from below is thought to be due to the approximate nature of the counterpoise correction. Bonding energies computed with Aaug-cc-pVXZDF basis sets (with the exception of Aaug-cc-pVQZ basis set) converge to the bonding energy from above and vary between −75.67 and −78.77 μhartrees. Results obtained at the LMP2 level of theory with the same basis sets are shown in Figures 19 and 20. In all cases, LMP2 bonding energies lie within 13 μhartrees of the MP2 bonding energies, most of them being slightly lower than bonding energies computed at the MP2 level of theory. Bonding energies computed with Acc-pVXZDF and A6-311G**DF basis sets still converge to the Ne_2 bonding energy from below, varying between −108.74 and −76.47 μhartrees. Bonding energies computed with Aaug-cc-pVXZ basis sets are almost saturated, varying between −80.77 and −82.38 μhartrees.

We have also computed the methane dimer bonding energy with several augmented and standard basis sets. The results are shown in Figures 21 and 22. As in the case of Ne_2, augmentation with a set of d and f functions improves bonding energies and the use of LMP2 substantially reduces the BSSE; nonetheless, some differences exist for the two dimers. The A6-311G**DF basis set performs better for the methane dimer than for Ne_2, it does not overestimate the $(CH_4)_2$ bonding energy, and its absolute as well as relative BSSE is smaller. Calculation of the $(CH_4)_2$ bonding energy with the Aaug-cc-pVDZDF basis set, which is quite efficient for Ne_2, shows a very large BSSE for $(CH_4)_2$. LMP2 bonding energies lie within 120 μhartrees of the MP2 bonding energies, all higher than energies computed at

the MP2 level of theory, which is opposite to the behavior for Ne_2 LMP2 bonding energies.

IV. Discussion

Overall, our calculations show that both tessellated and augmented basis sets are able to describe bonding in the weakly bound systems we tested: $^3H_2$, Ne_2 and $(CH_4)_2$. The fraction of recovered bonding energy in these dimers roughly scales with the number of augmenting or tessellating functions, irrespective of their type; however, tessellated basis sets have a larger BSSE than augmented basis sets, and they are also more linearly dependent. Additionally, augmented basis sets are much easier to use in conventional electronic structure codes, because one does not need to set up ghost centers for monomer and dimer calculations. Augmented basis sets seem to be more desirable and efficient for use in calculations of intermolecular interactions.

In this section we discuss a number of general issues that arise from the above studies of the calculation of intermolecular interactions for $^3H_2$, Ne_2, and $(CH_4)_2$. In so doing, we will focus on tessellation and augmentation of basis sets, BSSE, and linear dependence problems associated with basis sets.

Further Considerations on Tessellation. The original proposal of the tessellation approach was motivated by the idea that integrals over s functions are simpler and faster to evaluate than those over the higher angular momentum functions.[14] Therefore, one would expect that tessellating space around the nucleus with a large number of s functions optimized to describe intermolecular interactions rather than augmenting the nucleus centered basis set with higher angular momentum functions should result in substantial computational savings and make the computation of intermolecular interactions more accessible.

This hypothesis has, however, two flaws. First, placing a number of s basis functions in close proximity causes a severe linear dependency problem. Second, although it is true that individual two-electron integrals over s basis functions are simpler and therefore faster to evaluate than a single two-electron integral over functions with higher angular momentum, this simplification does not obtain for the modern computation of integrals involving large number of functions.

In modern two-electron integral codes, the calculation of integrals is done in batches. An integral batch consists of all the integrals for a unique combination of four shells. The number of integrals in each batch depends on the number of functions in each shell. For example, an (ss|ss) batch consists of only one integral, a (dp|ds) batch has a total of 75 or 108 integrals because it involves a d shell with 5 (or 6) components, a p shell with 3 components and an s shell with 1 component. Calculation of integrals in the batches allows for intermediate quantities to be computed and reused for many integrals within the batch avoiding their recalculation. This reduces the number of arithmetic operations needed per uncontracted integral within the batch and causes a relative increase in computational cost for integrals involving low quantum numbers such as (ss|ss) batches, because the cost of setting up each batch (which is quantum number independent) is divided over 81 possible integrals in a (pp|pp) batch but carried by only one integral in an (ss|ss) batch.[55]

