IMPLEMENTATION AND APPLICATIONS OF DENSITY-FITTED SYMMETRY-ADAPTED PERTURBATION THEORY

A Thesis
Presented to
The Academic Faculty

by

Edward G. Hohenstein

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
December 2011

IMPLEMENTATION AND APPLICATIONS OF DENSITY-FITTED SYMMETRY-ADAPTED PERTURBATION THEORY

Approved by:

Professor Kenneth R. Brown,
Committee Chair
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor C. David Sherrill, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Angelo Bongiorno
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Stephen C. Harvey
School of Biology
Georgia Institute of Technology

Professor Charles L. Liotta
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Robert L. Whetten
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: 14 July 2011

To my Grandmother

Celia Baeff

PREFACE

The role of theoretical chemistry is to offer insight into chemical phenomena that can not be easily explained through experiment alone, and to make predictions either ahead of or in conjunction with experiment. Nowhere is theoretical chemistry more important than in the understanding of noncovalent interactions. Due to the transient nature of many noncovalent complexes, experimental studies are fraught with difficulty. Additionally, noncovalent interactions often appear in large, complex systems (*e.g.* proteins and nucleic acids); isolating the effect of a specific interaction becomes a nearly impossible task. Using theoretical methods, it is possible to probe the nature of specific interactions within extended systems. It is also possible to study how changes to a molecule affect the way it interacts. This opens up the possibility, for example, to perform *in silico* drug design. At present, we are limited only by our ability to apply accurate theoretical methods to systems that are large enough to include all of the relevant effects.

The vast majority of recognized chemical problems could be solved with existing theoretical methods; their application to these problems depends only on the size of the problem and the scalability of current implementations of the theory. Unfortunately, these limitations are often quite prohibitive. This leads directly to a major thrust in theoretical chemistry research: the introduction of new approximations to theoretical methods. While these approximations take many different forms, the goal of producing scalable implementations of theoretical methods remains the same. This is coupled to developments in computer hardware that also increase the size of systems that can be studied. Together, algorithmic and technological advances are creating vast possibilities for the applications of theoretical chemistry.

ACKNOWLEDGEMENTS

I would like to thank Prof. David Sherrill for his support and guidance over the last four years. When I arrived at Georgia Tech, I had little exposure to theoretical chemistry, no programming experience, and I wasn't completely sold on the idea of being a theoretical chemist; Prof. Sherrill gave me a chance to learn some of the basics I was missing and put me in a position to succeed. Later, he became a sounding board for my ideas (good and bad) and was always willing to help out when I got stuck. He made sure that my efforts were rewarded with trips to conferences in Berkeley and Colorado. As I was approaching the end of my time at Georgia Tech, he made sure that I found the best postdoc I could. Without his help, the last four years would not have been nearly as enjoyable or successful. For all of this, I will be forever grateful.

I would also like to thank the students and postdocs of the Sherrill group. I have always enjoyed the atmosphere of friendly competition that exists in the group. In particular, I want to thank Michael Marshall and Dr. John Sears for their friendship and their critical ear. I want to thank Dr. Tait Takatani for our many projects together. Finally, I want to thank Robert Parrish for our time programming together and the many projects that it has made possible.

I want to thank the professors of Washington College who gave me the background I needed in chemistry and math to be a successful graduate student. I want to thank Prof. Louise Amick for encouraging me to pursue a math major. Prof. Anne Marteel-Parrish and Prof. Leslie Sherman for years of advice and support. Finally, I want to thank Prof. James R. Locker, my undergraduate research advisor. He saw that I had a growing interest in theoretical chemistry and made sure that there was a component of it in my research. The time I spent working with him convinced me to go to graduate school and study theoretical chemistry. Over the years, he has been a terrific friend and mentor; I especially enjoyed our fishing trips together.

Finally, I would like to thank my parents, Dee and Ed Hohenstein. Their years of love and support have gotten me to where I am today. They have been a constant source of stability, always making sure that I wind up where I need to be. They have always encouraged me to pursue the best opportunities that arise, even when is takes me far from home. I will never be able to repay all that they have given me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Noncovalent interactions play a vital role throughout much of chemistry. The understanding and characterization of these interactions is an area where theoretical chemistry can provide unique insight. While many methods have been developed to study noncovalent interactions, symmetry-adapted perturbation theory (SAPT) stands out as one of the most robust. In addition to providing energetic information about an interaction, it provides insight into the underlying physics of the interaction by decomposing the interaction energy into contributions from electrostatics, induction, exchange-repulsion, and dispersion. Therefore, SAPT is capable of not only answering questions about how strongly a complex is bound, but also why it is bound. This proves to be an invaluable tool for the understanding of noncovalent interactions in complex systems.

The wavefunction-based formulation of SAPT can provide qualitative results for large systems as well as quantitative results for smaller systems. In order to extend the applicability of this method, approximations to the two-electron integrals must be introduced. At low-order, the introduction of density fitting approximations allows SAPT computations to be performed on systems with up to 220 atoms and 2850 basis functions. Higher-orders of SAPT, which boast accuracy rivaling the best conventional theoretical methods, can be applied to systems with over 40 atoms. Additionally, higher-order SAPT benefits from approximations that attempt to truncate unnecessary unoccupied orbitals.

SAPT has proven especially useful in the study of heteroatom effects on π-π interactions. Here, benzene-pyridine and pyridine dimer complexes were used as a model for understanding the effect of nitrogen substitutions. SAPT computations implicate the introduction of a dipole, a reduction in polarizability, and a reduction in the spatial extent of the π orbitals for the changes in interaction energy. The indole-benzene complex contains many possible T-shaped configurations as well as several local minima on the π stacked potential energy surface. SAPT computations illustrate the origin of the energetic differences between all

of these geometries. Acene dimers are prototypes for π-π interactions in extended systems. The changes in these interactions with increasing linear acene length provide a glimpse into the nature of π-π interactions. Highly polarizable molecules and those containing high degrees of delocalization are often problematic for many theoretical methods; molecules that are both highly polarizable and delocalized can cause catastrophic failures in those methods. Different levels of SAPT are used to probe the problematic dispersion interactions in these types of complexes and locate the origin of some of these failures. Finally, the question of how substituents tune π-π interactions has been hotly contested amongst theoretical chemists for the last ten years. The most recent development has been the finding that both electron donating and electron withdrawing substituents increase the strength of the electrostatic interaction, which contradicts conventional wisdom. The application of SAPT clearly explains the origin of this surprising effect.

CHAPTER I

INTRODUCTION

1.1 Noncovalent Interactions

Noncovalent interactions are ubiquitous in chemistry. Solvation effects are governed by the noncovalent solvent-solvent and solvent-solute interactions.[40,222,223] Organic crystal structures and energetics are also determined through nonbonded interactions.[38,173,174,183] In biochemistry, the secondary structure of macromolecules contain significant contributions from noncovalent interactions. The structure of proteins involves interactions between side chains and the backbone.[23,28,142,181,182] Hydrogen bonding interactions give rise to the adenine-thymine (AT) and guanine-cytosine (GC) specificity observed in DNA. Stacking of base pairs contributes to the preference for different forms of DNA (*e.g.* A-form or B-form).[52,116,117,188,199,206] In addition to influencing secondary structure, noncovalent interactions govern the binding of small molecules to nucleic acids and proteins. Drug binding to proteins and nucleic acids as well as the intercalation of DNA are all dominated by noncovalent interactions.[18,53,88,126,129,132,135,150]

All noncovalent interactions are composed of the same four basic components: electrostatics, exchange-repulsion, induction, and dispersion. The distinction between different classes of noncovalent interactions is based on which effects dominate the interaction. For example, hydrogen bonding interactions are dominated by electrostatic effects, while π-π stacking interactions are dominated by dispersion effects. These individual components and their physical origins will be discussed in subsequent sections. In order to accurately describe noncovalent interactions, coupled-cluster with singles and doubles including perturbative triples [CCSD(T)][176] is relied upon. Beyond computing highly accurate interaction energies, it is often useful to obtain a decomposition of the energy into these four basic components. There are various energy decomposition techniques available,[12,27,36,86,111,122,152,170,207] but perhaps the most well-defined and robust is the symmetry-adapted perturbation theory

(SAPT).[111]

A review by Jeziorski *et al.* describes the development and applications of wavefunction-based SAPT.[111] In order to obtain accurate interaction energies from SAPT, there must be some account of the intramonomer electron correlation. The wavefunction-based formulation of SAPT including second-order intramonomer electron correlation corrections has been influential in the understanding of π-π interactions,[120,133,134,171,202,203] XH-π interactions,[181,217,218,220] and ion-π interactions.[46,119] Despite the successes of this method, current applications are typically limited to systems no larger than substituted benzene dimers. Recent work by Singh *et al.* included a particularly extensive application of wavefunction-based SAPT.[200] The largest system studied in this work was the benzene dimethyl-bipyridinium complex with a 6-31+G* basis. This system contains 40 atoms and roughly 500 basis functions; it is likely the largest wavefunction-based SAPT computation performed to date that includes some account of intramonomer electron correlation. In that work, however, only the electrostatic and exchange terms are corrected to account for intramonomer electron correlation.

1.1.1 Electrostatics

To a first approximation, electrostatic interactions between molecules can be described as the interactions between the permanent multipole moments of the monomers. This approximation is valid at long-range, but breaks down at short range as molecular orbital overlap increases and leads to significant amounts of charge penetration. A proper account of electrostatics requires the integration over the electron density of each monomer. The electrostatic interaction involves the electron repulsion between electrons of A and electrons of B, the electron-nuclear attraction between electrons of A and nuclei of B, the electron-nuclear attraction between electrons of B and nuclei of A, and the nuclear repulsion between nuclei of A and nuclei of B.[210]

Contributions to electrostatics from charge penetration are very important at short-range. To illustrate the origin of this effect, we will consider the interaction between two helium atoms. At long-range, the electrostatic interaction is zero, because the helium atoms

are spherically symmetric (thus they do not have permanent multipole moments). This allows all four contributions to electrostatics to be approximated as point charge interactions. At short-range, the orbitals of the two helium atoms overlap. This leads to deviations from idealized $1/r$ behavior in the electron-electron and electron-nuclear terms. Under the Born-Oppenheimer approximation, the nuclear-nuclear term is still treated with point charges. Each electron-nuclear term depends on one diffuse quantity (the helium $1s$ orbital), while the electron-electron term depends on two diffuse quantities. This leads to larger deviations in the electron-electron term (which becomes less repulsive) than in the electron-nuclear term (which becomes less attractive). The net effect is that the electrostatic interaction becomes attractive at short-range. This attractive interaction increases exponentially as the overlap between molecular orbitals increases.

1.1.2 Exchange-Repulsion

The primary repulsive component of the interaction energy is a result of exchange. This is a short-range effect resulting from the Pauli exclusion principle. As molecular orbitals overlap, electrons are repelled from the area between the monomers in order to satisfy the Pauli exclusion principle. This rearrangement of electrons raises the interaction energy. Exchange effects will be discussed elsewhere in the context of symmetry-adapted perturbation theory. The exchange-repulsion increases exponentially as molecular orbital overlap increases. In many force field type approaches, the exchange-repulsion is modeled as the $1/r^{12}$ term in a Lennard-Jones potential. Other, more robust, force fields will use the proper exponential form for exchange; even more infrequently, it is modelled with overlap integrals.[62,210]

1.1.3 Induction

Induction is, to a first approximation, the interaction between the permanent multipole moments of one monomer with the induced moments of the other. Additional, "self consistency" effects would include the interaction between induced moments. In general, any relaxation of a monomer wavefunction in response to the mean field of the other is classified as a type of induction. This definition includes intermolecular charge transfer as an induction interaction. Charge transfer in the context of SAPT will be discussed in detail later.

Many force field approaches do not include any account of induction. Only polarizable force fields attempt to capture the induction energy. Even then, the induction is partitioned into polarization, usually modelled with inducible atom-centered dipoles (and possibly higher order multipoles), and charge transfer. This partitioning into local and nonlocal induction contributions is not present in the quantum mechanical description of induction. Generally speaking, the inclusion of induction is most important for describing the interaction of polar molecules or charged species. This is especially true in cooperative hydrogen bonded complexes (*e.g.* HF trimer).

1.1.4 Dispersion

When the interaction of nonpolar molecules is considered, dispersion is usually the dominant attractive force. Dispersion is a result of interactions between instantaneous charge fluctuations. The leading contribution to dispersion is due to instantaneous dipole-dipole interactions, which gives rise to the $1/r^6$ dependence of the dispersion energy. From an electronic structure perspective, dispersion is purely a correlation effect; it can be thought of as correlating an electron on one monomer with an electron on the other monomer. As such, dispersion is not included in Hartree-Fock (HF) and poorly described (or neglected entirely) by standard density functionals. Empirical dispersion terms usually come from pairwise C_6/r^6 estimates, where C_6 comes from atomic polarizabilities and ionization potentials. This approach is found in most force fields and many dispersion corrected density functional theory (DFT) approaches. Higher-order dispersion terms, C_8/r^8 or C_{10}/r^{10}, are occasionally included.

