

EVALUATING THE EFFICACY OF SMALL BASIS SETS AND THE
COUNTERPOISE PROCEDURE TO REPRODUCE COMPLETE BASIS SET
LIMIT HIGHER-ORDER CORRELATION CORRECTIONS FOR WEAKLY
BOUND MOLECULAR CLUSTERS

by
Cara M. Thorne

A thesis submitted to the faculty of The University of Mississippi in partial
fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2014

Approved by

Advisor: Dr. Gregory S. Tschumper

Reader: Dr. Brian W. Hopkins

Reader: Dr. Nathan I. Hammer

© 2014
Cara M. Thorne
ALL RIGHTS RESERVED

Acknowledgements

I would like to acknowledge the Mississippi Center for Supercomputing Research for access to their computational resources and the National Science Foundation (CHE-0957317 and EPS-0903787) for funding this work. I would like to thank my advisor Dr. Tschumper for his guidance throughout this project and for giving me the opportunity to learn and conduct research in his lab since freshman year. I would also like to thank all the graduate students in the Tschumper Group, specifically Eric Van Dornshuld, who was always available to answer a question or to troubleshoot a problem. Finally, I would like to thank Dr. Hopkins and Dr. Hammer for being readers on my thesis.

Abstract

The efficacy of small basis sets and the counterpoise procedure to accurately reproduce higher-order correlation corrections, which are routinely employed to estimate CCSD(T) interaction energies at the complete basis set (CBS) limit using an additive scheme, was examined through the study of four dispersion-dominated dimer systems. The dimers of acetylene (HCCH), cyanogen (NCCN), diphosphorous (P_2) and 1,4-diphosphabutadiyne (PCCP) were studied in three different configurations: cross, parallel-displaced and t-shaped at the intermolecular distance corresponding to the minimum point on the CCSD(T) potential energy curve (PEC) for each system using MP2, CCSD and CCSD(T) methods in conjunction with 8 basis sets from Dunning’s correlation consistent set of basis sets both with and without diffuse functions. Six standard extrapolation techniques were used to produce CBS limit estimates of the higher-order correlation corrections based on the differences between CCSD(T) and MP2 interaction energies, denoted $\delta_{MP2}^{CCSD(T)}$, and seven standard extrapolation techniques were used to produce the corresponding differences between CCSD(T) and CCSD interactions energies, denoted $\delta_{CCSD}^{CCSD(T)}$. Triple- ζ quality basis sets are required to accurately reproduce the $\delta_{MP2}^{CCSD(T)}$ higher-order correlation correction to within $0.1 \text{ kcal mol}^{-1}$ while double- ζ quality basis sets with diffuse functions can reproduce the $\delta_{CCSD}^{CCSD(T)}$ higher-order correlation correction to within $0.1 \text{ kcal mol}^{-1}$. Only when the aDZ basis set was used to compute $\delta_{MP2}^{CCSD(T)}$ did the counterpoise procedure improve agreement of the higher-order correlation corrections to the CBS limit compared with non-corrected results. The counterpoise procedure rarely yielded a better description of the $\delta_{CCSD}^{CCSD(T)}$ higher-order correlation correction relative to the values obtained without the counterpoise procedure.

Contents

1	Introduction	1
1.1	Van der Waals Interactions	1
1.2	Motivation for Quantum Mechanics	2
1.2.1	Quantization	2
1.2.2	Wave-Particle Duality	5
1.3	The Schrödinger Equation	7
1.4	Approximate Computational Methods	8
1.4.1	Hartree-Fock Method	8
1.4.2	Correlated Methods	9
1.4.3	Basis Sets	10
1.4.4	Basis Set Superposition Error	11
1.4.5	Correcting for Basis Set Superposition Error	12
1.4.6	Higher-Order Correlation Corrections	13
2	Computational Methods	15
2.1	Model Systems	15
2.2	Level of Theory	17
2.3	Extrapolation to CBS Limit	17
3	Results and Discussion	22
4	Conclusions	24
	References	25

1 Introduction

1.1 Van der Waals Interactions

Relatively weak intermolecular interactions between non-bonded molecules and atoms can be collectively referred to as van der Waals forces or non-covalent interactions. The term includes instantaneous dipole - instantaneous dipole interactions which are present in the interactions between all atoms and molecules and are a result of long range electron correlation.¹ Van der Waals interactions are essential to protein structure from the folding of polypeptide chains into their secondary and tertiary structures to the association of multiple polypeptides to form quaternary structures. Although more energy is required to break a single covalent bond than an individual van der Waals interaction, it is the numerous weak interactions that predominate as a stabilizing force in protein structure.² An example of a quaternary van der Waals interaction that is vital to human health is the folding of hemoglobin. This protein consists of four protein subunits - two alpha and two beta chains - and a prosthetic heme group that interact with one another through a series of van der Waals interactions to form the functional conformation of the protein. Any deviation from this conformation would impair the ability of hemoglobin to bind oxygen which would in turn impair the ability of red blood cells, which are largely comprised of hemoglobin, to deliver oxygen to the body tissues via blood flow through the circulatory system.² These weak, non-covalent interactions also play a role in drug binding,³⁻⁶ crystal packing,^{7,8} directed self-assembly of nanomaterials⁹ and the structure and function of biomolecules in addition to proteins such as the double helical structure of DNA.¹⁰ In addition, non-covalent interactions are responsible for the very existence of the liquid phase

and all related effects such as solvent phenomena.¹ The ubiquity of these interactions makes their accurate characterization of vital importance. Accurate characterization of these interactions requires computational modeling using robust quantum mechanical electronic structure methods that include dynamic electron correlation.

1.2 Motivation for Quantum Mechanics

Many observations drove the development of quantum mechanics by demonstrating the inadequacy of classical physics to describe experimental outcomes. These observations include the blackbody effect, line spectra of atoms and molecules, heat capacities of solids, the photoelectric effect, and Compton scattering. These experimental observations required the introduction of two revolutionary concepts. First, physical quantities such as energy and momentum are not continuously variable but quantized, and second, the particle property of momentum and the wave property of wavelength are related and, consequently, matter exhibits wave-particle duality.

1.2.1 Quantization

A blackbody is an idealized physical body that absorbs all electromagnetic radiation regardless of its frequency or angle of incidence.¹¹ It is possible to measure the spectral density of radiation emitted from a blackbody as a function of frequency for a given temperature. When experimental results are plotted against theoretical curves predicted from classical theory for the same temperature, the two curves behave similarly at low frequencies but the theoretical curve increases infinitely with increasing frequency while the experimental curve reaches a maximum spectral density and then tapers off to zero with increasing frequency. The area under these curves gives the total energy per unit volume of the field of the blackbody. Thus, classical physics predicts that a blackbody will emit an infinite amount of energy at all temperatures above absolute zero!¹¹ Max Planck developed insights that led to an understanding

of blackbody radiation. He realized that he could only obtain agreement between experiment and theory if he assumed energy radiated by the blackbody was related to frequency by

$$E = nh\nu \tag{1}$$

where h is Planck's constant, n is a positive integer, ν is the frequency which is continuous and E is the energy which is quantized for a given ν . This was a radical departure from classical theory in which energy is continuous and can take on any value. The quantization of physical properties, such as energy, was substantiated by observations made in the field of atomic and molecular spectroscopy, and observations made with respect to the heat capacity of solids.