We have performed some timing tests using the computational package Molpro 2000.1,[42] which implements a very efficient scheme for calculation of two electron integrals.[56] As our testing system we used the methane molecule with the modified 10s Huzinaga basis set on hydrogen and s, p functions from the cc-pV6Z basis set on carbon. We have augmented the carbon and hydrogen basis sets with an increasing number of s, p, and d functions and compared the timings for computation of two electron integrals for these basis sets. For up to about 15 additional primitives, the angular momentum dependence of the timing is insignificant. For more than 15 additional functions, there is a clear impact of angular momentum. Up to about 30 additional functions, the time needed for calculation of two electron integrals is smallest for basis sets augmented with p functions, and longest for basis sets augmented with d functions. For more than 30 additional functions, timings for basis sets augmented with p functions were still the fastest, followed by basis sets augmented with d functions, and finally basis sets augmented with s functions. For example, for basis sets augmented with 45 functions, the time for the computation of two electron integrals for p augmentation functions was half that of s functions.

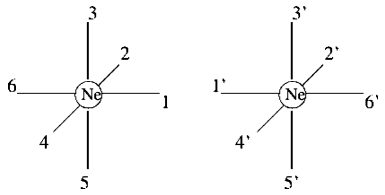

Figure 23. Ne_2 with indicated positions of tessellating functions in the vertexes of the octahedron.

Tessellation could still be viable if a specific (ss|ss) code were implemented or if p functions were used for tessellation. More efficient implementation of the s integral code or the use of p functions instead of s functions might not, however, bring the desired cost benefit because it seems that a relatively small number of tessellating functions is able to recover a substantial portion of the bonding energy. Moreover, one would still need to deal with the significant linear dependency problem.

Another issue is the relationship between tessellating and bond functions. Bond functions are auxiliary functions centered at the bond midpoints added to standard basis sets to improve the description of intermolecular interactions. The exponents of bond functions are chosen to maximize the correlation contribution to the bonding energy, but at the same time to keep the change in the Hartree−Fock interaction energy minimal. A set of {3s3p2d} bond functions with fixed exponents was found very useful to describe bonding in a variety of van der Waals systems. For a recent review of bond functions see ref 31.

The basic idea behind the use of tessellating functions is therefore very similar to that of bond functions. Both are off-centered and extend standard basis sets; however, some inherent differences between the two approaches exist. For example, due to their construction, bond functions cannot be associated with a particular molecule or atom in the weakly bound complex, so they cannot be used to describe bond dissociation potential curves. Tessellating functions, on the other hand, are directly associated with a particular atom. Also, because tessellating functions are placed symmetrically around the nucleus, they do not bias the basis sets in certain directions. To test whether each of the tessellating functions makes a contribution to the dimer bonding energy and not just those that extend in the bond direction, we have run test calculation on Ne_2 using portions of the Tspd-cc-pV6ZP$_v$ basis set. Figure 23 shows the labeling scheme and results are summarized in Table 2. Although the functions in the bond region make a larger contribution to the neon dimer bonding energy, other functions contribute as well. The presence of these nonbond functions also somewhat reduces the BSSE.

TABLE 2: BSSE-corrected Bonding Energies and BSSE for Ne_2[a]

tessellating functions used	bonding energy (μh)	BSSE (μh)
1-6, 1′-6′	−72.05	85.84
1, 1′	−67.96	152.39
1-5, 1′-5′	−71.84	96.37
2-5, 2′-5′	−56.25	31.06
2-6, 2′-6′	−59.00	33.95
6, 6′	−37.54	16.77
no tessellation	−32.91	8.62

[a] Energies and BSSE are computed on MP2 level of theory with the spd-cc-pV6Z basis set tessellated with p functions situated in different positions of the vertexes of the octahedron. Numbers in the first column denote tessellating functions present in the basis set; for more detailed description, see Figure 23.

Augmentation. Augmenting standard basis sets with d and f functions significantly improves computed bonding energies of van der Waals sytems. LMP2 is effective in reducing BSSE, and the reduction is most apparent with larger basis sets for which BSSE arises primarily at the correlated level. It is difficult to assess which of the standard basis sets is the most suited for such augmentation in terms of the computational cost and quality of the calculation. Among the basis sets tested, Aaug-cc-pVDZDF is the most efficient for Ne_2. For the methane dimer, the A6-311G**DF basis set is the most efficient. Suitability of a particular basis for one system does not imply the suitability of this basis set for a different van der Waals system. For example, although the A6-311G**DF basis set performs well for the methane dimer, it would not be our first choice for Ne_2 due to the overestimation of the Ne_2 bonding energy. Therefore, augmentation of medium-sized basis sets such as 6-311G** or aug-cc-pVDZ should always proceed with caution, especially when confronting a new system, one for which little or no data are available. Use of aug-cc-pVTZ or larger basis sets is desirable for more accurate results. If calculation with a larger basis set is prohibitive, medium size basis sets can be quite useful in obtaining qualitative insights into the behavior of new systems.