Three-body dispersion effects (here, body refers to electrons) can be important in the context of dimer interactions as well as trimer interactions. These effects can be rationalized qualitatively for helium trimer. If an electron on monomer A induces instantaneous dipoles on monomers B and C, the instantaneous dipoles on B and C will interact with each other. If the trimer is arranged in a compact, equilateral triangle, this will lead to a repulsive interaction. This qualitative description can be used to explain the form of the Axilrod-Teller-Muto dispersion term and its $1/r^9$ dependence.[11,156] It is important to note

that Axilrod-Teller-Muto dispersion assumes spherical monomers (*e.g.* noble gas trimers). Although this correction has been applied to molecular species, it is unlikely that this is a generally reliable approach. Where possible, a fully quantum mechanical treatment of three-body dispersion should be preferred.

1.2 Computing Noncovalent Interactions

Noncovalent interactions can be computed using either *supermolecular* or *perturbational* approaches. Supermolecular computations of the interaction energy involve the computation of the total dimer and monomer energies.

$$E_{\text{int}} = E_{\text{Dimer}} - E_{\text{Monomer}_\text{A}} - E_{\text{Monomer}_\text{B}} \tag{1}$$

Perturbational approaches treat the interaction as a perturbation and solve for the interaction energy directly (*i.e.* without computing the total energy of the dimer). The most rigorous perturbational approach is the symmetry-adapted perturbation theory (SAPT),[111] which will be discussed briefly in this section and at great length elsewhere. Interaction energies can be computed with either wavefunction-based (WFN) methods or approaches from density functional theory (DFT).

In accurate quantum mechanical computations, a high degree of electron correlation must be included to reliably account for dispersion interactions. The current standard for accurately computing the interaction energy within a small, noncovalently bound complex is coupled-cluster with singles and doubles including perturbative triples [CCSD(T)].[176] Unfortunately, the applicability of CCSD(T) is hindered by the formal $\mathcal{O}(N^7)$ complexity of the method (more specifically, $\mathcal{O}(o^3v^4)$, where o and v are the number of occupied and virtual orbitals, respectively); to describe noncovalent interactions in large systems, less computationally expensive methods must be employed. The recently developed[214] spin-component scaled (SCS) CCSD has been shown to produce results which closely match CCSD(T); however, the formal scaling of this approach remains high at $\mathcal{O}(N^6)$. Second-order Møller-Plesset perturbation theory (MP2) offers another approach for describing noncovalent interactions; with formal $\mathcal{O}(N^5)$ complexity, MP2 can be extended to much larger systems than are accessible with coupled-cluster methods. However, MP2 is substantially less accurate

than the more rigorous coupled-cluster approaches.[203] While spin-component scaling[44,67,87] significantly improves the accuracy of MP2 on average, there are nevertheless cases where the accuracy is not as good as desired, or the system is too large for the computation to be feasible. More efficient methods that can effectively treat noncovalent interactions are necessary if complexes with 100 atoms or more are to be studied routinely. One approach is to reduce the computational scaling of coupled-cluster methods, and this is being actively pursued by several research groups.[56,189,193,211] Another approach is to attempt to improve the reliability of methods which are already applicable to larger systems.

There are many approximate methods that attempt to correct the description of dispersion by MP2 or CCSD. The SCS methods, originally introduced by Grimme, attempt to correct the correlation energy by empirically scaling the same- and opposite-spin components.[67,214] There are several different parameterizations of MP2, including some specifically for noncovalent interactions;[44,87] unfortunately, it is not clear that any one set of parameters is reliable for all types of interactions.[65,213,215] The SCS-CCSD method[214] has been shown to be much more reliable than SCS-MP2 for treating all types of noncovalent interactions.[65] However, this method requires an iterative $\mathcal{O}(N^6)$ procedure to compute an interaction energy, which limits its applicability. The MP2.5 method of Hobza and co-workers[167] is comparable to SCS-CCSD with regard to accuracy, yet does not require an iterative $\mathcal{O}(N^6)$ energy evaluation. Of the reliable, approximate wavefunction based methods, the least computationally expensive [noniterative $\mathcal{O}(N^5)$] and the most physically justified is the MP2C method of Hesselmann.[84,165] This method attempts to correct the behavior of MP2 by evaluating dispersion with frequency-dependent polarizabilities from time-dependent density functional theory (TDDFT).

DFT is widely used today for examining a variety of chemical systems with dozens of atoms or more.[161] The application of DFT to noncovalently bound complexes has been limited due to the failure of most density functionals to describe dispersion interactions, which can be critical for noncovalent complexes. Dispersion interactions are inherently long-range electron correlation effects, which are not captured by the popular local or semi-local density functionals.[4,34,90,114,128,130,225] Several approaches exist for improving existing density

functionals to handle dispersion effects. Among the more physically motivated approaches, Röthlisberger and co-workers have added effective atom-centered nonlocal potentials which have been fit to benchmark *ab initio* data;[232,233] Langreth, Lundqvist, and co-workers have introduced a van der Waals density functional (vdW-DF) which adds nonlocal terms to the correlation energy functional;[43] and Becke has proposed a novel approach that formulates the dispersion interaction in terms of the dipole moment that would be created when considering an electron and its exchange hole.[14,15,112]

A more pragmatic and simple approach is to add empirical terms that model dispersion interactions[68,69,115,248,255] These DFT-D approaches require only computation of interactions between atoms pairs; therefore, the additional computational expense is negligible. The DFT-D approach of Grimme is a widely applicable method for correcting the performance of standard density functionals.[68,69,74] This method utilizes a damped R^{-6} term to model the dispersion interactions (Equations 2, 3, and 4).

$$E_{\mathrm{DFT-D}} = E_{\mathrm{DFT}} - s_6 \sum_{i=1}^{N-1} \sum_{j>i}^{N} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}), \tag{2}$$

$$C_6^{ij} = \sqrt{C_6^i C_6^j}, \tag{3}$$

$$f_{\mathrm{dmp}}(R_{ij}, R_r, d) = \frac{1}{1 + e^{-d(R_{ij}/R_r - 1)}}. \tag{4}$$

The recently developed -D3 correction[73] includes R^{-6} and R^{-8} terms computed with atomic dispersion coefficients that depend on the chemical environment (through the steric numbers of the atoms). A possible improvement to DFT-D methods is to apply the dispersion correction to long-range corrected functionals that are parameterized in the presence of the dispersion correction; the ωB97X-D is such a functional[35] and appears to be particularly robust for noncovalent interactions when used with augmented triple-ζ basis sets.[29] The so called "double hybrid" functionals contain an MP2-like term that accounts for dispersion. These functionals also appear well suited to describe noncovalent interactions. The B2PLYP and XYG3 functionals are two examples of double hybrid functionals that have been shown to perform well.[29,70,192,231,249]

1.2.1 Computational Considerations

In this section, we will highlight some of the important considerations regarding the computation of noncovalent interactions.

1.2.1.1 One-Particle Basis Sets

The proper choice of one-particle basis is inexorably tied to both the choice of method and the system of interest. In this section, we will limit our discussion to standard WFN approaches (Møller-Plesset perturbation theory, coupled-cluster theory, *etc.*), which have well understood convergence behavior (the basis set dependence of DFT methods will be mentioned briefly in other sections and can be found elsewhere in great detail[29]).

In general, the use of large, diffuse basis sets is recommended for computing noncovalent interactions.[78,105,123,225,227] This is because the dispersion energy, a dynamical electron correlation effect, is known to be slowly convergent with respect to the size of basis.[76] Additionally, the dispersion energy is related to the molecular polarizability, which is known to benefit from the inclusion of diffuse basis functions. Since large basis coupled-cluster computations cannot be routinely performed, extrapolation schemes are utilized to remove the basis set incompleteness error in the interaction energy with lower-scaling methods (usually second-order Møller-Plesset perturbation theory); higher-order correlation effects are then evaluated in a smaller basis.[101,204]

1.2.1.2 Basis Set Superposition Error

The use of a finite basis set for supermolecular computations of intermolecular interactions results in the so-called basis set superposition error (BSSE). Typically, the computation of a dimer energy uses atom-centered basis functions on the nuclei of both monomers. If the monomer computations use basis functions centered only on the nuclei of that monomer, the basis used to describe the dimer will have more flexibility. The most common solution to this problem is the Boys-Bernardi counterpoise (CP) correction, where the monomer energies are computed in a dimer centered basis,[21]

$$E_{\text{interaction}}^{\text{CP}} = E_{\text{dimer}}^{\text{dimer}} - E_{\text{monomer}_\text{A}}^{\text{dimer}} - E_{\text{monomer}_\text{B}}^{\text{dimer}}, \tag{5}$$

where the superscripts define the basis in which the energy is computed. This gives rise to the typical definition of BSSE:

$$E_{\mathrm{BSSE}} = (E^{\mathrm{dimer}}_{\mathrm{monomer_A}} - E^{\mathrm{monomer_A}}_{\mathrm{monomer_A}}) + (E^{\mathrm{dimer}}_{\mathrm{monomer_B}} - E^{\mathrm{monomer_B}}_{\mathrm{monomer_B}}). \quad (6)$$

This definition for BSSE is not exact and has proven to be a source of much contention.[136,177,230]

Typically, qualitative descriptions of BSSE involve one monomer "stealing" the basis functions of the other monomer. This description does not sufficiently reflect the complexity of BSSE and seems to suggest that the CP correction is an exact solution to BSSE. A better description of BSSE is that the finite basis sets used for the dimer and monomers are not uniformly flexible. If monomers are treated in a monomer-centered basis set, there is more flexibility to describe the monomer (in the dimer computation) in the dimer-centered basis. Unfortunately, if the monomers are treated in a dimer-centered basis, then there is less flexibility to treat the monomers in the dimer computation. This is a result of the Pauli exclusion principle; in a finite basis, there is more flexibility to describe N_A electrons than there is to describe $N_A + N_B$ electrons. As a result, the CP correction overcorrects the interaction energy for BSSE.

In certain circumstances, the convergence of unCP corrected interaction energies with respect to the size of the basis set is more rapid than the convergence of CP corrected interaction energies.[78,224] In some cases, the CP corrected interaction energies tend to underestimate estimated complete basis interaction energies with the same magnitude as unCP corrected results overestimate it. This has lead to the averaging of CP and unCP corrected results.[120] In general, the convergence of CP corrected interaction energies with respect to the basis set is smooth, but slow. The convergence of unCP corrected results is more unpredictable; it can be more rapid, slower, or oscillatory. Interaction energies that are obtained with methods that are free from BSSE (*i.e.* SAPT) typically follow the same convergence behavior as CP corrected energies.

1.2.1.3 Interaction and Binding Energies

There is an important distinction between the interaction energy of a complex and the binding energy of a complex. Unfortunately, these are often used interchangeably in the

literature. One main distinction involves the geometries of the monomers; the other distinction is a matter of convention. Interaction energies do not include the effect of geometric distortion and attractive interactions are defined to be negative.

$$E_{\text{interaction}} = E_{\text{dimer}}^{\text{complex}} - E_{\text{monomer}_\text{A}}^{\text{complex}} - E_{\text{monomer}_\text{B}}^{\text{complex}} \tag{7}$$

Here, the subscript labels each energy with its corresponding moiety and the superscript denotes the origin of each geometry. In contrast to interaction energies, binding energies include geometric effects and favorable binding energies are defined to be positive.

$$E_{\text{binding}} = E_{\text{monomer}_\text{A}}^{\text{isolated}} + E_{\text{monomer}_\text{B}}^{\text{isolated}} - E_{\text{dimer}}^{\text{complex}} \tag{8}$$

Interaction energies are used to describe the energetic changes of a complex at a fixed geometry due to noncovalent interactions. Using the partitioned Hamiltonian from intermolecular perturbation theory, the interaction energy is the difference between the dimer energy with contributions from the intermolecular operator, V, and without it. Binding energies are more physically motivated, they describe the change in energy when isolated monomers are brought together to form a complex. This includes the change in energy due to the distortion of the isolated monomer geometry to its geometry in the complex.

Often, it is possible to compute energies at a higher level of theory than geometries. In many cases, reasonable geometries can be obtained with methods that will not produce particularly reliable interaction energies. In these situations, it is useful to decouple the interaction energy and geometric effects that are contained in the binding energy.

$$E_{\text{binding}} = (E_{\text{monomer}_\text{A}}^{\text{isolated}} - E_{\text{monomer}_\text{A}}^{\text{complex}}) + (E_{\text{monomer}_\text{B}}^{\text{isolated}} - E_{\text{monomer}_\text{B}}^{\text{complex}}) - E_{\text{interaction}} \tag{9}$$

In this picture, the geometric distortion effects are treated separately, which allows the energetic changes resulting geometric considerations to be treated at a different level of theory from the interaction energy. It is important to note that a consistent level of theory MUST be used to compute the monomer and complex geometries and the geometric distortion energy. Otherwise, it is possible for $E_{\text{monomer}}^{\text{isolated}} - E_{\text{monomer}}^{\text{complex}}$ to be positive, which is unphysical. Additionally, this decoupling of geometric and energetic effects proves to be especially useful for correcting binding energies for BSSE. If BSSE corrected interaction

energies and geometric distortion energies are computed entirely in the monomer basis, a BSSE corrected binding energy can be obtained.