When light emitted from thermally excited electrons is dispersed with a prism or grating, it is seen to consist of collections of narrow, discrete lines at wavelengths characteristic to the species involved. Likewise, when atoms and molecules are exposed to a continuous spectrum of light, the transmitted light contains a series of dark bands corresponding to wavelengths also characteristic of the species involved. These discrete spectra are a direct consequence of the quantization energy in these atomic and molecular systems in which only photons of light with energies that exactly correspond to the difference between two energy levels of the electronic structure can be absorbed or emitted.¹² It should be noted that molecular spectra are more complex than atomic spectra because additional energy terms arise from vibrational and rotational degrees of freedom.¹² As a consequence of the increased number of modes, an increased number of allowed energy transitions exist which leads to the formation of complex bands seen in molecular spectroscopy. Despite this complexity, the same principles apply to atomic and molecular spectroscopy.

Based on the observations made from spectral data, Bohr extended Rutherford's picture of a planetary atom by proposing that electrons, which are bound by Coulombic attraction to the positively charged nucleus, move in discrete circular orbits in which

the angular momentum of the circular orbits is fixed in units of Planck's constant, h , and the energy of the orbits is also fixed and indexed by the principle quantum number.¹³ Bohr postulated that light could only be emitted when an electron makes a transition from a higher to a lower energy level with energy lost from the electron by an emission of a photon. This deviated from classical radiation theory in which stationary states can radiate energy to produce a continuous emission spectrum.¹³ Bohr's simple model would later provide the foundation for the more sophisticated modern quantum theory developed by Schrödinger, Heisenberg and others.

Additional evidence corroborating quantization of physical properties comes from the heat capacities of crystalline solids. Each atom in a crystal requires three coordinates to describe its position and thus has three degrees of freedom associated with it. A solid of N atoms therefore has $3N$ degrees of freedom, each with its own set of energy levels. Energy is gained or lost by molecules in molecular collisions. An estimate of the amount of energy gained or lost in a molecular collision is $k_b T$ where k_b is the Boltzmann constant and T is the absolute temperature. A molecule will only take up energy if the spacing between adjacent energy levels is approximately equal to $k_b T$. In classical mechanics, however, the equipartition theorem of energy states that in thermal equilibrium the same average energy is associated with each independent degree of freedom.¹⁴ As evidenced by the work of Dulong and Petit, who discovered that the heat capacity of many crystalline solids at room temperature is $3R$, where R is the ideal gas constant, predictions based on classical mechanics hold at high temperatures at which the quantum nature of energy levels becomes unimportant.¹¹ A large discrepancy exists between the prediction of heat capacities of solids by classical physics and experimentally obtained heat capacities at low temperatures due to the fact that the heat capacity of solids is an inherently quantized phenomenon that exhibits significant temperature dependence at low temperatures. At low temperatures, $k_b T$ is small, and there is insufficient thermal energy to access excited vibrational

states. As such, energy cannot be absorbed by the system and the heat capacity of the crystalline solid approaches zero as temperature approaches absolute zero; this is in accordance with the third law of thermodynamics.

1.2.2 Wave-Particle Duality

Despite the agreement with experiment, resistance to Planck's explanation of black-body radiation, that the energy of radiation comes in discrete packets or quanta, was strong until Einstein's explanation of the photoelectric effect gave Planck's hypothesis credence. The process of electron ejection from a solid surface by light is called the photoelectron effect. Classical theory predicts that only a small fraction of light incident over the entire surface of a solid will be absorbed by any single electron. Classical theory also predicts that electrons will only be emitted from the solid if the light is sufficiently intense and that the kinetic energy of each ejected electron increases with light intensity. Results from experiment negated these classical predictions. Experiment demonstrates that the number of emitted electrons is proportional to the light intensity, that their kinetic energy is independent of the light intensity and dependent on the light frequency, that no electrons are emitted unless the frequency of light is above a threshold frequency (regardless of the light intensity) and that even at very low light intensities (intensities so low that all the energy incident on the solid surface is only slightly greater than the threshold energy needed to eject a single electron) electrons are still emitted.¹¹ The inability of classical theory to explain experimental results once again resulted in the formation of a new theory. Albert Einstein hypothesized that the energy of light is proportional to its frequency:

$$E = \beta\nu \tag{2}$$

where β is a constant to be determined later. Not only did Einstein's hypothesis challenge classical electrodynamics, in which there is no relationship between the energy and the frequency of a light wave, but he determined that β is identical to Planck's constant h . Thus, the equation that relates the energy of light to its frequency is

$$E = h\nu \tag{3}$$

The result suggested that h is a "universal constant" that appears in seemingly unrelated phenomena. Its appearance in the context of the photoelectric effect caused Planck's hypothesis explaining blackbody radiation to gain greater acceptance. The fact that low intensities of light could cause the ejection of an electron led to the conclusion that all of the incident light energy can be concentrated in a single electronic excitation. The concept of a photon, a spatially localized packet of light, followed. Spatial localization is a characteristic of particles and thus it was concluded that light, a wave by classical definition, can exhibit wave-particle duality. Louis de Broglie reasoned that matter can also exhibit wave-particle duality. Diffraction of particles by a double slit later confirmed that particles such as electrons, neutrons and atoms also exhibit wave-particle duality which, along with quantization, is a property that distinguishes quantum mechanics from classical mechanics.¹¹

A second observation illustrating wave-particle duality is the Compton Effect. The Compton Effect concerns the inelastic scattering of x-rays by electrons. When an incident x-ray of wavelength λ interacts with a target object, the scattered x-rays travel at an angle to the original direction of propagation of the incident x-ray with a different wavelength, λ' , and less energy. This interaction also produces an ejected electron. To explain these experimental observations, it is necessary to describe the situation just as one would when explaining the effects of two particles (e.g. two billiard balls) colliding and scattering from one another. The x-ray scatters and causes an electron to be ejected from the target object in a direction that conserves

the momentum of the system. When momentum and energy conservation equations are taken into account, the scattering angles and the observed wavelength shift, $\Delta\lambda$, of the x-ray can be explained. The success of using the conservation of energy and momentum to describe the experimental data from the Compton Effect is powerful evidence that electromagnetic radiation has momentum like a particle and thus gives credence to the principle of wave-particle duality postulated by de Broglie.¹¹

1.3 The Schrödinger Equation

Building off the work of Louis de Broglie, Erwin Schrödinger developed the modern quantum theory that integrated wave-particle duality into classical physics. The Schrödinger equation is essentially a wave equation that can be used to describe particles. The non-relativistic, time-independent Schrödinger equation is the basis of all quantum mechanical calculations associated with this research project

$$\hat{H}\Psi = E\Psi \tag{4}$$

When this equation is applied to the case of atoms and molecules, E is the total energy of the system, Ψ is the n -electron wave function that depends on both the identities and positions of the nuclei as well as the coordinates of all the electrons, and \hat{H} is the Hamiltonian operator that specifies the kinetic (\hat{T}) and potential (\hat{V}) operators of all the nuclei and electrons.¹¹

$$\hat{H} = \hat{T}_{nuc} + \hat{T}_{elec} + \hat{V}_{nuc,nuc} + \hat{V}_{nuc,elec} + \hat{V}_{elec,elec} \tag{5}$$

Nuclei move much more slowly than electrons due to their greater mass. As a result, the nuclei can be approximated as stationary relative to the electrons. This assumption is the Born-Oppenheimer approximation, the result of which is that the kinetic energy of the nuclei is zero and the nuclear-nuclear repulsion term ($\hat{V}_{nuc,nuc}$) becomes

constant. This simplifies the Hamiltonian to Eq. 6:

$$\hat{H}_{elec} = \hat{T}_{elec} + V_{nuc,nuc} + \hat{V}_{nuc,elec} + \hat{V}_{elec,elec} \quad (6)$$

Once the Born-Oppenheimer approximation has been invoked, the electronic Schrödinger equation needs to be solved. There are no analytic solutions to the n -body problem when n is greater than two (where the two bodies are an electron and the nucleus).¹¹ As such, the Schrödinger equation, like methods in classical physics, must be solved numerically for systems containing more than a single electron. Exact solutions are therefore available for the energy levels of the hydrogen atom and single electron ions but not for multi-electron atoms. Thus, for most chemical applications, the Schrödinger equation must be solved by approximate methods.