Basis Set Superposition Error. Basis set superposition error (BSSE) has been known for a long time to hamper the computation of intermolecular interactions. BSSE arises from incompleteness in the basis set of the monomers; in the dimer calculation, monomers use each other's basis functions to improve their energies and thus artificially lower the energy of the dimer. For a review of BSSE see refs 7 and 8 and references therein.

Several schemes can be used to remove the BSSE from calculations. A posteriori schemes (such as counterpoise correction) remove BSSE after the supermolecular calculation has been completed, a priori schemes (e.g., Chemical Hamiltonian) aim to remove the BSSE from the computational model. For example, in the Chemical Hamiltonian approach (CHA), BSSE effects are removed by modifying one-electron Hamiltonian in a manner that ensures that the free monomer wave functions remain unchanged in the extended basis set used for supermolecular calculation.[57,58] CHA has been successfully used at the SCF, DFT and MP2 level of theory. The drawback of the method is that the CHA Hamiltonian is non-Hermitian and its applicability to different levels of theory is nontrivial.

Other examples of a priori schemes for removal of BSSE are self-consistent field for molecular interaction for two and multicomponent systems,[59−61] the constrained dimer function approach,[62] the strictly monomer molecular orbital SCF approach,[63] and the method of Muguet and Robinson that attempts to remove BSSE using a specific localization scheme for Hartree−Fock molecular orbitals.[64]

The most popular scheme for BSSE correction is the a posteriori counterpoise (CP) correction proposed independently by Jansen and Ross[12] and Boys and Benardi.[13] The basic idea is to correct inconsistency in the basis set by using exactly same set of functions for computation of monomer, as well as, dimer properties; that is, calculation on one monomer is done in the presence of basis functions of the other monomer. The complete set of basis functions used for the calculation on a dimer is often called the dimer centered basis set, and the set of basis functions used for the monomer is called the monomer centered basis set. CP correction scheme is in principle very simple and applicable at any level of theory using conventional quantum chemistry codes.

Much discussion is found in the literature concerning the accuracy of the CP procedure and alternative approaches, such as the virtual-only counterpoise procedure[65] and a variety of a priori approaches mentioned earlier, are suggested. Currently, the CP procedure is widely accepted as a useful tool for eliminating most of the BSSE, some even claim that the CP correction rigorously eliminates BSSE in a supermolecular calculation for closed-shell fragments.[7] Our experience with computing bonding energies of van der Waals dimers at the MP2 level of theory with augmented and tessellated basis sets suggests that although the CP correction is very accurate, it does not necessarily remove BSSE completely. This is especially noticeable in our calculations for Ne_2 with tessellated and augmented 6-311G** basis sets (see Figures 9 and 10). These basis sets have very large BSSE compared to tessellated or augmented spd-cc-pV6Z (Figures 7 and 8) with the same number of extending functions. They also overestimate bonding energies. This overestimation is caused by a portion of the BSSE that is not corrected by the CP procedure. Counter to this interpretation is the relatively large Tspd-cc-pV6Z$P_v^2{}_f$ basis set's MP2 BSSE of 846.37 μhartrees and a bonding energy of −81.89 μhartrees, whereas the smaller A6-311G**DF basis set has a MP2 BSSE of 513.54 μhartrees and a bonding energy of −97.07 μhartrees. In this particular case, the basis set with smaller BSSE overestimates the bonding energy whereas the basis set with the larger BSSE does not. (Note that the best available calculation on the MP2 level of theory gives −84.5 μhartrees for the Ne_2 bonding energy.[16]) Interestingly, a large part of the BSSE for the A6-311G**DF basis set, about 300 μhartrees, arises at the Hartree−Fock (HF) level, whereas BSSE with the Tspd-cc-pV6Z$P_v^2{}_f$ basis set arises almost exclusively at the correlated level (BSSE at HF level is 1 μhartree). Therefore we speculate that the large BSSE already present at the HF level indicates an imbalanced basis set and a possible (very small) undercorrection of BSSE by the CP procedure.