Binding energies are the preferred quantity for comparisons with experiment. Beyond what has been described above, zero-point vibrational energy (ZPVE) corrections should be included in the binding energy. For computations of thermodynamic quantities (ΔH, ΔG, *etc.*), finite temperature corrections from the molecular partition functions are required. These can be obtained from the computations required to obtain the ZPVE. Typically, harmonic potentials are used to obtain these corrections, however, noncovalent complexes may have strongly anharmonic potentials necessitating a more robust treatment.

1.2.2 Wavefunction-based Methods

In this section, we will highlight a few methods that were designed to treat noncovalent interactions.

1.2.2.1 Spin-Component Scaled Methods

In correlated wavefunction theories, the most important excited Slater determinants (beyond the Hartree–Fock reference determinant) are usually the doubly-excited ones. If the two excited electrons have the same spin, these correlation amplitudes might be called "same spin" terms, otherwise they might be called "opposite spin" terms. Noting that the quality of the energetic contributions from these two types of terms may differ in approximate *ab initio* methods, Grimme[67] proposed to simply scale these contributions by different factors. This led to the spin-component-scaled MP2 (SCS-MP2) method, with an energy defined as

$$E_{SCS-MP2} = E_{HF} + p_{SS}E^{corr}_{SS-MP2} + p_{OS}E^{corr}_{OS-MP2}, \tag{10}$$

where E_{HF} is the Hartree–Fock energy, E^{corr}_{SS-MP2} is the same-spin component of the MP2 correlation energy, E^{corr}_{OS-MP2} is the opposite-spin component of the MP2 correlation energy, and p_{SS} and p_{OS} are the two scale-factors. The addition of these two parameters means that the method acquires a somewhat "semi-empirical" flavor, although this is perhaps not an accurate description, because Grimme fit these parameters to a set of 51 high-quality CCSD(T)/cc-pVQZ reaction energies, rather than to experimental data. In this

sense, it is perhaps more accurate to describe this as a parameterized first-principles model. Because the parameters were obtained by fitting to reaction energies, it is not surprising that the method exhibits improved results for reaction energies. However, it also exhibits significantly improved results for noncovalent interactions, and this (along with improved predictions for atomization energies, geometries, and vibrational frequencies)[67] suggests that the scaling is truly a general improvement.

SCS-MP2 was highlighted as a promising alternative to expensive CCSD(T) computations for noncovalent interactions, especially after a very impressive demonstration of errors less than 0.2 kcal mol^{-1} for the very difficult benzene dimer.[89] Other studies showed definite improvements over canonical MP2 for various noncovalent interactions.[60,215] Hill and Platts re-parameterized the model specifically for noncovalent interactions in 2007,[87] introducing a variant denoted SCSN-MP2. Similarly, DiStasio and Head-Gordon introduced the spin-component-scaled for molecular interactions MP2, or SCS(MI)-MP2, which uses yet another pair of scale factors obtained by fits to the S22 set.[44] SCS(MI)-MP2 parameters were fit separately for each Dunning cc-pVXZ basis set, which can help improve results when using smaller basis sets; also, fitting to non-augmented basis sets means that the computations are less expensive.

Antony and Grimme[6] evaluated the reliability of SCS-MP2 for the JSCH test set.[117] Rather good results were obtained using SCS-MP2 in conjunction with triple-ζ basis sets (TZVPP or cc-pVTZ) without the need for counterpoise correction (due to favorable error cancellation). However, the authors noted that interaction energies for n-alkane dimers were poor for SCS-MP2, because canonical MP2 tends to do well for these, and the scaling procedure spoils the good agreement. Similarly, scaling also provides poorer results for H-bonded systems. Mixed results for SCS-MP2 vs. conventional MP2 were found by King for the ethylene dimer.[121] The three main variants of SCS-MP2 have also been evaluated against the NBC10 test set of potential energy curves,[198,215] and against the S22x5 test set,[65] where they exhibit errors of several tenths up to one kcal mol^{-1}. SCS(MI)-MP2 was found to be somewhat better than SCS-MP2 or SCSN-MP2 for these tests. All three SCS-MP2 variants underbound methane dimer by a significant fraction of the binding energy

(although the error remains only tenths of one kcal mol^{-1} on an absolute scale).[198] Evaluations of SCS(MI)-MP2 by Hobza and co-workers[179] for potential curves of 7 additional dimers provided similar results.

Reasoning that spin-component scaling could work even better for a more robust wavefunction, Takatani et al.[214] introduced spin-component scaled coupled-cluster singles and doubles (SCS-CCSD), in which the double excitation amplitudes are scaled just as in MP2. The two scaling parameters were fit to reaction energies, using most of the reactions employed by Grimme to fit SCS-MP2.[67] The resulting method is of very high quality, especially for noncovalent interactions. It is robust for cases where scaling degrades the quality of MP2 (as in the methane dimer),[214] and it exhibits a mean absolute deviation of only 0.24 kcal mol^{-1} for the S22A test set.[213] Hobza and co-workers have proposed five statistical criteria which are desirable for methods treating noncovalent interactions, and out of a large pool of promising approaches, the SCS-CCSD method was essentially tied with MP2C (see below) for providing the best performance for the S22x5 test set.[65] Even better performance by SCS-CCSD for noncovalent interactions was realized by Pitoňák, Řezáč, and Hobza by reparameterizing against the S22 test.[169] This new method, SCS(MI)-CCSD, has a remarkable mean absolute deviation of only 0.03 kcal mol^{-1} for the noncovalent databases of Zhao and Truhlar.[252] The disadvantage of the approach is, of course, that CCSD computations are expensive, scaling as $\mathcal{O}(o^2v^4)$. However, local correlation, density fitting, and other techniques can already reduce this cost significantly;[79,189,191] the local density-fitted CCSD implemented by Schütz and Manby has been used for systems as large as $(\mathrm{Gly})_{16}$.[190]

Pitoňák *et al.* also pursued scaling in conjunction with accurate wavefunction theories.[167] Their MP2.5 method improves upon MP2 by adding the third-order correlation energy, scaled by a fitted scaling factor. Their initial tests indicated that good results were obtained when adding one-half of the third-order term (hence the name MP2.5). This approach is somewhat less expensive than SCS-CCSD and provides results which are not quite as good, but which improve over various forms of SCS-MP2 for the S22x5 test set.[65] Roughly, the cost of MP2.5 is equivalent to a single CCSD iteration, with a computational cost $\mathcal{O}(o^2v^4)$. MP2.5 computations were reported for the porphine dimer, demonstrating that

it is applicable to much larger systems than those accessible by CCSD(T).[167]

1.2.2.2 MP2C

The poor performance of MP2 for π-π interactions is related to its overestimation of the dispersion energy, which can be written as

$$E_{disp}^{(2)} = -\frac{1}{2\pi}\int_0^\infty d\omega \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4 \chi_A(\mathbf{r}_1, \mathbf{r}_3, \omega)\chi_B(\mathbf{r}_2, \mathbf{r}_4, \omega)\frac{1}{r_{12}}\frac{1}{r_{34}}, \tag{11}$$

where χ is a monomer response function. In MP2, these are uncoupled Hartree–Fock (UCHF) response functions. In the context of SAPT(DFT), accurate dispersion energies are obtained by evaluating Equation 11 with coupled response functions obtained from time-dependent density functional theory (TDDFT).[147] Heßelmann has proposed to use the difference between $E_{disp}^{(2)}$ evaluated with UCHF and TDDFT response functions (ΔMP2C = $E_{disp}^{(2)}$[TDDFT]$-E_{disp}^{(2)}$[UCHF]) as a correction to supermolecular MP2 interaction energies.[84] The resulting coupled MP2 method (MP2C) is more physically justified than MP2 and provides similar accuracy to SCS-CCSD and MP2.5.[65] Additionally, the evaluation of the ΔMP2C correction scales as $\mathcal{O}(n^4)$, therefore, in principle, this method is only as expensive as a supermolecular MP2 interaction energy, $\mathcal{O}(n^5)$.[165] At present, this method appears to be the most promising for accurately treating π-π interactions in extended systems.

1.2.2.3 SAPT

In SAPT, the dimer Hamiltonian is partitioned into contributions from the Fock operator of each monomer (F), the interaction between the monomers (V), and the fluctuation potential of each monomer (W).

$$H = F_A + F_B + V + W_A + W_B \tag{12}$$

The interaction energy can be written as a perturbation series,

$$E_{int} = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(E_{pol}^{(nkl)} + E_{exch}^{(nkl)}), \tag{13}$$

where n denotes the order in V, and k and l denote the order in W_A and W_B, respectively. Here, the E_{pol} terms originate from the *polarization expansion* and E_{exch} are repulsive terms resulting from the antisymmetry of the wavefunction with respect to the exchange

of electrons between monomers. A more complete description of SAPT can be found in Reference 111.

The SAPT series can be carried out to various degrees of completeness depending on the size of the system being studied and the accuracy desired. Historically, several truncations of this series have been defined.

$$E_{SAPT0} = E_{HF} + E^{(20)}_{disp} + E^{(20)}_{exch-disp} \tag{14}$$

$$E_{SAPT2} = E_{SAPT0} + E^{(12)}_{elst,r} + \epsilon^{(1)}_{exch}(2) + {}^{t}E^{(22)}_{ind} + {}^{t}E^{(22)}_{exch-ind} \tag{15}$$

$$E_{SAPT} = E_{SAPT0} + E^{(12)}_{elst,r} + E^{(13)}_{elst,r} + \epsilon^{(1)}_{exch}(\mathrm{CCSD}) + {}^{t}E^{(22)}_{ind} + {}^{t}E^{(22)}_{exch-ind} + E^{(21)}_{disp} + E^{(22)}_{disp} \tag{16}$$

Notation for the SAPT terms follows from Equation 13, where the $E^{(vw)}$ defines the order in V and in $W_A + W_B$. E_{HF} is the Hartree–Fock interaction energy. In Equation 15, $\epsilon^{(1)}_{exch}(2) = E^{(11)}_{exch} + E^{(12)}_{exch}$, whereas in Equation 16, $\epsilon^{(1)}_{exch}(\mathrm{CCSD})$ refers to the intramonomer correlation correction to exchange evaluated with converged CCSD amplitudes. The r subscript indicates that contributions due to orbital response are included.

The methods defined in Equations 14-16 reflect the mid–1990s state of the art in both computer implementation and theoretical development of SAPT. Since then, advances in both areas have made other groupings of SAPT terms (and the inclusion of additional higher-order terms) more practical:

$$E_{SAPT2+} = E_{SAPT2} + E^{(21)}_{disp} + E^{(22)}_{disp}, \tag{17}$$

$$E_{SAPT2+(3)} = E_{SAPT2+} + E^{(13)}_{elst,r} + E^{(30)}_{disp}, \tag{18}$$

$$E_{SAPT2+3} = E_{SAPT2+} + E^{(13)}_{elst,r} + E^{(30)}_{disp} + E^{(30)}_{exch-disp} + E^{(30)}_{ind-disp} + E^{(30)}_{exch-ind-disp}. \tag{19}$$

The introduction of density fitting has made the evaluation of the intramonomer correlation corrections more efficient[94] (although this performance gain is lost if CCSD amplitudes are employed as in equation 16). To compute accurate dispersion energies, one must include the triples contribution, which scales as $\mathcal{O}(N^7)$. Fortunately, natural orbitals can be used to dramatically reduce the cost of this term.[95] Following from Reference 162, the triples contribution to $E^{(30)}_{exch-disp}$ is neglected due to its $\mathcal{O}(N^7)$ scaling and relative unimportance within $E^{(30)}_{exch-disp}$.