1.4 Approximate Computational Methods

1.4.1 Hartree-Fock Method

The variational method is a commonly used numerical procedure in both classical and quantum physics. The idea behind the variational method is to guess a “trial” wavefunction for a problem that consists of adjustable parameters. These parameters are adjusted until the energy of the trial wavefunction is minimized. The motivation for this procedure comes from the Variational Theorem which states that the energy of any trial wavefunction is always an upper bound to the exact ground state energy.¹⁵ Thus, any variations to the trial wavefunction that result in a decrease in its energy are getting the approximate energy closer to the exact answer. Hartree-Fock theory is a variational method.

Hartree-Fock theory recovers 99% of the overall energy of the system but important chemistry occurs in the remaining 1%. This “missing” energy is the electron correlation energy that is defined as the difference between the exact electronic energy

of the system and the Hartree-Fock energy.¹⁶ The Hartree-Fock method misses this energy because it partially ignores electron correlation. In a system of n interacting bodies the instantaneous motions of the bodies are correlated. Mean-field approximations, such as Hartree-Fock theory, neglect the instantaneous correlated motions of the bodies and instead assume that each electron interacts with an average charge distribution due to the other $n-1$ electrons.¹¹ This is incorrect. Fortunately, procedures exist such that one can converge to “the right answer”, the answer most congruent with experimental results and theoretical predictions. The first component of convergent approaches in quantum chemistry treats the electron correlation problem (see the horizontal axis in Figure 1).⁹

1.4.2 Correlated Methods

Correlated methods are post Hartree-Fock methods. Many-body perturbation theory (MBPT), of which Møller-Plesset (MP) perturbation theory is a specific case, and coupled-cluster (CC) theory are two of the most popular approaches developed to systematically improve the computational description of electron correlation that is neglected by Hartree-Fock theory.⁹ The missing electron correlation energy is recovered by constructing the wavefunction out of many different electron configurations that are generated by exciting electrons from the occupied orbitals of the Hartree-Fock reference configuration to unoccupied orbitals. Expansions (a perturbative power series expansion in MBPT or an exponential expansion in CC for example) are used to construct approximate many-electron wavefunctions.⁹ When all possible excitations have been incorporated then the exact solution to the non-relativistic Schrödinger equation has been obtained. For instance, $MP2 \rightarrow MP3 \rightarrow MP4$, where the numbers indicate the order in the perturbation expansion, is one possible progression toward the full configuration interaction (FCI) limit. However, this series does diverge occasionally from the FCI limit.^{17,18} The CC expansion, including single (S), double

(D), triple (T), etc. excitations/substitutions, always converges to the FCI limit. The “gold standard” of convergent quantum chemistry is the CCSD(T) method, which includes all single and double substitutions as well as a perturbative approximation of the connected triples, since it generally provides results that are chemically reliable and can be applied to moderately sized systems.⁹ CCSD(T) is the most reliable method for describing non-covalent/van der Waals interactions that can still be applied to systems containing approximately two dozen atoms. The accuracy afforded by this method, however, comes with a substantial monetary and temporal computational cost. The unfavorable $\mathcal{O}(N^7)$ scaling of computational cost where N represents the size of the system, means that this robust method can only be applied to relatively small systems of approximately twelve non-hydrogen atoms.

1.4.3 Basis Sets

The second component of convergent approaches in quantum chemistry improves the atomic orbital (AO) basis sets (see the vertical axis in Figure 2). In quantum chemistry, a basis set refers to a set of non-orthogonal one particle functions used to build molecular orbitals (MOs).¹⁹ AOs are represented by atom-centered Gaussian functions that are relatively simple to compute and mimic analytical wavefunctions for one-electron atoms (i.e. the s , p , and d atomic orbitals). Quantum chemists are interested in approximating a molecular wavefunction Ψ . As such, molecular orbitals (MOs) are built out of AOs in the linear combination of atomic orbitals to form molecular orbitals (LCAO-MO) approximation. A larger basis set only improves the results to a computation (i.e. gets one closer to “the right answer”) if the series of basis sets is designed to converge systematically to the complete basis set (CBS) limit. Dunning’s set of correlation consistent basis sets are constructed to converge smoothly to the CBS limit.²⁰⁻²² These basis sets are typically denoted cc-pVXZ and refer to a Dunning correlation consistent, polarized valence, X -zeta basis where X denotes the

maximum angular momentum of the Gaussian atomic orbitals in the basis set and is also referred to as the cardinal number of the basis (D for double- ζ basis set, T for triple- ζ basis set, etc.).⁹

1.4.4 Basis Set Superposition Error

In studying weakly bound clusters, one is often interested in calculating the supermolecule interaction energy. To do so, the energy difference between a cluster of homogeneous monomers and the non-interacting fragments is calculated using Eq. 7.

$$E_{int} = E[(molecule)_n] - nE[molecule] \quad (7)$$

In the more general case of a heterogeneous cluster composed of N fragments (denoted f_i where i is the index), the interaction energy is calculated using Eq. 8.

$$E_{int} = E[cluster] - E[f_1] - E[f_2] - E[f_3] - \dots - E[f_N] \quad (8)$$

Whenever energies of the monomer fragments are compared to the energy of the cluster using the procedure outlined in Eq. 7 or Eq. 8 and these energies are computed using small, finite basis sets, an inconsistency is introduced. Essentially, the n monomers that comprise a cluster use a larger basis set when the computation is performed on the cluster than when the computation is performed on the individual monomers.⁹ Consider the interaction between monomers A and B to form the dimer AB. As the intermolecular distance between monomers A and B decreases, monomer B can use additional basis functions from monomer A to describe its electron distribution and vice versa. The result is an artificial strengthening of intermolecular interactions and a stabilization of the interaction energies as intermolecular distance decreases. It is not the addition of extra accessible basis functions that causes this error but the inconsistent treatment of the basis set for each monomer as the inter-

molecular distance is varied that amounts to a problem (i.e. performing computations on the separate monomers in one basis set and performing the computations on the now dimerized monomers in a different basis set).²³ This inconsistency was noted as early as 1968²⁴ and was later termed basis set superposition error (BSSE).²⁵