Another way to partially remove BSSE is to use local correlation methods such as local MP2 (LMP2).[66−69] Removal of BSSE is a byproduct of these methods, their primary purpose being the reduction of computational cost associated with the treatment of electron correlation. Reduction in computational cost is achieved in two ways: (1) pair correlation between the distant orbitals is neglected (or, alternatively, this correlation is treated at a lower level); (2) virtual space for a given pair is restricted to a subset of atomic orbitals localized in the spatial vicinity of the correlated pair. Restriction of the virtual space also prevents basis functions located on distant centers from contributing with their tails to improve basis set flexibility, and as a result, the BSSE is reduced.[70,71] Although local correlation

methods will eliminate incremental BSSE arising at the correlated level, they do not impact HF level BSSE.

The positive effect of the LMP2 method on BSSE is also confirmed in this work. Nine different basis sets, aug-cc-pVXZ, aug-cc-pVXZ (X = D, T, Q, 5), and 6-311G**, are augmented with a set of p and d functions optimized with respect to the Ne_2 intermolecular bonding energy at the MP2 level of theory. Figures 17−20 show Ne_2 bonding energies as well as BSSE for these basis sets computed at the MP2 and LMP2 theory level. The effect of LMP2 on bonding energy is relatively small—LMP2 bonding energies are between 97 and 119% of MP2 bonding energies, indicating that augmented basis sets are almost saturated with respect to the intermolecular interaction. The impact of LMP2 on BSSE is significant—BSSE computed with the LMP2 method is reduced to 5−78% of the BSSE computed at the MP2 level of theory. LMP2 has an especially favorable effect when used with the augmented aug-cc-pVXZ series, in which BSSE is reduced to 5−18% of the BSSE computed with MP2.

Similar results are obtained for a mixture of standard (cc-pVDZ, 6-311G**, aug-cc-pVDZ, cc-pVTZ, aug-cc-pVTZ) and augmented (A6-311G**DF, Aaug-cc-VDZDF, Aaug-cc-pVTZDF) basis sets for the methane dimer. Figures 21 and 22 show methane dimer bonding energy and BSSE computed at the MP2 and LMP2 levels of theory with the above basis sets. Two small standard basis sets (cc-pVDZ, 6-311G**) experience a significant reduction in bonding energy as well as BSSE: LMP2 bonding energy is 32−41% of the MP2 bonding energy and LMP2 BSSE is 23−30% of the MP2 BSSE. Bonding energies computed with all other basis sets are not significantly affected (LMP2 recovers between 85 and 96% of the MP2 bonding energy), whereas BSSE is substantially reduced (BSSE computed at the LMP2 level of theory is between 9 and 53% of the MP2 BSSE).

One could also mitigate BSSE by using large, almost complete basis sets. Unfortunately, this approach is costly in terms of the computational time that scales with the number of basis functions N as N^4 or N^6 for correlated methodologies. Moreover, this approach creates linear dependency problems which are discussed next.

Linear Dependence. Calculations performed with large basis sets, especially those containing diffuse functions or those with several closely spaced sets of off-centered functions, often suffer from numerical instabilities due to linear dependency (i.e., two or more functions spanning almost the same physical space). This problem has been known to arise for calculations on periodic systems[72] and calculations using a large set of bond functions[73] and is also observed for tessellated basis sets.

Two main problems exist for nearly linearly dependent basis sets, one at the HF level and the other for correlated calculations. Solution of the HF equations typically involves construction of an orthogonalizing transformation matrix $\mathbf{S}^{-1/2}$ ($\mathbf{S}$ is an overlap matrix). For near linear dependence in the basis set, eigenvalues of the $\mathbf{S}$ matrix will approach zero and construction of $\mathbf{S}^{-1/2}$ involves dividing by quantities that are nearly zero. $\mathbf{S}^{-1/2}$ becomes almost singular, leading to problems in numerical precision. Near linear dependence also leads to very large molecular orbital (MO) coefficients for virtual orbitals. This is problematic for correlated calculations that require transformed integrals. For example, if a MO coefficient is 1000, the product of four such coefficients is 10^{12}. Because two-electron integrals are at best evaluated to an accuracy of 10^{-14}, the transformed integrals will have very large numerical errors and the resulting energies may behave abnormally.