Often a δE_{HF} term is defined as the difference between the HF interaction energy and SAPT terms that it contains:

$$\delta E_{HF}^{(2)} = E_{int}^{HF} - \left(E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,r}^{(20)} + E_{exch-ind,r}^{(20)} \right), \tag{20}$$

$$\delta E_{HF}^{(3)} = E_{int}^{HF} - \left(E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,r}^{(20)} + E_{exch-ind,r}^{(20)} + E_{ind}^{(30)} + E_{exch-ind}^{(30)} \right). \tag{21}$$

This term is sometimes included as a way to capture some higher-order terms not explicitly evaluated by SAPT. In second-order SAPT, it is often helpful to include $\delta E_{HF}^{(2)}$, as it incorporates certain third-order terms. However, for third-order SAPT methods, it is sometimes better to use a pure SAPT approach and to omit the $\delta E_{HF}^{(3)}$ correction.[162] The δE_{HF} term contains higher-order induction and exchange-induction interactions as well as some unphysical exchange effects.[109,153] In dimers containing polar molecules, the higher-order induction effects dominate the δE_{HF} term and it should be included in the interaction energy. For nonpolar molecules, the induction effects are relatively unimportant and more accurate results can be obtained by omitting this term.[162]

A useful byproduct of the perturbative expansion of the interaction energy is the decomposition into physically meaningful components (i.e., electrostatics, exchange, induction, and dispersion).

$$E_{electrostatic} = E_{elst}^{(10)} + E_{elst,r}^{(12)} + E_{elst,r}^{(13)} \tag{22}$$

$$E_{exchange} = E_{exch}^{(10)} + E_{exch}^{(11)} + E_{exch}^{(12)} \tag{23}$$

$$E_{induction} = E_{ind,r}^{(20)} + E_{exch-ind,r}^{(20)} + E_{ind}^{(30)} + E_{exch-ind}^{(30)} + {}^{t}E_{ind}^{(22)} + {}^{t}E_{exch-ind}^{(22)} + \delta E_{HF}^{(3)} \tag{24}$$

$$E_{dispersion} = E_{disp}^{(20)} + E_{disp}^{(30)} + E_{disp}^{(21)} + E_{disp}^{(22)} + E_{exch-disp}^{(20)} + E_{exch-disp}^{(30)} \tag{25}$$

The above grouping of $E_{exch-ind}$ and $E_{exch-disp}$ terms with induction and dispersion, respectively, rather than exchange is somewhat arbitrary and represents a chemical interpretation of the terms rather than a mathematical one. The E_{ind} and E_{disp} terms are artificially lowered in energy by unphysical contributions that violate the Pauli exclusion principle. Their exchange counterparts remove these unphysical terms and allow for a more intuitive interpretation of the SAPT decomposition. Rigorously, the $E_{ind-disp}^{(30)}$ term cannot be grouped as induction or dispersion. Physically, this term can be interpreted as the change in dispersion

due to the induced orbital deformation, so, for simplicity, $E^{(30)}_{ind-disp}$ and $E^{(30)}_{exch-ind-disp}$ will be added to $E_{dispersion}$.

The zeroth-order wavefunction in SAPT is a product of monomer wavefunctions. The monomer wavefunctions can be computed in either a "monomer centered basis" (MCB) or a "dimer centered basis" (DCB), where the latter includes the basis functions from the other monomer (as used in counterpoise correction). In practice, the difference in the choice of basis affects only a few parts of the computation (ignoring the obvious effect on basis set completeness). One of these effects is on exchange; there are two approaches to deriving the exchange terms. One approach uses interaction density matrices and leads to equations valid in either choice of basis, the other approach uses second quantization and leads to equations valid only in the dimer centered basis. Another important consequence of the choice of basis is related to the inclusion of charge-transfer excitations. In the dimer centered basis, the induction terms will include charge-transfer. In the monomer centered basis, the space required for charge-transfer excitations is no longer included. Misquitta and Stone have shown that the charge-transfer energy can be computed with SAPT as:[209]

$$E_{charge-transfer} = \left(E^{(20)}_{ind,r}(\mathrm{DCB}) + E^{(20)}_{exch-ind,r}(\mathrm{DCB})\right) - \left(E^{(20)}_{ind,r}(\mathrm{MCB}) + E^{(20)}_{exch-ind,r}(\mathrm{MCB})\right). \tag{26}$$

Hybrids of the dimer and monomer centered basis, where a basis contains all the functions of one monomer and select functions from the other, are also valid. It is also possible to include bond functions in addition to the monomer or dimer centered basis.

1.3 Approximations for Electron Repulsion Integrals

The use of approximate representations of two-electron integrals has become popular as a means to speed up their evaluation and reduce storage requirements. There are several closely related approaches to approximate two-electron integrals. The two discussed in this work are the density fitting approximation (DF, also called resolution-of-the-identity or RI)[47,48,55,178,228,236,239,244] and the Cholesky decomposition (CD).[9,16,125,186] In practice, both methods approximate the two-electron, four-index quantities as a linear combination

of three-index quantities. The DF approach utilizes three-index integrals where one index corresponds to a pre-optimized auxiliary basis set of atom-centered Gaussians (the auxiliary basis is typically a few times larger than the size of the chosen AO basis set). This is very similar to the pseudospectral approximation,[57,139] which evaluates the third index on a grid. The CD approach guarantees the AO integrals to a certain accuracy, and is independent of particular electronic structure method. It has a slightly larger overhead associated with the computation of AO three-index quantities as compared to DF, but the result is an unbiased, method-independent approximation of the two-electron integrals A general comparison of DF and CD methods has recently been published by Weigend, Kattannek, and Ahlrichs.[237]

1.3.1 Density Fitting

The two-electron integrals in the DF approximation are given by

$$(\mu\nu|\rho\sigma) \approx \sum_{PQ} (\mu\nu|P)[J^{-1}]_{PQ}(Q|\rho\sigma). \tag{27}$$

The two index quantity, $[J^{-1}]_{PQ}$, is the inverse of the Coulomb metric evaluated in an auxiliary basis set:

$$[J]_{PQ} = \int P(\mathbf{r}_1)\frac{1}{r_{12}}Q(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2. \tag{28}$$

Ignoring any sparsity due to large distances between centers, there are $\mathcal{O}(N_{df}N_{ao}^2)$ three-index integrals in the DF approach, compared to $\mathcal{O}(N_{ao}^4)$ two-electron integrals. It is convenient to rewrite Equation 27 with different three-index quantities,

$$\widetilde{(\mu\nu|Q)} = \sum_{P} (\mu\nu|P)[J^{-\frac{1}{2}}]_{PQ}, \tag{29}$$

$$(\mu\nu|\rho\sigma) \approx \sum_{Q} \widetilde{(\mu\nu|Q)}\widetilde{(Q|\rho\sigma)} \tag{30}$$

In SAPT, the three index terms must be transformed into the molecular orbital basis of each monomer,

$$\widetilde{(ij|Q)} = \sum_{\mu\nu} C_{\mu i}^{M\dagger}\widetilde{(\mu\nu|Q)}C_{\nu j}^{N} \tag{31}$$

where C^M and C^N represent the SCF coefficient matrices of monomers M and N, and where i and j are MOs resulting from the Hartree-Fock computations on monomers M and

N, respectively. All the two-electron integrals necessary in SAPT can be formed from the $\widetilde{(ij|Q)}$ quantities through the MO basis analogue of Equation 30.

1.3.2 Cholesky Decompositions

The intermediate quantities that result from a Cholesky decomposition (CD), $L^Q_{\mu\nu}$, are analogous to the $\widetilde{(\mu\nu|Q)}$ three center integrals:

$$(\mu\nu|\rho\sigma) \approx \sum_Q L^Q_{\mu\nu} L^Q_{\rho\sigma}. \tag{32}$$

The Cholesky vectors are defined recursively through

$$L^{\rho\sigma}_{\rho\sigma} = \sqrt{(\rho\sigma|\rho\sigma) - \sum_{n=0}^{\rho\sigma_{max}-1} L^n_{\rho\sigma} L^n_{\rho\sigma}}, \tag{33}$$

$$L^{\rho\sigma}_{\mu\nu} = \frac{1}{L^{\rho\sigma}_{\rho\sigma}} \left((\mu\nu|\rho\sigma) - \sum_{n=0}^{\rho\sigma_{max}-1} L^n_{\mu\nu} L^n_{\rho\sigma} \right), \mu\nu \neq \rho\sigma. \tag{34}$$

The Cholesky vectors can be transformed for SAPT using an analogue of Equation 31.

$$L^Q_{ij} = \sum_{\mu\nu} C^{M\dagger}_{\mu i} L^Q_{\mu\nu} C^N_{\nu j} \tag{35}$$

The atomic orbital two-electron integrals can be approximated to an accuracy of δ using the Cholesky procedure. If the $(\rho\sigma|\rho\sigma)$ integrals are ordered from largest to smallest, only the Cholesky vectors where $L^{\rho\sigma}_{\rho\sigma} L^{\rho\sigma}_{\rho\sigma} > \delta$ need to be formed. A further approximation can be made to the Cholesky decomposition by including only $\rho\sigma$ pairs where ρ and σ are centered on the same nucleus. This is referred to as a one-center Cholesky decomposition (1C-CD).[7,8] Although formally it removes the error bound on the approximate integrals, in practice, the error made in the one-center approximation is minimal.

1.3.3 Density Fitting in Symmetry-Adapted Perturbation Theory

The DF approximation has been applied to SAPT in the context of SAPT(DFT).[27,86,170] This approximation reduces the bottleneck associated with the evaluation of the dispersion term from the SAPT(DFT) computation and allows the method to be applicable to larger systems. SAPT(DFT) computations have been performed on the benzene dimer using an aug-cc-pVQZ basis set, which includes more than 1500 basis functions.[86] One of the largest

systems studied with the SAPT(DFT) method is the $2(H_2)$-C_{60} complex with a TZVPP basis set by Korona *et al.*[127] There are several considerations for SAPT(DFT) computations that are not necessary in wavefunction based SAPT. To produce reasonable interaction energies, SAPT(DFT) requires an asymptotic correction. One of the more widely used corrections requires the ionization potentials of the monomers. The evaluation of the dispersion interaction in SAPT(DFT) scales as $\mathcal{O}(N^6)$, and density fitting can reduce this to $\mathcal{O}(N^4)$ for pure density functionals, and to $\mathcal{O}(N^5)$ with hybrid functionals (including Hartree-Fock exchange). The hybrid functionals are usually more accurate,[27] but the available implementations of DF-SAPT(DFT) do not include exact (Hartree-Fock) exchange when the dispersion term is evaluated.[25,151] To circumvent this problem, the dispersion energy can be evaluated with an LDA kernel; this introduces less than 1% deviation from the dispersion evaluated with GGA kernels.[148,170] This approximation recovers the accuracy of the hybrid functionals, through the use of the hybrid GGA orbitals, while still scaling $\mathcal{O}(N^4)$. Finally, the inclusion of induced-multipole induced-multipole interactions in SAPT is not done explicitly, but rather by computing a δHF term. This term is computed as the difference between the SAPT0 energy (excluding dispersion) and the HF interaction energy. This cannot be computed from SAPT(DFT), and if it is needed, a separate wavefunction-based SAPT0 computation is required.

Recently, we have developed an SAPT program that uses the density fitting (DF) approximation to evaluate the necessary two-electron integrals.[93] We demonstrated that great speedups could be achieved simply by replacing the conventional integral transformation with one employing approximate DF integrals. The possibility of using the three-index integrals directly in the energy evaluations was not exploited in our initial implementation. For computations with less than 1000 basis functions, this is not necessary and, in some cases, it can be more efficient not to use the three-index integrals. However, when large SAPT0 computations are considered (i.e., >1500 basis functions), the disk I/O associated with four-index arrays begins to dominate, and it is advantageous to introduce three-index quantities where ever possible. Additionally, we have performed the largest SAPT computations to date that account for the triples correction to dispersion with this DF-SAPT

program.[94] While these advances have made SAPT computations much faster, the computations nevertheless remain very demanding if triples are included, and for such computations, our DF-SAPT program remains limited to systems roughly the size of a nucleic acid base pair (with a double-ζ basis set). The triples dispersion correction includes terms that scale as $\mathcal{O}(o^3v^4)$ and $\mathcal{O}(o^4v^3)$ (where o is the number of occupied orbitals and v is the number of virtual or unoccupied orbitals). Since the number of virtual orbitals is usually much larger than the number of occupied orbitals (this is required for an accurate description of dispersion), the overall scaling will be $\mathcal{O}(o^3v^4)$.

1.4 Prototypical Noncovalent Complexes

In this section we will discuss prototypes for different classes of noncovalent interactions. We will address the important interaction components for each complex and considerations for analyzing SAPT results for these types of interactions.

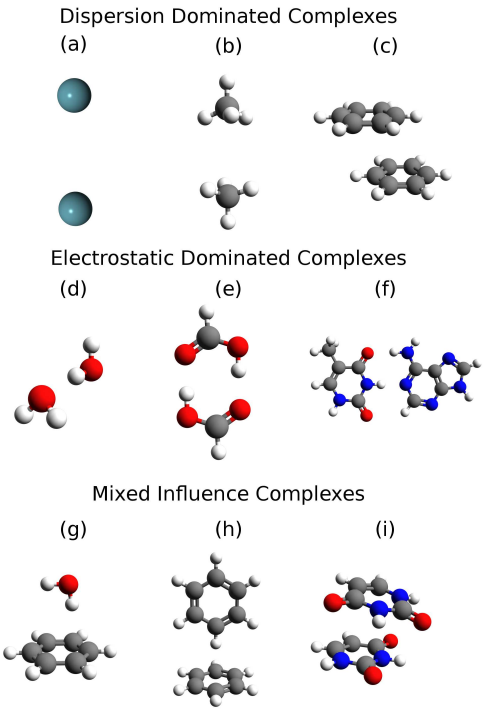
1.4.1 Dispersion Dominated Complexes

Certain noncovalent complexes that are characterized by the presence of strong London dispersion forces and the absence of strong electrostatic or inductive forces are classified as dispersion dominated; these complexes typically include neutral, nonpolar molecules. The interactions between rare gases, hydrocarbons, and base pair stacking in DNA all fall under the heading of dispersion dominated. These interactions are typically weaker than electrostatic dominated interactions and are non-directional.