1.4.5 Correcting for Basis Set Superposition Error

Jansen and Ros in 1969²⁶ and S.F. Boys and F. Bernardi in 1970²⁷ independently developed the “counterpoise procedure”, which is the most common method to remove BSSE. The goal of counterpoise (CP) procedure is to compute the energy of the individual monomers in the basis set of the cluster so as to treat the basis set of each monomer consistently as intermolecular distance is varied. The method essentially makes the errors in the energy calculations for the two separate monomers nearly equal to the corresponding errors in the dimer cluster calculation. The errors are then expected to cancel when the energy difference between the cluster and the non-interacting monomers set at an infinite intermolecular distance, which are now calculated using the cluster basis set, is calculated.²⁷ Boys and Bernardi noted that the CP procedure requires additional computational time, compared with methods that do not involve the CP procedure, to compute system interaction energies but that the extra effort is “well worth while if it gives the possibility that much more accurate values of the intermolecular interactions may be obtained.”²⁷ There is an ongoing debate in the scientific community regarding the efficacy of the CP procedure to correct for BSSE associated with interaction energy calculations of dispersion dominated systems. Some groups profess their validity stating that counterpoise corrected values are superior for dispersion dominated systems.^{28,29} Other groups question this conclusion stating that CP corrections make the results of these dispersion dominated calculations worse more often than they make them better.³⁰

1.4.6 Higher-Order Correlation Corrections

In order to extend the CCSD(T) method to larger systems, many groups are employing an additive approach to estimate the CCSD(T) interaction energies with larger basis sets.^{1,31-41} In this approach, the correction term for a high level method, the second term in Eq. 9, is obtained using a relatively small basis set, denoted B, and is added to the CBS limit interaction energy of a lower level method denoted Low:

$$E_{\text{int}}^{\text{High/CBS}} \approx E_{\text{int}}^{\text{Low/CBS}} + \delta_{\text{Low/B}}^{\text{High/B}} \quad (9)$$

The higher level method is generally CCSD(T) while the lower level method is generally MP2 or CCSD, both of which are associated with a smaller scaling term, $\mathcal{O}(N^5)$ and $\mathcal{O}(N^6)$ respectively, and can therefore be applied to larger systems than can CCSD(T). The CCSD(T) correction term represents the difference between the CCSD(T) and MP2 or CCSD interaction energies computed using a common basis set. The CCSD(T) correction term describes correlation effects that are omitted in the MP2 and CCSD methods and thus produces a solution that is closer to the FCI limit than the interaction energy of the lower level method alone. In previous work performed by the Tschumper Group,³⁰ it was noted that triple- ζ and double- ζ quality basis sets were required to accurately reproduce the estimated CBS limit higher order correlation corrections, $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ and $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ respectively, to within $0.1 \text{ kcal mol}^{-1}$ of the CBS limit for dispersion dominated systems. However, the conclusions drawn were subject to criticism as their validity was contingent on the quality of the CBS limit estimates in the paper which were calculated using aTZ/aQZ extrapolations. It has been noted in previous works,⁴² that the non-monotonic convergence of higher order correlation corrections to the CBS limit with increasing basis set size induces errors in certain extrapolation schemes. The aTZ/aQZ extrapolation scheme was implicated in the above work. The goal of the work presented in this thesis is to produce

high quality estimates of the complete basis set (CBS) limit for the higher order correlation corrections of small yet challenging dispersion dominated dimers. The CBS limits used in Ref. 30 will be compared to the high quality estimates produced in this work to ascertain their caliber and to determine if the conclusions reached in Ref. 30 regarding the ability of some basis sets and the CP procedure to reproduce the higher-order correlation corrections to within benchmark standards are valid.

2 Computational Methods

2.1 Model Systems

A drawback of the correlated electronic structure techniques (CC and MBPT) is that their computational demands - namely memory, CPU time and disk space - increase sharply with the size of the system.⁹ So, although interest lays in dispersion dominated systems for their biological significance, it is not feasible to perform demanding computations that are required to obtain accurate extrapolation results to the CBS limit on large biomolecules. As such, prototype systems are studied. The benzene dimer can mimic the major intermolecular stabilizing forces found in interactions between large biomolecules such as the nucleotides of DNA. Therefore, the benzene dimer is a widely studied prototype of π - π interactions. Computational costs involved in studying benzene are still high, however. Even smaller prototype systems, such as acetylene, are used to study π -type interactions. Rigid, linear monomer geometries were adopted for all computations. Experimental bond lengths were taken from Herzberg^{43,44} for acetylene (HCCH) $r(\text{CC}) = 1.2030 \text{ \AA}$ and $r(\text{CH}) = 1.0600 \text{ \AA}$; for diphosphorous (P_2) $r(\text{PP}) = 1.8943 \text{ \AA}$; and for cyanogen (NCCN) $r(\text{CC}) = 1.3839 \text{ \AA}$ and $r(\text{CN}) = 1.1578 \text{ \AA}$. 1,4-diphosphabutadiene (PCCP) has been observed experimentally but structural characterization was not possible. Thus, no experimentally obtained geometrical parameters have been reported in the literature.⁴⁵ As such, bond lengths for PCCP used in this work were adopted from a past paper from the group³⁰ in which a low-level geometry optimization with the B2PLYP-D dispersion corrected double-hybrid functional with the cc-pVDZ basis set was used. The ge-

ometry optimization employed a loose convergence criterion of $1.5 \times 10^{-3} E_h \text{ bohr}^{-1}$ and was calculated with `Gaussian 09`.⁴⁶ The bond lengths obtained for PCCP using the aforementioned method ($r(\text{CC})=1.35560 \text{ \AA}$ and $r(\text{CP})=1.58597 \text{ \AA}$) were found to be consistent with CCSD(T) optimizations with correlation consistent triple- ζ basis sets.⁴⁷

The dimer structures were computed in three different configurations: cross (X), t-shaped (T) and parallel-displaced (PD). These prototype configurations are depicted in Figure 2 for $(\text{PCCP})_2$, but are applicable to all dimers studied in this work.

The D_{2d} X-configuration (Figure 2a) is defined by one intermolecular parameter, the intermolecular distance (R) between the mid-points of the central bond of each monomer. The torsional angle between the monomers is exactly 90 degrees. The C_{2v} T-configuration (Figure 2b) is also defined by one intermolecular parameter. The intermolecular distance (R) is measured from the mid-point of the central bond of one monomer, which is perpendicular to the molecular C_2 rotational axis of symmetry, to the nearest atom in the second monomer, which is collinear with the molecule’s C_2 rotational axis of symmetry. The C_{2h} parallel-displaced configuration (Figure 2c) is defined by two intermolecular parameters. R again indicates the separation between the monomers and is measured, in this case, as the distance between the C-C axes (or P-P axes for the diphosphorous monomers) of the two linear monomers in each system. The second intermolecular parameter is the horizontal slip distance (R_S) and is a measure of the horizontal distance between the mid-points of the central bonds of each monomer. R_S was fixed at 3.00 \AA for the acetylene dimer $(\text{HCCH})_2$, 2.80 \AA for the cyanogen dimer $(\text{NCCN})_2$, 2.31 \AA for the diphosphorous dimer $(\text{P}_2)_2$, and 2.66 \AA for the 1,4-diphosphabutadiene dimer $(\text{PCCP})_2$. These values correspond roughly to the average of the MP2 and CCSD(T) optimized R_S parameters were obtained with triple- ζ quality basis sets with diffuse functions.³⁰ In this work, R was also fixed (Table 1). These values correspond to the intermolecular distance at which an energy

minima occurred on the CCSD(T)/CBS potential energy curve.