Only a few publications in the literature discuss linear dependence or near linear dependence and the ways to avoid it.[72−74] The severity of this problem is usually measured by the size of the smallest eigenvalue of the overlap matrix $\mathbf{S}$. The smaller the eigenvalue, the greater the linear dependence. Serious numerical instabilities arise when the eigenvalues are of the order 10^{-8} or smaller, although sometimes even larger eigenvalues (of the order 10^{-7} and 10^{-6}) generate unreliable results depending on the computational method used.

One way to deal with near linear dependence is to simply omit one or more of the most diffuse basis functions from the basis set, or alternatively omit linear combinations of basis functions that correspond to small eigenvalues of the overlap matrix. Some quantum chemical program packages (e.g., Gaussian) do this automatically for the user. This approach works very well in most cases but can cause spurious results when applied to the calculation of intermolecular interactions using the supermolecular approach. To compute the bonding energy by the supermolecular method, one usually computes the energy of monomers using a monomer centered basis set and the energy of dimer with dimer centered basis set. This inconsistency in the basis is the cause of BSSE when one uses unsaturated basis sets (see the discussion in the previous section) and is usually not considered to be an issue in calculations using large, almost saturated basis sets. If the basis sets used in a computation are nearly linearly dependent, different linear combinations of basis functions may be deleted from the monomer and dimer centered basis sets making the calculation inconsistent and the computed bonding energies no longer reliable.

This problem can be somewhat alleviated by using the dimer centered basis set for calculations on the monomer in the same way one uses the dimer centered basis set to compute counterpoise corrections. This ensures that the same basis set is used in both monomer and dimer calculations and therefore that the same linear combinations corresponding to small eigenvalues of the $\mathbf{S}$ matrix are deleted by the program. The bonding energy computed from such a calculation might be smaller than expected, because the basis functions removed by the program are not completely redundant and do provide contribution to the bonding energy of the system. Alternatively, one could turn off the automatic removal of the basis functions by the program and risk the unreliability of a numerically unstable calculation resulting from the nearly linearly dependent basis set. Obviously, it would be best to avoid using linearly dependent basis sets altogether.

One is often tempted to believe that using faster computers, parallel processing, larger disk space, and computer memory will help push computational limits further, by enabling calculations at higher levels of theory with larger basis sets. In reality, this is a rather oversimplified view. Gaussian basis sets that have been used in quantum mechanical computations for the past fifty years with great success suffer from one serious shortcoming: a large number of Gaussian functions is required to describe systems with high levels of accuracy, especially for calculations of nonbonded interactions for which dispersion plays an important role. For example, one of the most accurate computations of $(NH_3)_2$ at the MP2 level of theory used 1024 basis functions.[75] Our largest computation on methane dimer with the Tspd-cc-pV6ZP_v basis set on carbon and THuzP_v basis set on hydrogen uses a total of 466 primitives and the smallest eigenvalue of the $\mathbf{S}$ matrix is 10^{-7}. The methane dimer basis set could still be extended to give a more accurate description of the system, but due to the problems with linear dependence

as well as the computational cost, we feel that using more than 500 primitives is not advisable. In general, due to the fact that Gaussian basis functions are not orthogonal, increasing the number of Gaussian functions in a basis set leads to near linear dependency, which causes the computation to be numerically unstable. We believe that learning how to deal with the problem of linear dependence in the basis set, whether at the programming level (by finding more efficient ways to treat numerical instabilities) or modeling level (by discovering functions with better spatial properties to replace Gaussians) is one of the challenges computational chemistry faces in future years.

V. Conclusion

In this work the role of basis sets in ab initio calculations of intermolecular interactions is explored. The focus of this study is on three systems in which dispersion interactions dominate: $^3H_2$, Ne_2 and $(CH_4)_2$.

Standard basis sets are augmented and tessellated with functions optimized with respect to the BSSE-corrected intermolecular bonding energy of each system. Relatively few tessellating or augmenting functions recover a significant portion of the intermolecular bonding energy. To obtain reliable results, both tessellated and augmented basis sets require underlying standard basis sets that provide a good description of the core and valence electrons. Tessellated basis sets with a large number of s functions do not offer an advantage in terms of the computational speed in comparison with the use of higher angular momentum atom-centered functions. Augmented atom-centered basis sets are less linearly dependent and easier to work with than tessellated basis sets, and therefore more desirable for use in the calculations.