1.4.1.1 Argon Dimer

The argon dimer is a prototypical dispersion bound complex. Here, we will consider a near equilibrium geometry: 3.75 Å separation. The primary attractive interactions are London dispersion forces; the lack of permanent multipole moments results in weak electrostatic interactions (at long range) and weak inductive forces. SAPT2+(3)/aug-cc-pVTZ computations (Table 1) quantify the forces at play in the argon dimer. Indeed, the dispersion energy is more than a factor of 3 larger than the electrostatic energy and more than a factor of 20 larger than the induction energy. At first glance, the electrostatic energy is larger

Figure 1: The geometries of prototypical noncovalent complexes studied in this work: (a) argon dimer, (b) methane dimer, (c) π-stacked benzene dimer, (d) water dimer, (e) formic acid dimer, (f) Watson-Crick adenine thymine, (g) water-benzene, (h) T-shaped benzene dimer, and (i) π-stacked uracil dimer.



than might have been expected. This is a result of charge penetration; orbital overlap between the argon atoms leads to a net attractive electrostatic interaction. It is interesting to note that the leading induction term, $E_{ind,r}^{(20)}$, is almost exactly cancelled by the leading exchange-induction term, $E_{exch-ind,r}^{(20)}(S^2)$. This behavior is fairly typical for interactions between nonpolar molecules (*e.g.* compare Tables 1 and 3). For this reason, we recommend classifying exchange-induction terms as part of the total induction interaction. This is even more important when third-order interactions are considered; $E_{ind}^{(30)}$ and $E_{exch-ind}^{(30)}$ terms often diverge, however, their sum remains physically reasonable.

Argon dimer makes a useful test case, because it is both small and difficult to describe with low levels of theory. Its small size allows the convergence of each of the SAPT terms to be analyzed with respect to the size of the basis set (see Table 2). Immediately, it can be seen that the electrostatic and exchange terms are relatively insensitive to the choice

Table 1: SAPT2+(3)/aug-cc-pVTZ results for dispersion dominated complexes in kcal mol^{-1}.

	$(Ar)_2$	$(CH_3)_2$	$(C_6H_6)_2$ Stacked
$E^{(10)}_{elst,r}$	-0.12	-0.14	-2.86
$E^{(12)}_{elst,r}$	-0.02	-0.03	0.07
$E^{(13)}_{elst,r}$	0.01	0.01	0.27
$E^{(10)}_{exch}(S^2)$	0.42	0.53	9.13
$E^{(11)}_{exch}(S^2)$	0.01	0.02	0.57
$E^{(12)}_{exch}(S^2)$	0.02	0.04	-0.75
$E^{(20)}_{ind,r}$	-0.15	-0.06	-4.29
$E^{(22)}_{ind}$	-0.02	-0.01	0.28
$E^{(20)}_{exch-ind,r}(S^2)$	0.15	0.06	4.01
$E^{(22)}_{exch-ind}(S^2)$	0.02	0.01	-0.27
δHF	-0.02	-0.02	-0.64
$E^{(20)}_{disp}$	-0.58	-0.96	-12.00
$E^{(30)}_{disp}$	0.01	0.02	0.41
$E^{(21)}_{disp}$	0.10	0.00	3.71
$E^{(22)}_{disp}$(SDQ)	0.01	0.06	0.68
$E^{(22)}_{disp}$(T)	-0.08	-0.12	-2.74
$E^{(20)}_{exch-disp}(S^2)$	0.03	0.06	1.74
E_{elst}	-0.14	-0.15	-2.52
E_{exch}	0.45	0.59	8.96
E_{ind}	-0.02	-0.03	-0.90
E_{disp}	-0.50	-0.94	-8.21
E_{SAPT0}	-0.27	-0.53	-4.91
E_{SAPT2}	-0.25	-0.50	-4.99
$E_{\mathrm{SAPT2+}}$	-0.22	-0.56	-3.35
$E_{\mathrm{SAPT2+(3)}}$	-0.21	-0.53	-2.67

of basis; the induction terms are slightly more sensitive. For the argon dimer, reasonable estimates for the electrostatic, exchange, and induction terms can be obtained in a small, cc-pVTZ basis. The dispersion energy, however, converges very slowly with respect to basis set size and is extremely sensitive to the inclusion of diffuse basis functions. The aug-cc-pVTZ basis (100 functions) recovers a larger fraction of the dispersion energy than the large, cc-pV5Z basis (190 function). Consequently, diffuse basis functions must be included to effectively study dispersion bound complexes.

It is also interesting to examine the convergence and relative importance of the intramonomer correlation corrections. The intramonomer corrections to electrostatics, exchange, and induction are relatively unimportant, however, their convergence with respect to basis differs from the rate of the leading terms significantly. For example, the $E_{exch}^{(12)}(S^2)$ term changes by 0.06 kcal mol^{-1} while the $E_{exch}^{(10)}$ term does not change at all. This is because electron correlation effects are slowly convergent (*i.e.* the $E_{exch}^{(12)}(S^2)$ term) while $E_{exch}^{(10)}$ is a Hartree-Fock-like term that converges rapidly with respect to basis. The change in sign is a result of $E_{exch}^{(12)}(S^2)$ being composed of five separate terms, each with their own convergence behavior. The intramonomer corrections to dispersion are much more important. At aug-cc-pV5Z, the net intramonomer correction should be 0.04 kcal mol^{-1}. Unfortunately, the terms that scale $\mathcal{O}(N^6)$, $E_{disp}^{(21)}$ and $E_{disp}^{(22)}$(SDQ), are not a good approximation to the intramonomer correction. The expensive ($\mathcal{O}(N^7)$) $E_{disp}^{(22)}$(T) term must be included for a balanced treatment of the intramonomer electron correlation.

1.4.1.2 Methane Dimer

The methane dimer interaction is composed of an even larger fraction of dispersion than the argon dimer interaction. The methane dimer configuration considered here is reported in Ref. 117. In addition to being larger, the dispersion energy in the methane dimer has different properties with regard to the inclusion of intramonomer electron correlation. In this case, the intramonomer correction to dispersion tends to stabilize the complex; again, the $E_{disp}^{(22)}$(T) correction is important for the accurate determination of the dispersion energy. It is interesting to note that the exchange and dispersion energies for the methane dimer are significantly larger than for the argon dimer, but the electrostatics are nearly equivalent. This is a result of the permanent multipole moments possessed by methane, but not argon. A multipole based electrostatic analysis would predict a repulsive potential, which would be repulsive by 0.11 kcal mol^{-1} at the geometry considered here. The charge penetration effects are able to overcome this repulsion resulting in a net attractive electrostatic term.

Table 2: SAPT results for the argon dimer computed with various basis sets.[a] Energies are reported in kcal mol^{-1}.

	DZ	TZ	QZ	5Z	aDZ	aTZ	aQZ	a5Z
$E^{(10)}_{elst,r}$	-0.12	-0.13	-0.12	-0.12	-0.13	-0.12	-0.12	-0.12
$E^{(12)}_{elst,r}$	-0.02	-0.02	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01
$E^{(13)}_{elst,r}$	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01
$E^{(10)}_{exch}$	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
$E^{(10)}_{exch}(S^2)$	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
$E^{(11)}_{exch}(S^2)$	-0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02
$E^{(12)}_{exch}(S^2)$	0.03	0.01	0.00	-0.01	0.05	0.02	0.00	-0.01
$E^{(20)}_{ind,r}$	-0.13	-0.16	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15
$E^{(22)}_{ind}$	-0.02	-0.02	-0.01	-0.01	-0.03	-0.02	-0.01	-0.01
$E^{(20)}_{exch-ind,r}(S^2)$	0.13	0.16	0.15	0.15	0.15	0.15	0.15	0.15
$E^{(22)}_{exch-ind}(S^2)$	0.02	0.02	0.01	0.01	0.03	0.02	0.01	0.01
δHF	-0.02	-0.01	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02
$E^{(20)}_{disp}$	-0.15	-0.36	-0.49	-0.54	-0.46	-0.58	-0.62	-0.65
$E^{(30)}_{disp}$	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01
$E^{(21)}_{disp}$	0.03	0.09	0.10	0.10	0.09	0.10	0.10	0.10
$E^{(22)}_{disp}$(SDQ)	0.00	0.01	0.02	0.03	-0.01	0.01	0.03	0.04
$E^{(22)}_{disp}$(T)	-0.01	-0.03	-0.06	-0.07	-0.04	-0.08	-0.09	-0.10
$E^{(20)}_{exch-disp}(S^2)$	0.01	0.02	0.03	0.03	0.03	0.03	0.04	0.04
E_{elst}	-0.13	-0.14	-0.13	-0.13	-0.15	-0.14	-0.13	-0.13
E_{exch}	0.44	0.44	0.44	0.43	0.48	0.45	0.44	0.43
E_{ind}	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
E_{disp}	-0.11	-0.27	-0.40	-0.44	-0.39	-0.50	-0.54	-0.56
E_{SAPT0}	0.15	-0.06	-0.18	-0.23	-0.16	-0.27	-0.31	-0.33
E_{SAPT2}	0.16	-0.05	-0.18	-0.23	-0.13	-0.25	-0.30	-0.33
$E_{\mathrm{SAPT2+}}$	0.18	0.01	-0.12	-0.17	-0.09	-0.22	-0.27	-0.29
$E_{\mathrm{SAPT2+(3)}}$	0.18	0.01	-0.11	-0.16	-0.08	-0.21	-0.25	-0.27

[a]XZ abbreviates cc-pVXZ and aXZ abbreviates aug-cc-pVXZ.

1.4.1.3 π-Stacked Benzene Dimer

The π-π interactions present in the benzene dimer are unique from both a geometric and an electronic structure perspective. The planar geometry of benzene allows the rings to get close together and interact strongly. Although the total interaction energy of the benzene dimer is only 5 times larger than the methane dimer, the individual components range from 8-30 times larger in the benzene dimer. This is a direct result of the benzene dimer geometry.

The dispersion energy is much more difficult to compute in the benzene dimer. This is a result of the motion of electrons in the π-orbitals. London dispersion forces originate from instantaneous charge fluctuations; in the case of the benzene dimer, these fluctuations can be very large due to the polarizability of the π-orbitals. The charge fluctuations must be allowed to damp each other, or their effect will be significantly overestimated. As a result, the intramonomer correlation corrections to dispersion for the benzene dimer are destabilizing by roughly 2 kcal mol^{-1}.

1.4.2 Electrostatic Dominated Complexes

Electrostatic dominated complexes typically involve two polar molecules. The most common examples include hydrogen bonds, but electrostatic dominated complexes are not limited to those involving hydrogen bonds. Halogen bonded complexes and those with strong dipolar interactions can all be classified as electrostatic dominated. Ionic complexes are not typically included in discussions of noncovalent interactions, although these are strongly bound by electrostatic interactions. Electrostatic dominated complexes, especially those with hydrogen bonds, often contain significant contributions from induction.

1.4.2.1 Water Dimer

The water dimer is a common example of a complex bound by electrostatic interactions. At this geometry (from Ref. 117), the electrostatic interactions are more than 3 times stronger than the induction and dispersion forces. In contrast to dispersion bound complexes, these electrostatics can be qualitatively described in a multipole picture. The multipole interactions would account for roughly -7.5 kcal mol^{-1} of the total electrostatics, however, this implies that charge penetration effects are responsible for the remaining 1 kcal mol^{-1}. The induction forces in the water dimer are much more important than they are in dispersion bound complexes. These strong inductive forces give rise to cooperativity in larger hydrogen bonded clusters. Dispersion is non-negligible and roughly equivalent to induction for the water dimer. The non-dispersion intramonomer correlation corrections are much more important for the water dimer than they are for dispersion bound complexes. In particular, the $E^{(12)}_{exch}(S^2)$ term is worth 1 kcal mol^{-1}.

Table 3: SAPT2+(3)/aug-cc-pVTZ results for electrostatic dominated complexes in kcal mol^{-1}.