2.2 Level of Theory

The basis sets employed in this study are Dunning’s correlation consistent basis sets. Basis sets augmented with diffuse functions (aug-cc-pVXZ or aXZ; $X=D, T, Q, 5, 6$) as well as those without diffuse functions (cc-pVXZ or XZ; $X=T, Q, 5$) were employed. Energy point calculations for each of the four homogeneous dimers at three configurations were performed with each of the aforementioned basis sets at the MP2, CCSD and CCSD(T) levels of theory. Computations for counterpoise corrections were performed for every system for comparison. Strict thresholds and convergence criteria were applied due to the weak nature of the interactions between dimers. These criteria ensured that the correlation energies converged to at least $1 \times 10^{-8} E_h$. In addition, the frozen core approximation was adopted for all calculations. The Molpro⁴⁸ software package was used to perform all energy point computations.

2.3 Extrapolation to CBS Limit

Correlation consistent basis sets, which are built to systematically converge to the CBS limit as the cardinal number of the basis set (X) increases, were employed in this study. Therefore, it is possible to use extrapolation techniques to estimate the CBS limit. The three-parameter exponential function suggested by Feller⁴⁹ was used to extrapolate to the Hartree-Fock (HF) CBS limit energy from the corresponding aXZ HF energies.

$$E_{\text{HF}}^{\text{aXZ}} = E_{\text{HF}}^{\text{CBS}} + a \exp(-bX) \quad (10)$$

For a sequence of three basis sets, the fit can be reduced to an algebraic expression. Best results are obtained when the most accurate data points, those calculated with the largest basis sets, are used in the extrapolation.⁵⁰ For this study, the sequence of

three basis sets used was aXZ where $X=Q, 5, 6$.

$$E_{\text{HF}}^{\text{CBS}} = E_{\text{HF}}^{\text{a6Z}} - \frac{(E_{\text{HF}}^{\text{a6Z}} - E_{\text{HF}}^{\text{a5Z}})^2}{E_{\text{HF}}^{\text{a6Z}} - 2E_{\text{HF}}^{\text{a5Z}} + E_{\text{HF}}^{\text{aQZ}}} \quad (11)$$

Other extrapolation techniques must be employed to describe the convergence behavior of the electron correlation energy (E_{corr}) as this energy value converges more slowly than the HF energy. The following expression, a manipulation of a formula suggested by Helgaker et al.,⁵⁰ is popular in extrapolation procedures of the electron correlation energy since it only uses the two largest basis sets and thus reduces to an algebraic expression. One need only to insert the correlation energies of the two largest basis sets, denoted by the cardinal numbers X_{max} and $X_{\text{max}} - 1$, into the equation to calculate an estimated electron correlation CBS limit.

$$E_{\text{corr}}^{\text{CBS}} = \frac{E_{\text{corr}}^{X_{\text{max}}}(X_{\text{max}})^3 - E_{\text{corr}}^{X_{\text{max}}-1}(X_{\text{max}} - 1)^3}{(X_{\text{max}})^3 - (X_{\text{max}} - 1)^3} \quad (12)$$

In this study, two different extrapolations were performed using the above equation: one in which $X_{\text{max}} = 5$ and $X_{\text{max}} - 1 = 4$ and another in which $X_{\text{max}} = 6$ and $X_{\text{max}} - 1 = 5$. Extrapolations employing only two data points can be subject to error if there is an outlier in the data. Extrapolations using four data points are more resilient in this regard. Thus, four additional extrapolation schemes, each fitting their respective parameters using the four most accurate data points obtained with the aug-cc-pVXZ family of basis sets ($X=T, Q, 5, 6$), were used to obtain estimates of the CBS limit electron correlation energy. Equation 13, is a simple two-parameter inverse cubic function suggested by Helgaker et al.⁵⁰

$$E_{\text{corr}} = E_{\text{corr}}^{\text{CBS}} + \frac{b}{X^3} \quad (13)$$

An incarnation of Helgaker’s formula was also used

$$E_{\text{corr}} = E_{\text{corr}}^{\text{CBS}} + \frac{b}{(X + \frac{1}{2})^3} \quad (14)$$

as well as a two-parameter fit to a quartic polynomial (*Schwartz4*) and a three-parameter fit to a sixth-degree polynomial (*Schwartz6*):⁵¹

$$E_{\text{corr}} = E_{\text{corr}}^{\text{CBS}} + \frac{a}{(X + \frac{1}{2})^4} + \frac{b}{(X + \frac{1}{2})^6} \{b = 0 \text{ Schwartz4, } b \neq 0 \text{ Schwartz6} \quad (15)$$

An additional extrapolation scheme for coupled cluster methods based on a generalized form of a two-point extrapolation was developed by Schwenke.⁵²

$$E_{\text{corr}}^{\text{CBS}} = (E_{\text{corr}}^{X_{\text{max}}} - E_{\text{corr}}^{X_{\text{max}}-1})F + E_{\text{corr}}^{X_{\text{max}}-1} \quad (16)$$

In the above equation, $E_{\text{corr}}^{X_{\text{max}}}$ and $E_{\text{corr}}^{X_{\text{max}}-1}$ are the correlation energies evaluated with the two largest available basis sets. In this study, the basis sets used were a6Z and a5Z such that $X_{\text{max}} = 6$ and $X_{\text{max}} - 1 = 5$. The term F in Equation 16 is a function that can potentially take on any value, however, Schwenke used accurate estimates of the correlation energies of a few small molecules along with a least-squares fit to produce a set of coefficients to use as F in the extrapolation of coupled cluster energies.⁵³ In this study, an F value of 2.2656206 was used to extrapolate to the CCSD CBS limit electron correlation energy, while an F value of 2.2479617 was used to extrapolate to the (T) CBS limit electron correlation energy. The CCSD(T) CBS limit electron correlation energy was computed by summing the CCSD and (T) CBS limit values.

Each of the seven extrapolation schemes discussed above was performed using data corrected with the CP procedure and data not corrected with the CP procedure (nonCP). However, the Schwenke extrapolation procedure was not originally designed for correlation energies from Møller-Plesset perturbation theory. Thus, a total of 12

extrapolation schemes were performed per system, six with CP corrected data and six with non-CP corrected data, when the energy points were computed at the MP2 level of theory and 14 extrapolation schemes were performed per system when the energy points were computed at the coupled cluster level of theory.

Interaction energies were then obtained in a two-step procedure. First, the HF CBS limit energy computed using Equation 11 and the CBS limit electron correlation energies, obtained using Equations 12 through 16, were summed to calculate the total energy of each monomer and dimer of every system under study:

$$E_{\text{total}}^{\text{CBS}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr}}^{\text{CBS}} \quad (17)$$

Then the energies of the monomers were subtracted from the dimer energies to yield the CBS limit interaction energies as in Equation 8. As 12 extrapolated values for the MP2 CBS limit correlation energy (and 14 values for the coupled cluster CBS limit correlation energies) were obtained for each structure, the procedure outlined above yields 12 (or 14) values for the CBS limit interaction energy per system. The mean of these 12 (or 14) values is taken to be the CBS limit interaction energy. A higher-order correlation correction term at the CBS limit is calculated by taking the difference between the CCSD(T) interaction energy at the CBS limit and the interaction energy of a lower level method, MP2 or CCSD, at the CBS limit:

$$\delta_{\text{Low/CBS}}^{\text{CCSD(T)/CBS}} = E_{\text{int}}^{\text{CCSD(T)/CBS}} - E_{\text{int}}^{\text{Low/CBS}} \quad (18)$$