Our results further suggest that the counterpoise correction, applied to the closed shell van der Waals clusters at the MP2 level of theory is very accurate, removing nearly all BSSE, although it is not exact. LMP2 theory is efficient in removing that portion of the BSSE that arises at the correlated level.

Linear dependence of the basis sets is a real problem for accurate ab initio calculation of intermolecular interactions, and we believe more attention needs to be focused on solving this fundamental issue. Current approaches to eliminating linear dependencies from basis sets do so by deleting a linear combination of basis functions corresponding to the smallest eigenvalues of the overlap matrix **S.** This approach inadvertently causes supermolecular calculations to be inconsistent because different combinations of functions may be deleted from monomer and dimer basis sets. Therefore, it is recommended that the same set of functions (i.e., the dimer centered basis set) be used for calculations on both monomer and dimer, even if the monomer basis set is saturated. Use of dimer centered basis sets in all calculations will ensure the removal of the same sets of functions for each calculation.

Overall, at present, no shortcut exists to reliable computation of intermolecular interactions.

References and Notes

(1) Szalewicz, K.; Jeziorski, B. Symmetry-Adapted Perturbation Theory of Intermolecular Interactions. In *Molecular Interactions*; Sheiner, S, Ed.; Wiley Tutorial Series in Theoretical Chemistry; Wiley: Chichester, England, 1997; p 1.

(2) Chalasinski, G.; Szczesniak, M. *Chem. Rev.* **2000**, *100*, 4227.

(3) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

(4) Perez-Jorda, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.

(5) Cybulski, M. S.; Seversen, C. E. *J. Chem. Phys.* **2005**, *122*, 014117.

(6) Kestner, N. R. *J. Chem. Phys.* **1968**, *48*, 252.

(7) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1973**, *59*, 4557.

(8) Van Duijneveldt, F. B. Basis Set Superposition Error. In *Molecular Interactions*; Sheiner, S., Ed.; Wiley Tutorial Series in Theoretical Chemistry; Wiley: Chichester, England, 1997; p 81.

(9) Kestner, N. R.; Combariza, J. E. Basis Set Superposition Errors: Theory and Practice. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1999; Vol. 13, p 99.

(10) van Lenthe, J. H.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Adv. Chem. Phys.* **1987**, *69*, 521.

(11) Gutowski, M.; Chalasinski, G. *J. Chem. Phys.* **1992**, *98*, 5540.

(12) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

(13) Jansen, H. B.; Ros, P. *Chem. Phys. Lett.* **1969**, *3*, 140.

(14) Boys, F.; Benardi, F. *Mol. Phys.* **1970**, *19*, 553.

(15) Rappé, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.

(16) Dunning, T. H. *J. Phys. Chem. A* **2000**, *104*, 9062.

(17) Woon, D. E. *J. Chem. Phys.* **1993**, *100*, 2838.

(18) Novoa, J. J.; Planas, M.; Rovira, M. C. *Chem. Phys. Lett.* **1996**, *251*, 33.

(19) Chalasinski, G.; Szczesniak, M. *Chem. Rev.* **1994**, *94*, 1723.

(20) Tsuzuki, S.; Uchimaru, T.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **1996**, *252*, 206.

(21) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 1975.

(22) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995.

(23) Sadlej, A. J. *Theor. Chim. Acta* **1991**, *79*, 123.

(24) Sadlej, A. J. *Theor. Chim. Acta* **1992**, *81*, 339.

(25) Kello, V.; Sadlej, A. J. *Theor. Chim. Acta* **1992**, *83*, 351.

(26) Tsuzuki, S.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **1998**, *102*, 2091.

(27) van Duijneveldt-van de Rijdt, J. G. C. M.; Duijneveldt, F. B. *J. Chem. Phys.* **1999**, *111*, 3812.

(28) Tao, F.-M. *J. Chem. Phys.* **1993**, *98*, 2481.

(29) Tao, F.-M.; Pan, Y.-K. *Chem. Phys. Lett.* **1992**, *194*, 162.

(30) Tao, F.-M. *J. Chem. Phys.* **1994**, *100*, 3645.

(31) Tao, F.-M. *Int. Rev. Phys. Chem.* **2001**, *20*, 617.

(32) Neumark, G. F. Ph.D. Thesis, Columbia University, New York, 1951.

(33) Kimball, G. E.; Neumark, G. F. *J. Chem. Phys.* **1956**, *26*, 1285.