	$(H_2O)_2$	$(CHO_2H)_2$	WC A-T
$E^{(10)}_{elst,r}$	-8.33	-33.66	-27.27
$E^{(12)}_{elst,r}$	0.06	1.37	0.52
$E^{(13)}_{elst,r}$	0.16	0.08	0.17
$E^{(10)}_{exch}(S^2)$	6.98	35.33	27.82
$E^{(11)}_{exch}(S^2)$	0.09	0.18	0.15
$E^{(12)}_{exch}(S^2)$	1.01	5.05	3.25
$E^{(20)}_{ind,r}$	-3.01	-21.09	-14.13
$E^{(22)}_{ind}$	-0.45	-3.06	-1.97
$E^{(20)}_{exch-ind,r}(S^2)$	1.67	11.13	8.03
$E^{(22)}_{exch-ind}(S^2)$	0.25	1.61	1.12
δHF	-0.86	-6.91	-4.94
$E^{(20)}_{disp}$	-2.55	-10.99	-11.74
$E^{(30)}_{disp}$	0.09	0.38	0.40
$E^{(21)}_{disp}$	0.04	0.94	1.25
$E^{(22)}_{disp}$(SDQ)	-0.04	-0.38	-0.09
$E^{(22)}_{disp}$(T)	-0.43	-2.11	-2.49
$E^{(20)}_{exch-disp}(S^2)$	0.47	2.25	2.06
E_{elst}	-8.11	-32.20	-26.58
E_{exch}	8.08	40.55	31.22
E_{ind}	-2.40	-18.30	-11.88
E_{disp}	-2.41	-9.91	-10.61
E_{SAPT0}	-5.63	-23.93	-20.17
E_{SAPT2}	-4.67	-18.78	-17.10
$E_{\mathrm{SAPT2+}}$	-5.10	-20.33	-18.43
$E_{\mathrm{SAPT2+(3)}}$	-4.85	-19.87	-17.85

It is also possible to assess the convergence rates of different SAPT terms with respect to basis set for the water dimer. The electrostatic energy in the water dimer is much more sensitive to the size of the basis than it is in the argon dimer. This results from the description of the multipole moments of the water molecules. These interactions are less sensitive to the presence of diffuse functions than are purely dispersion bound complexes. Even so, the convergence rate of cc-pVXZ basis sets seems to be one cardinal number slower than the aug-cc-pVXZ series. This is better than the dispersion complexes, where the convergence

rate was slower by two cardinal numbers. Unfortunately, the aug-cc-pVTZ basis (184 functions) is nearly as large as the cc-pVQZ basis (230 functions). As interesting aside, there have been reports of the convergence rates of CP and unCP corrected interactions energies for the water dimer that show the unCP corrected results are well converged at haug-cc-pVQZ, where diffuse functions are removed from hydrogen atoms.[51,224] Remembering that SAPT energies are free from BSSE, it is obvious from Table 4 that this observation is a result of fortuitous error cancellation.

Table 4: SAPT results for the water dimer computed with various basis sets.[a] Energies are reported in kcal mol^{-1}.

	DZ	TZ	QZ	5Z	aDZ	aTZ	aQZ	a5Z
$E^{(10)}_{elst,r}$	-9.10	-8.53	-8.41	-8.36	-8.39	-8.33	-8.34	-8.34
$E^{(12)}_{elst,r}$	-0.26	-0.06	0.01	0.04	0.03	0.06	0.05	0.05
$E^{(13)}_{elst,r}$	0.32	0.24	0.18	0.16	0.17	0.16	0.15	0.15
$E^{(10)}_{exch}$	7.45	7.16	7.06	7.02	7.04	7.03	7.02	7.01
$E^{(10)}_{exch}(S^2)$	7.39	7.11	7.01	6.97	6.99	6.98	6.97	6.97
$E^{(11)}_{exch}(S^2)$	-0.02	0.08	0.11	0.12	0.03	0.09	0.12	0.13
$E^{(12)}_{exch}(S^2)$	1.24	0.95	0.84	0.79	1.35	1.01	0.86	0.79
$E^{(20)}_{ind,r}$	-2.80	-2.99	-3.00	-3.00	-2.87	-3.01	-3.01	-3.01
$E^{(22)}_{ind}$	-0.46	-0.41	-0.39	-0.38	-0.53	-0.45	-0.41	-0.39
$E^{(20)}_{exch-ind,r}(S^2)$	1.61	1.70	1.67	1.66	1.56	1.67	1.66	1.66
$E^{(22)}_{exch-ind}(S^2)$	0.26	0.24	0.22	0.21	0.28	0.25	0.23	0.22
δHF	-0.84	-0.89	-0.90	-0.91	-0.90	-0.91	-0.92	-0.92
$E^{(20)}_{disp}$	-1.50	-2.23	-2.49	-2.61	-2.22	-2.55	-2.64	-2.67
$E^{(30)}_{disp}$	0.03	0.07	0.10	0.12	0.04	0.09	0.13	0.14
$E^{(21)}_{disp}$	0.04	0.07	0.05	0.04	0.07	0.04	0.03	0.03
$E^{(22)}_{disp}$(SDQ)	-0.05	-0.01	0.02	0.03	-0.11	-0.04	0.00	0.02
$E^{(22)}_{disp}$(T)	-0.17	-0.33	-0.40	-0.43	-0.34	-0.43	-0.45	-0.45
$E^{(20)}_{exch-disp}(S^2)$	0.35	0.43	0.47	0.49	0.41	0.47	0.49	0.50
E_{elst}	-9.04	-8.35	-8.22	-8.16	-8.20	-8.11	-8.14	-8.14
E_{exch}	8.67	8.20	8.01	7.93	8.42	8.13	7.99	7.93
E_{ind}	-2.22	-2.37	-2.40	-2.42	-2.46	-2.45	-2.44	-2.44
E_{disp}	-1.30	-2.00	-2.25	-2.37	-2.16	-2.41	-2.44	-2.43
E_{SAPT0}	-4.83	-5.35	-5.60	-5.71	-5.39	-5.63	-5.73	-5.75
E_{SAPT2}	-4.06	-4.55	-4.81	-4.93	-4.22	-4.67	-4.89	-4.96
$E_{\mathrm{SAPT2+}}$	-4.25	-4.82	-5.14	-5.30	-4.61	-5.10	-5.30	-5.37
$E_{\mathrm{SAPT2+(3)}}$	-3.90	-4.52	-4.86	-5.02	-4.40	-4.85	-5.03	-5.08

[a]XZ abbreviates cc-pVXZ and aXZ abbreviates aug-cc-pVXZ.

1.4.2.2 Formic Acid Dimer

The formic acid dimer differs from the water dimer in that it contains two hydrogen bonds and that the hydrogens are much more acidic. This results in a stronger interaction by roughly a factor of 4; only the induction term grows disproportionately between the water dimer and formic acid dimer. The intramonomer corrections become much more important for the formic acid dimer. In this case, the corrections to electrostatics and exchange are destabilizing by 6.5 kcal mol^{-1}; this is an unusually large intramonomer correction. The self consistency effects captured by the δHF term are also remarkably important at -7 kcal mol^{-1}. This illustrates the importance of higher-order induction effects; the inclusion of third-order induction would reduce the δHF term to -4.5 kcal mol^{-1}.

1.4.2.3 Watson-Crick Adenine-Thymine

The Watson-Crick structure of adenine-thymine also contains two hydrogen bonds, but the hydrogens are not as acidic as they are in formic acid. Additionally, adenine-thymine contains long-range dispersion interactions between the π systems. This dimer has decreased electrostatics, exchange, and induction relative to the formic acid dimer resulting from longer hydrogen bond distances (1.67 Å in the formic acid dimer compared to 1.82Å and 1.92 Å in adenine-thymine). The dispersion, however, increases as a result of the long-range interactions. It has been noted that certain methods that neglect long-range dispersion (namely, M05-2X and M06-2X) will systematically underestimate the interaction of hydrogen bonded base pairs.[91]

1.4.3 Mixed Influence Complexes

Another class of interactions exist that contain roughly equivalent contributions from electrostatics and dispersion. XH-π interactions, such water-benzene, ammonia-benzene, or HCN-benzene, are all prototypical mixed influence complexes. Detailed SAPT analysis of certain π stacking interactions has shown them to be mixed influence complexes rather than dispersion dominated as originally thought. The T-shaped benzene dimer, on the other hand, was thought to be of mixed influence, but is actually dominated by dispersion.

Table 5: SAPT2+(3)/aug-cc-pVTZ results for mixed influence complexes in kcal mol^{-1}.

	H_2O-C_6H_6	$(C_6H_6)_2$ T-shaped	U-U Stacked
$E^{(10)}_{elst,r}$	-2.97	-2.09	-8.99
$E^{(12)}_{elst,r}$	0.21	-0.04	0.24
$E^{(13)}_{elst,r}$	0.05	0.13	0.23
$E^{(10)}_{exch}(S^2)$	3.00	4.26	10.72
$E^{(11)}_{exch}(S^2)$	0.12	0.13	0.21
$E^{(12)}_{exch}(S^2)$	0.18	0.15	0.73
$E^{(20)}_{ind,r}$	-1.31	-1.17	-4.71
$E^{(22)}_{ind}$	-0.06	-0.11	-0.20
$E^{(20)}_{exch-ind,r}(S^2)$	0.68	0.91	3.56
$E^{(22)}_{exch-ind}(S^2)$	0.03	0.08	0.15
δHF	-0.33	-0.39	-0.54
$E^{(20)}_{disp}$	-3.28	-6.03	-13.83
$E^{(30)}_{disp}$	0.11	0.19	0.33
$E^{(21)}_{disp}$	0.53	1.37	3.59
$E^{(22)}_{disp}$(SDQ)	0.08	0.34	-0.47
$E^{(22)}_{disp}$(T)	-0.67	-1.27	-3.14
$E^{(20)}_{exch-disp}(S^2)$	0.34	0.63	1.38
E_{elst}	-2.72	-2.00	-8.51
E_{exch}	3.30	4.54	11.66
E_{ind}	-0.99	-0.68	-1.73
E_{disp}	-2.90	-4.77	-12.14
E_{SAPT0}	-3.88	-3.88	-12.40
E_{SAPT2}	-3.40	-3.66	-11.26
$E_{\mathrm{SAPT2+}}$	-3.47	-3.22	-11.29
$E_{\mathrm{SAPT2+(3)}}$	-3.31	-2.91	-10.73

1.4.3.1 Water-Benzene

The water-benzene complex contains what is often described as a π hydrogen bond, where the π system of benzene is the hydrogen bond acceptor. This leads to favorable electrostatic interactions between the positively charged hydrogen and the benzene's negatively charged π electrons. Due to the proximity of the water molecule to benzene, there is also a significant dispersion force. The hydrogen is able to polarize the π orbitals of benzene leading to significant induction. The electrostatic interactions in this complex are well described as interacting multipoles; the contribution from charge penetration is approximately -0.2 kcal

mol^{-1}.

1.4.3.2 T-shaped Benzene Dimer

The T-shaped benzene dimer had been classified as a mixed influence complex based on chemical intuition; there are CH-π and favorable quadrapole-quadrapole interactions present in this complex. Subsequent analysis with SAPT shows that this complex is actually dominated by dispersion. It is interesting to note that the electrostatic energy is larger in the stacked benzene dimer than in the T-shaped benzene dimer. This is a result of the decreased charge penetration in the T-shaped configuration. In the T-shaped configuration, charge penetration accounts for roughly -1.35 kcal mol^{-1} of the total electrostatic energy (-2.1 kcal mol^{-1}). In contrast, the stacked configuration is repulsive electrostatically on the basis of multipole interactions. In this case, the charge penetration (-4.2 kcal mol^{-1}) is entirely responsible for the attractive electrostatic interaction (-2.9 kcal mol^{-1}). The attractive components of the benzene dimer interaction (electrostatics, induction and dispersion) are roughly 71% dispersion in the stacked configuration; in the T-shaped configuration, the relative amount of dispersion is reduced (64%), but not reduced significantly.

1.4.3.3 π-Stacked Uracil Dimer

Another complex that was inappropriately classified is the stacked uracil dimer. In this case, it was classified as dispersion dominated as a result of its stacked geometry. This complex has strong electrostatic interactions due to the four NH$\cdots$O contacts that are present. Since these electrostatic interactions are not in their ideal geometry, they are not as strong as they would be in a hydrogen bonded uracil dimer. In addition, there is roughly -3.4 kcal mol^{-1} of charge penetration present in this complex. As a result, the attractive components of the stacked uracil dimer interaction are composed of roughly 55% dispersion. If charge penetration or the "classical" electrostatic components were neglected, this would shift the apparent nature of the interaction sufficiently to classify this interaction as dispersion dominated (according to the scheme in Ref. 65).

CHAPTER II

DENSITY FITTING IN SYMMETRY-ADAPTED PERTURBATION THEORY

2.1 Notation and Definitions

The notation used here is chosen to remain consistent with the literature describing many-body SAPT. Certain terms will not be defined below, but rather in the text before they are used. We will use the Einstein summation convention where sums over repeated indices are implied. Any terms that are listed, but not defined, will be addressed subsequently.