This mean CBS limit is the value to which all other higher-order correlation corrections, obtained with the smaller basis sets mentioned above, will be compared in this study. Higher-order correlation corrections computed using an array of basis sets have been compared to the CBS limit values (Equation 19) to determine which basis sets can reproduce the CBS limit higher-order correlation corrections to within 0.1 kcal

mol⁻¹ and to gauge the efficacy of CP corrections in estimating these higher-order correlation corrections used in the additive approach to estimate the gold standard CCSD(T) interaction energies using lower level methods as in Equation 9.

$$\Delta\delta_{\text{Low}/B}^{\text{CCSD(T)}/B} = \delta_{\text{Low}/\text{CBS}}^{\text{CCSD(T)}/\text{CBS}} - \delta_{\text{Low}/B}^{\text{CCSD(T)}/B} \quad (19)$$

3 Results and Discussion

CBS limit estimates in Ref. 30 were computed using an extrapolation scheme that employed augmented triple and quadruple - ζ quality basis sets. Energetic properties calculated using these estimates include the CCSD(T) and MP2 interaction energies at the intermolecular distance associated with the lowest energy point on the CCSD(T) CBS potential energy curve. Cases have been reported where extrapolated aTZ/aQZ values were not sufficiently accurate for benchmark energetics. From Table 1, it can be seen that this is not the case for the CCSD(T) or MP2 interaction energies for the systems and configurations studied in Ref. 30 and in this work. Table 1 illustrates the quality of the CCSD(T) and MP2 interaction energy CBS limits used in this work. For the CCSD(T) interaction energy, the largest deviation from the mean CBS limit value, computed by taking the average of 14 extrapolation schemes employing basis sets as large as a6Z, was 0.047 kcal mol⁻¹. The largest deviation from the mean MP2 interaction energy CBS limit, computed by taking the average of 12 extrapolation schemes employing basis sets as large as a6Z, was 0.030 kcal mol⁻¹. When CBS limit CCSD(T) interaction energies from Ref. 30 were compared to CCSD(T) interaction energies from this work, the deviations between the two values never exceeded 0.105 kcal mol⁻¹ and were 0.048 kcal mol⁻¹ on average. Meanwhile, the deviations between the CBS limit MP2 interaction energies from Ref. 30 and this work were 0.059 kcal mol⁻¹ at most and 0.028 kcal mol⁻¹ on average. These differences are within benchmark accuracy of 0.1 kcal mol⁻¹ and speak to the quality of the aTZ/aQZ CBS limit estimates for CCSD(T) and MP2 interaction energies from Ref. 30.

A more important property with respect to the additive approach are higher order

correlation corrections. Table 2 illustrates the quality of the higher order correlation correction CBS limits used in this work. The 14 extrapolation procedures used to calculate the $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ higher order correlation corrections for each dimer and configuration never deviated from their average by more than $0.0079 \text{ kcal mol}^{-1}$ while the 12 extrapolation procedures used to calculate the $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher order correlation corrections never deviated from their average by more than $0.0257 \text{ kcal mol}^{-1}$.

When $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ higher order correlation corrections from Ref. 30 were compared with corresponding values from this work, deviations were $0.002 \text{ kcal mol}^{-1}$ on average and $0.005 \text{ kcal mol}^{-1}$ at most. These differences are negligible. The deviations associated with the $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher order correlation corrections are $0.027 \text{ kcal mol}^{-1}$ on average and $0.047 \text{ kcal mol}^{-1}$ at most. While these deviations associated with the $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher order correlation corrections are larger than those for the $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ higher order correlation corrections by an order of magnitude, they are still within benchmark accuracy. As such, aTZ/aQZ extrapolated $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ and $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher order correlation corrections used in Ref. 30 are accurate estimations of the CBS limit.

These results indicate that the conclusions drawn in Ref. 30 regarding the basis set dependence of higher order correlation corrections, namely that aTZ and aDZ basis sets are required to consistently reproduce the CBS limit $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ and $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher order correlation corrections over a wide range of configurations and intermolecular distances, hold. It was also mentioned in Ref. 30 that the CP procedure to correct for BSSE rarely improves agreement with CBS limit higher order correlation corrections and often increases deviations from the CBS limit. Only when the aDZ basis set was used to compute $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ did the CP procedure consistently improve agreement of the higher order correlation corrections to the CBS limit. As the accuracy of the CBS limit estimates used in Ref. 30 have been confirmed here, these conclusions regarding the CP procedure for calculating higher order correlation corrections in

dispersion-dominated systems should also hold.

4 Conclusions

Four small homogeneous dimers have been used to determine the basis set dependence of higher order correlation corrections and to examine the efficacy of the CP procedure in estimating these same higher-order correlation corrections that are often employed in additive schemes to estimate the CCSD(T) interaction energies at the CBS limit. To pinpoint the $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ and $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ higher-order correlation correction CBS limits for the dispersion dominated systems studied here, 12 (or 14) extrapolation procedures using basis sets as large as a6Z were employed. After establishing the quality of the estimated CBS limits used in Ref. 30 calculated using an aTZ/aQZ extrapolation procedure, it was determined that the conclusions made in Ref. 30 are valid. Therefore, for the systems of dispersion dominated dimers studied here, aDZ and aTZ quality basis sets are required to reproduce the $\delta_{\text{CCSD}}^{\text{CCSD(T)}}$ and $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ higher-order correlation corrections, respectively, to within $0.1 \text{ kcal mol}^{-1}$ of the CBS limit. Also, CP corrections for BSSE rarely improve agreement with the CBS limit higher-order correlation corrections. When improvements, compared to non-CP corrected data, are made under the CP procedure, the improvements are, at most, on the order of a few hundredths of a kcal mol^{-1} and generally only occur when the aDZ basis set is used. These small gains in precision do not justify the increased computational and temporal resources required to carry out the CP procedure.

References

- [1] P. Hobza, R. Zahradnik, and K. Müller-Dethlefs, *Collect. Czech. Chem. Commun.*, 2006, **71**(4), 443–531, DOI: 10.1135/cccc20060443.
- [2] C. Nelson, *Lehninger Principles of Biochemistry*, W.H. Freeman and Company, New York, 2008.
- [3] J. G. Hill and J. A. Platts, *Chem. Phys. Lett.*, 2009, **479**(4–6), 279–283, DOI: 10.1016/j.cplett.2009.08.021.
- [4] S. Li, V. R. Cooper, T. Thonhauser, B. I. Lundqvist, and D. C. Langreth, *J. Phys. Chem. B*, 2009, **113**(32), 11166–11172, DOI: 10.1021/jp905765c.
- [5] M. Kolář, T. Kubař, and P. Hobza, *J. Phys. Chem. B*, 2010, **114**(42), 13446–13454, DOI: 10.1021/jp1019153.
- [6] J. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. M. Jr., *J. Chem. Theory Comput.*, 2011, **7**(3), 790–797, DOI: 10.1021/ct100563b.
- [7] R. Podeszwa, B. M. Rice, and K. Szalewicz, *Phys. Rev. Lett.*, 2008, **101**(11), 115503/1–115503/4, DOI: 10.1103/PhysRevLett.101.115503.
- [8] R. Podeszwa, B. M. Rice, and K. Szalewicz, *Phys. Chem. Chem. Phys.*, 2009, **11**(26), 5512–5518, DOI: 10.1039/b902015b.