(34) Frost, A. A. The Floating Spherical Gaussian Orbital Method. In *Methods of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Modern Theoretical Chemistry; Plenum: New York, 1977; Vol. 3, p 29.

(35) Archibald, R. M.; Armstrong, D. R.; Perkins, P. G. *J. Chem. Soc., Faraday Trans. 2* **1974**, *70*, 1557.

(36) Spangler, D.; Christoffersen, R. E. *Int. J. Quantum. Chem.* **1978**, Suppl. 5, 127.

(37) Pakiari, A. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 155.

(38) Pakiari, A. H.; Keshavarz, M. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *337*, 155.

(39) Whitten, J. L. *J. Chem. Phys.* **1963**, *39*, 349.

(40) Whitten, J. L. *J. Chem. Phys.* **1966**, *44*, 369.

(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision x.x; Gaussian, Inc.: Pittsburgh, PA, 1998.

(42) Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Knowles, P. J.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Werner, H.-J. MOLPRO, version 2000.1 and 2002.6, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles.

(43) Huzinaga, S. J. *Chem. Phys.* **1965**, *42*, 1293.

(44) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(45) Kolos, W.; Wolniewicz, L. *J. Chem. Phys.* **1965**, *43*, 2429.

(46) Wilson, A. K.; van Mourik, T.; Dunning, T. H., Jr. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 339.

(47) Aziz, R. A.; Slaman, M. J. *Chem. Phys.* **1989**, *130*, 187.

(48) Matthews, G. P.; Smith, E. B. *Mol. Phys.* **1976**, *32*, 1719.

(49) Reid, B. P.; O'Loughlin, M. J.; Soarks, R. K. *J. Chem. Phys.* **1985**, *83*, 5656.

(50) Böhm, H. J.; Ahlrichs, R.; Scharf, P.; Schiffer, H. *J. Chem. Phys.* **1984**, *81*, 1389.

(51) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(52) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(53) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358.
(54) Dunning, T. H., Jr.; Woon, D. E. *J. Chem. Phys.* **1994**, *100*, 1975.
(55) Hegarty, D. Evaluation and Processing of Integrals. In *Advanced Theories and Computational Approaches to the Electronic Structure of Molecules*; Dykstra, C. E., Ed.; NATO ASI Series C; D. Reidel Publishing: Dordrecht, Holland, 1984; Vol. 133, p 39.
(56) Lindh, R.; Ryu, U.; Liu, B. *J. Chem. Phys.* **1991**, *95*, 5889.
(57) Mayer, I. *Int. J. Quantum Chem.* **1998**, *70*, 41.
(58) Mayer, I.; Valliron, P. *J. Chem. Phys.* **1998**, *109*, 3360.
(59) Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*, 157.
(60) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251.
(61) Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. *J. Chem. Phys.* **2001**, *115*, 3553.
(62) Sadlej, A. J. *J. Chem. Phys.* **1991**, *95*, 6705.
(63) Cullen, J. M. *Int. J. Quantum Chem.* **1991**, *Suppl. 25*, 193.
(64) Muguet, F. F.; Robinson, G. W. *J. Chem. Phys.* **1995**, *102*, 3648.
(65) Daudey, J. P.; Claverie, P.; Marieu, J. P. *Int. J. Quantum Chem.* **1974**, *8*, 1.
(66) Saebo, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213.
(67) Hetzer, G.; Pulay, P.; Werner, H.-J. *Chem. Phys. Lett.* **1998**, *290*, 143.
(68) Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691.
(69) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H.-J. *J. Chem. Phys.* **2000**, *113*, 9443.
(70) Schütz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997.
(71) Pedulla, J. M.; Vila, F.; Jordan, K. D. J. *Chem. Phys* **1996**, *105*, 11091.
(72) Gruneich, A.; Hess, B. A. *Theor. Chem. Acc.* **1998**, *100*, 253.
(73) van Mourik, T.; Vos, R. J.; van Lenthe, J. H.; van Duijneveldt, F. B. *Int. J. Quantum Chem.* **1997**, *63*, 805.
(74) Wallis, A.; McElwain, D. L. S.; Pritchard, H. O. *Int. J. Quantum Chem.* **1969**, *3*, 711.
(75) Stålring, J.; Schütz, M.; Lindh, R.; Karlström, D.; Widmark, P.-O. *Mol. Phys.* **2002**, *100*, 3389.