Indices

μ, ν, ρ, σ Index atomic orbitals

i, j, k, l Index all molecular orbitals

a Indexes occupied molecular orbitals of monomer A

b Indexes occupied molecular orbitals of monomer B

r Indexes virtual molecular orbitals of monomer A

s Indexes virtual molecular orbitals of monomer B

P, Q Index auxiliary (DF) basis functions

L, M Index Laplace quadrature points

Miscellaneous

N_A Number of electrons of monomer A

N_B Number of electrons of monomer B

V_0 Intermolecular nuclear repulsion energy

ϵ_i Orbital energy

$C_{\mu i}$ HF coefficient matrix

$A_{\mu i}$ HF coefficient matrix (of monomer A)

$B_{\mu i}$ HF coefficient matrix (of monomer B)

$\Delta_{ij}^{kl} = \frac{1}{\epsilon_i+\epsilon_j-\epsilon_k-\epsilon_l}$ Energy denominator

d_{ij}^L Laplace factorization of an energy denominator ($\Delta_{ij}^{kl} \approx d_{ik}^L d_{jl}^L$)

Integrals

S_i^j Overlap integrals: $(i|j)$

$(\nu_A)_i^j$ Nuclear attraction integrals (to the nuclei of monomer A): $(i|\nu_A|j)$

$(\nu_B)_i^j$ Nuclear attraction integrals (to the nuclei of monomer B): $(i|\nu_B|j)$

ν_{ik}^{jl} Two electron integrals: $(ij|kl)$ or $\langle ik|jl\rangle$

$(\omega_A)_i^j = (\nu_A)_i^j + 2\nu_{ia}^{ja}$ Electrostatic potential integrals (of monomer A)

$(\omega_B)_i^j = (\nu_B)_i^j + 2\nu_{ib}^{jb}$ Electrostatic potential integrals (of monomer B)

$\tilde{\nu}_{ik}^{jl}$ Generalized two electron integrals

$[J]_{PQ} = \int P(\mathbf{r}_1)\frac{1}{r_{12}}Q(\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2$ The Coulomb metric

$C_{ij}^P = (ij|Q)[J^{-1/2}]_{PQ}$ Symmetric factorization of DF integrals (bare)

$g_{kl}^{ij} = 2\nu_{kl}^{ij} - \nu_{lk}^{ij}$ Antisymmetrized two-electron integrals

A_{ij}^P Generalized symmetric factorization of DF integrals (monomer A)

B_{ij}^P Generalized symmetric factorization of DF integrals (monomer B)

Amplitudes

s_a^r CPHF coefficients of monomer A

s_b^s CPHF coefficients of monomer B

$\overset{(0)}{s}{}_a^r = (\omega_B)_a^r \Delta_a^r$ Zeroth-order induction amplitudes of monomer A

$\overset{(0)}{s}{}_b^s = (\omega_A)_b^s \Delta_b^s$ Zeroth-order induction amplitudes of monomer B

$t_{ab}^{rs} = \nu_{ab}^{rs}\Delta_{ab}^{rs}$ Dispersion amplitudes

$t_{a_1a_2}^{r_1r_2} = \nu_{a_1a_2}^{r_1r_2}\Delta_{a_1a_2}^{r_1r_2}$ Monomer A doubles amplitudes

$t_{b_1b_2}^{s_1s_2} = \nu_{b_1b_2}^{s_1s_2}\Delta_{b_1b_2}^{s_1s_2}$ Monomer B doubles amplitudes

$\theta_{a_1a_2}^{r_1r_2} = 2t_{a_1a_2}^{r_1r_2} - t_{a_1a_2}^{r_2r_1}$ Antisymmetrized monomer A doubles amplitudes

$\theta_{b_1b_2}^{s_1s_2} = 2t_{b_1b_2}^{s_1s_2} - t_{b_1b_2}^{s_2s_1}$ Antisymmetrized monomer B doubles amplitudes

$\overset{(2)}{t}{}_{a_1a_2}^{r_1r_2}$ Second-order monomer A doubles amplitudes

$\overset{(2)}{t}{}_{b_1b_2}^{s_1s_2}$ Second-order monomer B doubles amplitudes

$\overset{(2)}{\theta}{}_{a_1a_2}^{r_1r_2} = 2\overset{(2)}{t}{}_{a_1a_2}^{r_1r_2} - \overset{(2)}{t}{}_{a_1a_2}^{r_2r_1}$ Antisymmetrized second-order monomer A doubles amplitudes

$\overset{(2)}{\theta}{}_{b_1b_2}^{s_1s_2} = 2\overset{(2)}{t}{}_{b_1b_2}^{s_1s_2} - \overset{(2)}{t}{}_{b_1b_2}^{s_2s_1}$ Antisymmetrized second-order monomer B doubles amplitudes

$t_{r_1}^{a_1}$ Second-order monomer A singles amplitudes

$t_{s_1}^{b_1}$ Second-order monomer B singles amplitudes

Common Intermediates

$\tilde{a}^P = A_{aa}^P$ Diagonal occupied-occupied generalized DF integrals of monomer A

$\tilde{b}^P = B_{bb}^P$ Diagonal occupied-occupied generalized DF integrals of monomer B

$T_{ar}^P = t_{ab}^{rs} B_{bs}^P$ Generalized dispersion intermediates

$T_{bs}^P = t_{ab}^{rs} A_{ar}^P$ Generalized dispersion intermediates

$\Theta_{a_1 r_1}^P = \theta_{a_1 a_2}^{r_1 r_2} A_{a_2 r_2}^P$

$\Theta_{b_1 s_1}^P = \theta_{b_1 b_2}^{s_1 s_2} B_{b_2 s_2}^P$

$\overset{(2)}{\Theta}{}_{a_1 r_1}^P = \overset{(2)}{\theta}{}_{a_1 a_2}^{r_1 r_2} A_{a_2 r_2}^P$

$\overset{(2)}{\Theta}{}_{b_1 s_1}^P = \overset{(2)}{\theta}{}_{b_1 b_2}^{s_1 s_2} B_{b_2 s_2}^P$

$\tilde{V}_{r_1}^{a_1}$ Monomer A exchange-induction integrals

$\tilde{V}_{s_1}^{b_1}$ Monomer B exchange-induction integrals

$\tilde{V}_{a_1 b_1}^{r_1 s_1}$ Exchange-dispersion integrals

${}^A G_{r_1 s_1}^{a_1 b_1} = g_{a_1 r_2}^{r_1 a_2} t_{a_2 b_1}^{r_2 s_1}$

${}^B G_{r_1 s_1}^{a_1 b_1} = g_{b_1 s_2}^{s_1 b_2} t_{a_1 b_2}^{r_1 s_2}$

2.1.1 Generalized Two Electron Integrals

The exchange terms in SAPT present unique challenges for the introduction of DF integrals. The exchange interactions can be formulated in terms of second-quantization or interaction density matrices.[154,155] The second-quantization approach leads to equations which depend only on Coulomb type two-electron integrals $[(AA|BB)$, where $A(B)$ refers to any index on monomer A (B)] and are amenable to the introduction of DF integrals.[94] The interaction density matrix approach leads to more complex equations that include Coulomb, exchange $[(AB|AB)]$, and hybrid $[(AA|AB)]$ integrals. However, this approach involves integrals with fewer virtual indices than the second-quantization approach. Subsequently, the exchange corrections derived with the interaction density matrix approach have a lower computational scaling. However, the interaction density matrix approach uses generalized two-electron integrals rather than the bare two-electron integrals found in the

second-quantization approach. The generalized two-electron integrals are given as:[110]

$$\tilde{\nu}_{ik}^{jl} = \nu_{ik}^{jl} + (i|j)\,\frac{(k|\nu_A|l)}{N_A} + (k|l)\,\frac{(i|\nu_B|j)}{N_B} + (i|j)\,(k|l)\,\frac{V_0}{N_A N_B}. \tag{36}$$

Under the DF approximation, the generalized two-electron integrals are written as:

$$\tilde{\nu}_{ik}^{jl} \approx C_{ij}^{P} C_{kl}^{P} + (i|j)\,\frac{(k|\nu_A|l)}{N_A} + (k|l)\,\frac{(i|\nu_B|j)}{N_B} + (i|j)\,(k|l)\,\frac{V_0}{N_A N_B}. \tag{37}$$

Our initial implementation of DF-SAPT evaluated generalized two-electron integrals as shown above; the approximate four-index integrals were formed and then dressed with one-electron contributions. It is possible to define DF intermediates that include the one-electron contributions by adding three additional entries to the auxiliary index:

$$\begin{aligned}
A_{ij}^{1} &= C_{ij}^{1} \\
&\vdots \\
A_{ij}^{N_{df}} &= C_{ij}^{N_{df}} \\
A_{ij}^{N_{df}+1} &= (i|j) \\
A_{ij}^{N_{df}+2} &= \frac{(i|\nu_B|j)}{N_B} \\
A_{ij}^{N_{df}+3} &= (i|j)\sqrt{\frac{V_0}{N_A N_B}},
\end{aligned} \tag{38}$$

and

$$\begin{aligned}
B_{ij}^{1} &= C_{ij}^{1} \\
&\vdots \\
B_{ij}^{N_{df}} &= C_{ij}^{N_{df}} \\
B_{ij}^{N_{df}+1} &= \frac{(i|\nu_A|j)}{N_A} \\
B_{ij}^{N_{df}+2} &= (i|j) \\
B_{ij}^{N_{df}+3} &= (i|j)\sqrt{\frac{V_0}{N_A N_B}}.
\end{aligned} \tag{39}$$

Now, the length of the auxiliary index is equal to the number of DF basis functions plus three ($N_{aux} = N_{df} + 3$); from this point forward, we will neglect the difference between N_{aux}

and N_{df} for generalized DF integrals when considering scalings. This allows the generalized two-electron integrals to be formed directly as:

$$\tilde{\nu}_{ik}^{jl} \approx A_{ij}^{P} B_{kl}^{P}. \tag{40}$$

In practice, the A and B quantities are not explicitly stored; rather, the C type DF integrals are stored and the three additional auxiliary indices are appended as necessary when these integrals are read into memory. The memory requirements for the DF integrals needed to evaluate Eqn. 40 for two-electron integrals with bra-ket symmetry [i.e. $(ij|ij)$ type integrals] would be, in principle, doubled. However, in practice, this doubling is only encountered for $(ab|ab)$ integrals, where the $o^2 N_{df}$ DF integrals can easily be stored in memory. The $(as|as)$ and $(rb|rb)$ integrals appear only once, and the symmetry of the DF integrals cannot be exploited in the most efficient factorization of those exchange terms. Finally, the $(rs|rs)$ two-electron integrals do not appear in the SAPT exchange corrections. Therefore, the potential loss of symmetry in the formation of generalized two-electron integrals is not a problem.

2.1.2 Second-order Singles Amplitudes

Several intramonomer correlation corrections require the second-order singles amplitudes.[110]

$$t_{r_1}^{a_1} = \left[\theta_{r_3 r_2}^{a_3 a_1} \nu_{a_3 r_1}^{r_3 r_2} - \theta_{r_3 r_1}^{a_3 a_2} \nu_{a_3 a_2}^{r_3 a_1}\right] \Delta_{a_1}^{r_1} \tag{41}$$

These can be efficiently evaluated through the introduction of DF intermediates.

$$t_{r_1}^{a_1} = \left[\Theta_{a_1 r_2}^{P} C_{r_1 r_2}^{P} - \Theta_{a_2 r_1}^{P} C_{a_1 a_2}^{P}\right] \Delta_{a_1}^{r_1} \tag{42}$$

These two terms are a subset of those required during the evaluation of the $E_{elst,r}^{(12)}$ correction. It is possible to store the second-order singles amplitudes during the evaluation of this term to avoid redundant work.

2.1.3 Second-order Doubles Amplitudes

The most demanding step in the evaluation of intramonomer electron correlation effects in SAPT2 is the formation of the second-order double excitation amplitudes,[110]

$$\overset{(2)}{t}{}^{r_1 r_2}_{a_1 a_2} = \big[\nu^{r_3 r_4}_{r_1 r_2} t^{a_1 a_2}_{r_3 r_4} + \nu^{a_1 a_2}_{a_3 a_4} t^{a_3 a_4}_{r_1 r_2} + \nu^{a_2 r_3}_{r_2 a_3} \theta^{a_1 a_3}_{r_1 r_3} + \nu^{a_1 r_3}_{r_1 a_3} \theta^{a_2 a_3}_{r_2 r_3} - \nu^{a_2 r_3}_{a_3 r_2} t^{a_1 a_3}_{r_1 r_3} - \nu^{a_1 r_3}_{a_3 r_1} t^{a_2 a_3}_{r_2 r_3} - \nu^{a_1 r_3}_{a_3 r_2} t^{a_3 a_2}_{r_1 r_3} - \nu^{a_2 r_3}_{a_3 r_1} t^{a_1 a_3}_{r_3 r_2}\big] \Delta^{r_1 r_2}_{a_1 a_2}. \tag{43}$$

Unfortunately, the formation of the second-order amplitudes is not particularly amenable to the introduction of DF intermediates. The third and fourth terms can be rewritten with the DF integrals and the Θ intermediates as

$$\Theta^P_{a_1 r_1} C^P_{a_2 r_2} + C^P_{a_1 r_1} \Theta^P_{a_2 r_2}, \tag{44}$$

which reduces the scaling of these terms from $\mathcal{O}(o^3 v^3)$ to $\mathcal{O}(o^2 v^2 N_{df})$. In principle, the fifth and sixth terms could also be evaluated in $\mathcal{O}(N^5)$ if density fitting integrals and Laplace transformed energy denominators were introduced.

$$-\nu^{a_2 r_3}_{a_3 r_2} C^P_{a_1 r_1} C^P_{a_3 r_3} d^L_{a_1 r_1} d^L_{a_3 r_3} - \nu^{a_1 r_3}_{a_3 r_1} C^P_{a_2 r_2} C^P_{a_3 r_3} d^L_{a_2 r_2} d^L_{a_3 r_3} \tag{45}$$

However, in practice, this is simply evaluated conventionally with $\mathcal{O}(o^3 v^3)$ work. Including the contribution from the v^4 integrals in the first term scales as $\mathcal{O}(o^2 v^4)$. A conventional SAPT program must have access to this large group of integrals in order to evaluate this term. There are two types of v^4 integrals needed for SAPT (r^4 and s^4); additionally, to evaluate the $E^{(30)}_{disp}$ term the $r^2 s^2$ integrals are required. While DF cannot improve the scaling of steps with v^4 dependence, DF can still improve the efficiency of the SAPT algorithm. For example, each of the r^4 integrals are needed only once during the SAPT computation (this is true of all v^4 integrals); when they are approximated with DF, it is practical to form these integrals "on the fly" (from the B^P three index DF integrals) and evaluate their contribution to $\overset{(2)}{t}{}^{a_1 a_2}_{r_1 r_2}$ without storing them. The temporary batches of integrals should be as large as the system's memory will allow.