- [9] G. S. Tschumper, in *Reviews in Computational Chemistry*, John Wiley & Sons, Inc., 2009; pp. 39–90.
- [10] J. Černý, M. Kabeláč, and P. Hobza, *J. Am. Chem. Soc.*, 2008, **130**(47), 16055–16059, DOI: 10.1021/ja805428q.
- [11] R. Engel, *Physical Chemistry*, Pearson Education, Inc., Boston, 2013.
- [12] Bemath, *Spectra of Atoms and Molecules*, Oxford University Press, Oxford, 1995.
- [13] R. Eisberg, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, John Wiley & Sons, Inc., New York, 1985.
- [14] McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, 2008.
- [15] C. D. Sherrill, A brief review of elementary quantum chemistry, 2001.
- [16] P. O. Löwdin, *Adv. Chem. Phys.*, 1959, **2**(207), DOI: 10.1002/9780470143483.ch7.
- [17] M. L. Leininger, W. D. Allen, H. F. Schaefer, and C. D. Sherrill, *J. Chem. Phys.*, 2000, **112**(21), 9213–9222, DOI: 10.1063/1.481764.
- [18] J. Olsen, O. Christiansen, H. Koch, and P. J. ergensen, *J. Chem. Phys.*, 1996, **105**(12), 5082–5090, DOI: 10.1063/1.472352.
- [19] C. D. Sherrill, Basis sets in quantum chemistry, 2001.
- [20] T. Dunning, *J. Chem. Phys.*, 1989, **90**(2), 1007–1023, DOI: 10.1063/1.456153.
- [21] D. E. Woon and T. D. Jr., *J. Chem. Phys.*, 1993, **98**(2), 1358–1371, DOI: 10.1063/1.464303.
- [22] R. A. Kendall, T. H. D. Jr., and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**(9), 6796–6806, DOI: 10.1063/1.462569.

- [23] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, *Chem. Rev.*, 1994, **94**(7), 1873–1885, DOI: 10.1021/cr00031a007.
- [24] N. R. Kestner, *J. Chem. Phys.*, 1968, **48**(252), 252–257, DOI: 10.1063/1.1667911.
- [25] B. Liu and A. D. McLean, *J. Chem. Phys.*, 1973, **59**(8), 4557–4558, DOI: 10.1063/1.1680654.
- [26] H. B. Jansen and P. Ros, *Chem. Phys. Lett.*, 1969, **3**(3), 140–143, DOI: 10.1016/0009-2614(69)80118-1.
- [27] S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**(4), 553–566, DOI: 10.1080/00268977000101561.
- [28] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, *J. Phys. Chem. A*, 2009, **113**(38), 10146–10159, DOI: 10.1021/jp9034375.
- [29] G. Chalasiński, D. J. Funk, J. Simons, and W. H. Breckenridge, *J. Chem. Phys.*, 1987, **87**(6), 3569–3579, DOI: 10.1063/1.453002.
- [30] E. J. Carrell, C. M. Thorne, and G. S. Tschumper, *J. Chem. Phys.*, 2012, **136**(1), 014103/1–014103/11, DOI: 10.1063/1.3671950.
- [31] M. O. Sinnokrot and C. D. Sherrill, *J. Phys. Chem. A*, 2004, **108**(46), 10200–10207, DOI: 10.1021/jp0469517.
- [32] J. G. Hill, J. A. Platts, and H. J. Werner, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4072–4078, DOI: 10.1039/b608623c.
- [33] T. Janowski and P. Pulay, *Chem. Phys. Lett.*, 2007, **447**(1–3), 27–32, DOI: 10.1016/j.cplett.2007.09.003.
- [34] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**(17), 1985–1993, DOI: 10.1039/b600027d.

- [35] S. Tsuzuki and T. Uchamaru, *Current Organic Chemistry*, 2006, **10**(7), 745–762, DOI: 10.2174/138527206776818937.
- [36] B. W. Hopkins, A. M. ElSohly, and G. S. Tschumper, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1550–1558, DOI: 10.1039/b616878g.
- [37] K. L. Copeland, J. A. Anderson, A. R. Farley, J. R. Cox, and G. S. Tschumper, *J. Phys. Chem. B*, 2008, **112**(45), 14291–14295, DOI: 10.1021/jp805528v.
- [38] D. M. Bates, J. A. Anderson, P. Oloyede, and G. S. Tschumper, *Phys. Chem. Chem. Phys.*, 2008, **10**(19), 2775–2779, DOI: 10.1039/b718720c.
- [39] A. M. ElSohly, B. W. Hopkins, K. L. Copeland, and G. S. Tschumper, *Mol. Phys.*, 2009, **107**(8–12), 923–928, DOI: 10.1080/00268970802695404.
- [40] D. M. Bates and G. S. Tschumper, *J. Phys. Chem. A*, 2009, **113**(15), 3555–3559, DOI: 10.1021/jp8105919.
- [41] J. Rezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**(8), 2427–2438, DOI: 10.1021/ct2002946.
- [42] M. S. Marshall, L. A. Burns, and C. D. Sherrill, *J. Chem. Phys.*, 2011, **135**(19), 194102/1–194102/10, DOI: 10.1063/1.3659142.
- [43] G. Herzberg, *Molecular Spectra and Molecular Structure: Spectra of Diatomic Molecules*, R.E. Krieger Pub. Co, Malabar, 1989.
- [44] G. Herzberg, *Molecular Spectra and Molecular Structure: Electronic Spectra and Electronic Structure of Polyatomic Molecules*, Van Nostrand, Princeton, 1950.
- [45] M. Brönstrup, J. Gottfriedsen, I. Kretzschmar, S. J. Blanksby, H. Schwarz, and H. Schumann, *Phys. Chem. Chem. Phys.*, 2000, **2**(10), 2245–2250, DOI: 10.1039/b000225i.

- [46] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. C. X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09 Revision B.01, 2009.
- [47] B. Hajgató, N. N. Pham-Tran, T. Veszprémi, and M. T. Nguyen, *Phys. Chem. Chem. Phys.*, 2001, **3**(23), 5158–5164, DOI: 10.1039/b106927f.
- [48] H.-J. Werner, P. J. Knowles, F. R. Manby, M. Schütz, P. Celani, G. Knizia, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and A. Wolf, Molpro, version 2010.1, a package of ab initio programs, 2010.
- [49] D. Feller, *J. Chem. Phys.*, 1993, **98**(9), 7059–7071, DOI: 10.1063/1.464749.
- [50] T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.*, 1997, **106**(23), 9639–9646, DOI: 10.1063/1.473863.

- [51] J. M. L. Martin, *Chem. Phys. Lett.*, 1996, **259**(5–6), 669–678, DOI: 10.1016/0009-2614(96)00898-6.
- [52] D. W. Schwenke, *J. Chem. Phys.*, 2005, **122**(1), 014107/1–014107/7, DOI: 10.1063/1.1824880.
- [53] J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, *J. Chem. Phys.*, 2009, **131**(19), 194105/1–194105/13, DOI: 10.1063/1.3265857.

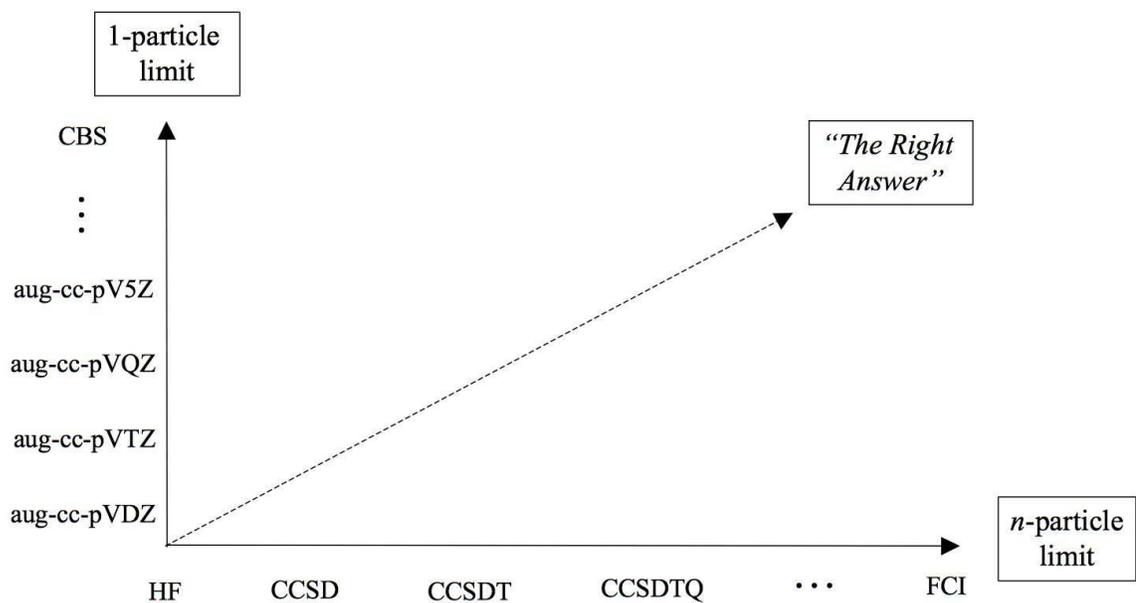


Figure 1: An example of a convergent quantum chemistry scheme that employs a systematic expansion of the size of the basis set to approach the complete basis set (CBS) limit and methods that systematically recover more of the electron correlation energy to approach the full configuration interaction (FCI) limit.

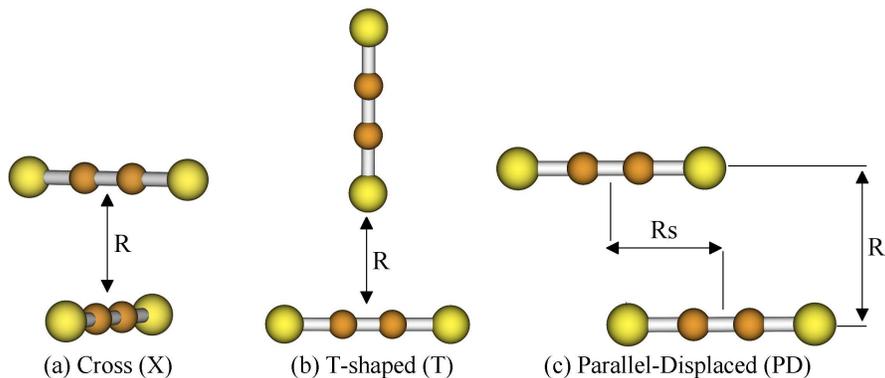


Figure 2: Three dimer configurations adopted in this study and the parameters used to characterize the separation between fragments.

Table 1: A comparison of CCSD(T) and MP2 CBS limit interaction energies (in kcal mol⁻¹) from Ref. 30 and this work along with the intermolecular distance (R in Å) associated with the lowest energy point on the CCSD(T) CBS curve. The maximum absolute and average absolute deviations (MAD and AAD in kcal mol⁻¹) associated with the differences between the interaction energies from Ref. 30 and this work are listed at the bottom of the table.

System	R _{min}	CCSD(T) $E_{\text{int}}/\text{CBS}$		MP2 $E_{\text{int}}/\text{CBS}$	
		Ref. 30 ^c	This work ^a	Ref. 30 ^c	This work ^b
(HCCH)₂					
X	4.1	-0.179	-0.183 ± 0.006	-0.341	-0.334 ± 0.003
PD	2.9	-1.381	-1.360 ± 0.015	-1.573	-1.534 ± 0.009
T	2.7	-1.456	-1.487 ± 0.024	-1.624	-1.622 ± 0.013
(NCCN)₂					
X	3.5	-0.444	-0.520 ± 0.047	-0.975	-1.015 ± 0.030
PD	3.3	-1.702	-1.756 ± 0.033	-2.398	-2.422 ± 0.027
T	3.1	-1.873	-1.954 ± 0.044	-2.414	-2.455 ± 0.040
(P₂)₂					
X	4.3	-0.857	-0.894 ± 0.014	-1.697	-1.725 ± 0.024
PD	3.9	-0.985	-1.001 ± 0.018	-1.823	-1.831 ± 0.025
T	4.3	-0.622	-0.619 ± 0.009	-1.127	-1.107 ± 0.012
(PCCP)₂					
X	3.5	-2.439	-2.544 ± 0.038	-5.092	-5.151 ± 0.023
PD	3.7	-2.380	-2.443 ± 0.032	-4.665	-4.686 ± 0.020
T	3.6	-1.661	-1.746 ± 0.037	-2.887	-2.934 ± 0.019
MAD:		0.105		0.059	
AAD:		0.048		0.028	

^a Average of 14 extrapolation schemes ± maximum deviation from the mean

^b Average of 12 extrapolation schemes ± maximum deviation from the mean

^c Obtained from Supporting Data at http://quantum.chem.olemiss.edu/053_DATA/

Table 2: A comparison of CBS limit higher-order correlation corrections for all dimers and configurations from Ref. 30 and this work. The maximum absolute and average absolute deviations (MAD and AAD) associated with the differences between the higher order correlation corrections Ref. 30 and this work are listed at the bottom of the table (all values in kcal mol⁻¹).

System	$\delta_{\text{CCSD/CBS}}^{\text{CCSD(T)/CBS}}$		$\delta_{\text{MP2/CBS}}^{\text{CCSD(T)/CBS}}$	
	Ref. 30	This work ^a	Ref. 30	This work ^b
(HCCH)₂				
X	-0.133	-0.1331 ± 0.0009	0.162	0.1512 ± 0.0052
PD	-0.200	-0.1967 ± 0.0017	0.193	0.1736 ± 0.0097
T	-0.189	-0.1867 ± 0.0025	0.167	0.1353 ± 0.0143
(NCCN)₂				
X	-0.381	-0.3796 ± 0.0026	0.530	0.4969 ± 0.0150
PD	-0.349	-0.3478 ± 0.0025	0.696	0.6681 ± 0.0132
T	-0.275	-0.2704 ± 0.0061	0.541	0.5031 ± 0.0156
(P₂)₂				
X	-0.456	-0.4560 ± 0.0034	0.841	0.8304 ± 0.0109
PD	-0.490	-0.4878 ± 0.0045	0.838	0.8289 ± 0.0148
T	-0.301	-0.2978 ± 0.0024	0.505	0.4875 ± 0.0108
(PCCP)₂				
X	-1.125	-1.1252 ± 0.0071	2.654	2.6073 ± 0.0257
PD	-1.106	-1.1020 ± 0.0079	2.285	2.2437 ± 0.0202
T	-0.648	-0.6459 ± 0.0054	1.226	1.1882 ± 0.0192
MAD:	0.005		0.047	
AAD:	0.002		0.027	

^a Average of 14 extrapolation schemes ± maximum deviation from the mean

^b Average of 12 extrapolation schemes ± maximum deviation from the mean