2.1.4 Exchange-Induction Integrals

Certain collections of exchange integrals are convenient to form, because they can be reused to evaluate several different corrections. The integrals needed to evaluate the $E^{(20)}_{exch-ind,r}$ term also appear in the $E^{(12)}_{exch}K^f_2$ and $E^{(30)}_{exch-ind}$ corrections.[162]

$$\begin{aligned}
\tilde{V}^{r_1}_{a_1} =& \tilde{\nu}^{b_1 r_1}_{a_1 b_1} \\
&+ 2S^{b_1}_{a_1}\tilde{\nu}^{a_2 r_1}_{a_2 b_1} - S^{b_1}_{a_2}\tilde{\nu}^{a_2 r_1}_{a_1 b_1} \\
&+ 2S^{b_1}_{a_2}\tilde{\nu}^{r_1 a_2}_{a_1 b_1} - S^{b_1}_{a_1}\tilde{\nu}^{r_1 a_2}_{a_2 b_1} \\
&+ 2S^{r_1}_{b_2}\tilde{\nu}^{b_2 b_1}_{a_1 b_1} - S^{r_1}_{b_2}\tilde{\nu}^{b_1 b_2}_{a_1 b_1} \\
&- 2S^{b_1}_{a_1}S^{r_1}_{b_2}\tilde{\nu}^{a_2 b_2}_{a_2 b_1} - 2S^{b_2}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{a_2 b_1}_{a_1 b_1} + S^{b_1}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{a_2 b_2}_{a_1 b_1} \\
&- 2S^{b_1}_{a_2}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_1 b_1} - 2S^{b_2}_{a_1}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_1}_{a_2 b_1} + S^{b_1}_{a_1}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_2 b_1}
\end{aligned} \tag{46}$$

This quantity can be efficiently evaluated with the introduction of the generalized DF integrals. In order to apply these integrals to the $E^{(12)}_{exch}K^f_2$ correction, they must be stored separately with the last six terms (those that are quadratic in S) scaled by a factor of 2.

2.1.5 Exchange-Dispersion Integrals

A similar quantity can be formed to evaluate the $E^{(20)}_{exch-disp}$ term and reused to evaluate a portion of the $E^{(30)}_{exch-disp}$ term.[162]

$$\begin{aligned}
\tilde{V}^{a_1 b_1}_{r_1 s_1} =& \tilde{\nu}^{s_1 r_1}_{a_1 b_1} \\
&+ S^{s_1}_{a_1}(2\tilde{\nu}^{a_2 r_1}_{a_2 b_1} - \tilde{\nu}^{r_1 a_2}_{a_2 b_1}) + S^{s_1}_{a_2}(2\tilde{\nu}^{r_1 a_2}_{a_1 b_1} - \tilde{\nu}^{a_2 r_1}_{a_1 b_1}) \\
&+ S^{r_1}_{b_1}(2\tilde{\nu}^{s_1 b_2}_{a_1 b_2} - \tilde{\nu}^{b_2 s_1}_{a_1 b_2}) + S^{r_1}_{b_2}(2\tilde{\nu}^{b_2 s_1}_{a_1 b_1} - \tilde{\nu}^{s_1 b_2}_{a_1 b_1}) \\
&+ S^{r_1}_{b_1}S^{b_2}_{a_2}\tilde{\nu}^{a_2 s_1}_{a_1 b_2} - 2S^{r_1}_{b_2}S^{b_2}_{a_2}\tilde{\nu}^{a_2 s_1}_{a_1 b_1} - 2S^{r_1}_{b_1}S^{b_2}_{a_1}\tilde{\nu}^{a_2 s_1}_{a_2 b_2} + 4S^{r_1}_{b_2}S^{b_2}_{a_1}\tilde{\nu}^{a_2 s_1}_{a_2 b_1} \\
&+ S^{s_1}_{a_1}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_2 b_1} - 2S^{s_1}_{a_2}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_1 b_1} - 2S^{s_1}_{a_1}S^{a_2}_{b_1}\tilde{\nu}^{r_1 b_2}_{a_2 b_2} + 4S^{s_1}_{a_2}S^{a_2}_{b_1}\tilde{\nu}^{r_1 b_2}_{a_1 b_2} \\
&+ S^{r_1}_{b_2}S^{s_1}_{a_2}\tilde{\nu}^{a_2 b_2}_{a_1 b_1} - 2S^{r_1}_{b_1}S^{s_1}_{a_2}\tilde{\nu}^{a_2 b_2}_{a_1 b_2} - 2S^{r_1}_{b_2}S^{s_1}_{a_1}\tilde{\nu}^{a_2 b_2}_{a_2 b_1} \\
&+ S^{a_2}_{b_1}S^{b_2}_{a_1}\tilde{\nu}^{r_1 s_1}_{a_2 b_2} - 2S^{a_2}_{b_1}S^{b_2}_{a_2}\tilde{\nu}^{r_1 s_1}_{a_1 b_2} - 2S^{a_2}_{b_2}S^{b_2}_{a_1}\tilde{\nu}^{r_1 s_1}_{a_2 b_1}
\end{aligned} \tag{47}$$

A more efficient formulation of the $E^{(20)}_{exch-disp}$ correction is described below for use in SAPT0 computations. When higher-order term are included, however, the efficient formulation is not necessary due to its $\mathcal{O}(o^2v^2N_{df})$ scaling.

2.1.6 $g_{ra}^{ar} \times t_{rs}^{ab}$

This contraction appears several times in the intramonomer correlation corrections to dispersion. The evaluation of this contraction canonically scales $\mathcal{O}(o^3v^3)$. The introduction of DF integrals can only factor the coulomb portion of the g integrals. If Laplace transformed energy denominators are applied, the DF integrals can be applied in the dispersion amplitudes and the entire term can be factored in $\mathcal{O}(N^5)$.[110]

$$ {}^{A}G_{r_1 s_1}^{a_1 b_1} = g_{a_1 r_2}^{r_1 a_2} t_{a_2 b_1}^{r_2 s_1} \tag{48} $$

$$ X_{a_1 r_1}^{P,L} = (2\nu_{r_1 r_2}^{a_1 a_2} - \nu_{a_2 r_2}^{a_1 r_1}) C_{a_2 r_2}^{P} d_{a_2 r_2}^{L} \tag{49} $$

$$ {}^{A}G_{r_1 s_1}^{a_1 b_1} = X_{a_1 r_1}^{P,L} C_{b_1 s_1}^{P} d_{b_1 s_1}^{L} \tag{50} $$

This factorization scales $\mathcal{O}(o^2v^2N_{df})N_\delta$, which is a significant improvement over the canonical evaluation. This intermediate can be used in the $E_{disp}^{(21)}$, $E_{disp}^{(211)}$, and $E_{disp}^{(22)}$(D) corrections.

2.2 DF-SAPT0

2.2.1 DF Integral Formation

The formation of the density-fitted integrals is the first step in the SAPT0 procedure and one that must be implemented efficiently. First the AO DF integrals, $(\mu\nu|P)$, must be generated. Then, they must be transformed by the inverse square root of the Coulomb metric:

$$ C_{\mu\nu}^{P} \leftarrow (\mu\nu|Q)[J^{-1/2}]_{PQ}. \tag{51} $$

The square brackets around the Coulomb metric are to clarify that we need the PQth element of the inverse square root of J, not the inverse square root of the PQth element of J. The resulting symmetric factorization of the AO DF integrals must now be transformed to the MO basis. For SAPT0 computations, there are 9 unique types of MO DF integrals required. In general,

$$ C_{ij}^{P} \leftarrow C_{\mu i} C_{\nu j} C_{\mu\nu}^{P}, \tag{52} $$

where, $C_{a_1 a_2}^{P}, C_{a_1 r_1}^{P}, C_{r_1 r_2}^{P}, C_{b_1 b_2}^{P}, C_{b_1 s_1}^{P}, C_{s_1 s_2}^{P}, C_{a_1 b_1}^{P}, C_{a_1 s_1}^{P}$ and $C_{r_1 b_1}^{P}$ must be formed. The most expensive step of this process is the multiplication of the AO DF integrals by $J^{-1/2}$, which scales $\mathcal{O}(N_{ao}^2 N_{df}^2)$. The AO to MO transformation scales $\mathcal{O}(N_{ao}^2 N_{mo} N_{df})$; this is less

expensive by roughly a factor of six in typical cases, because $N_{df} \approx 3N_{ao}$ and there is a two-fold permutational symmetry in the AO DF integrals that can be exploited in the multiplication by $J^{-1/2}$, but not the AO to MO transformation.

Algorithm 1 Pseudocode to compute the symmetric factorization of the MO DF integrals.

```
for μ_shell < #AOShells do
  for ν_shell <= μ_shell do
    for P_shell < #DFShells do
      if Above the Schwarz tolerance then
        Compute (μ_shell ν_shell | P_shell)
      end if
    end for
    if Out of memory OR last μ_shell ν_shell then
      C^P_μν ← (μν|Q)[J^{-1/2}]_PQ
      Write C^P_μν block, transposed, to disk
    end if
  end for
end for
for P_block < #Pblocks do
  Read C^P_μν block
  for P < #P_block do
    Unpack C^P_μν permutational symmetry
    C^P_ij ← A_μi A_νj C^P_μν
    Store C^P_{a1a2}, C^P_{a1r1}, C^P_{r1r2}
    C^P_ij ← A_μi B_νj C^P_μν
    Store C^P_{a1b1}, C^P_{a1s1}, C^P_{r1b1}
    C^P_ij ← B_μi B_νj C^P_μν
    Store C^P_{b1b2}, C^P_{b1s1}, C^P_{s1s2}
  end for
  Write C^P_{a1a2}, C^P_{a1r1}, C^P_{r1r2} block to disk
  Write C^P_{a1b1}, C^P_{a1s1}, C^P_{r1b1} block to disk
  Write C^P_{b1b2}, C^P_{b1s1}, C^P_{s1s2} block to disk
end for
```

This algorithm computes the symmetric factorization of the MO DF integrals with one pass through the AO DF integrals. The memory requirement for this algorithm is minimal ($2N_{df}^2$ will always be sufficient) and arrays on disk are only written and read once. The bottleneck associated with this procedure is the disk based transposition of $C_{\mu\nu}^P$. There is roughly $3N_{ao}^2 N_{df}$ disk storage required; this does not begin to become prohibitive on many computers until $N_{ao} > 2000$.

2.2.2 $E^{(10)}_{elst,r}$

The leading electrostatic term is given by:[110]

$$E^{(10)}_{elst,r} = 4\tilde{\nu}^{a_1 b_1}_{a_1 b_1}. \tag{53}$$

This can be evaluated simply with generalized DF integrals:

$$E^{(10)}_{elst,r} = 4A^P_{a_1 a_1} B^P_{b_1 b_1} = 4\tilde{a}^P \tilde{b}^P. \tag{54}$$

2.2.3 $E^{(10)}_{exch}(S^2)$

The leading exchange term, under the S^2 approximation, is given by:[110]

$$\begin{aligned} E^{(10)}_{exch}(S^2) = &- 2\tilde{\nu}^{b_1 a_1}_{a_1 b_1} - 2S^{b_1}_{a_2}(2\tilde{\nu}^{a_1 a_2}_{a_1 b_1} - \tilde{\nu}^{a_2 a_1}_{a_1 b_1}) - 2S^{a_1}_{b_2}(2\tilde{\nu}^{b_2 b_1}_{a_1 b_1} - \tilde{\nu}^{b_1 b_2}_{a_1 b_1}) \\ &+ 4S^{b_1}_{a_2} S^{a_2}_{b_2} \tilde{\nu}^{a_1 b_2}_{a_1 b_1} + 4S^{b_2}_{a_2} S^{a_1}_{b_2} \tilde{\nu}^{a_2 b_1}_{a_1 b_1} - 2S^{b_1}_{a_2} S^{a_1}_{b_2} \tilde{\nu}^{a_2 b_2}_{a_1 b_1}. \end{aligned} \tag{55}$$

The $E^{(10)}_{exch}(S^2)$ term can be rewritten in terms of generalized DF integrals and a few simple intermediates,

$$X_{b_1 b_2} = S^{b_1}_{a_2} S^{a_2}_{b_2} \tag{56}$$

$$X_{a_1 a_2} = S^{b_2}_{a_2} S^{a_1}_{b_2}, \tag{57}$$

in a form that can be efficiently implemented without forming four-index arrays.

$$\begin{aligned} E^{(10)}_{exch}(S^2) = &- 2A^P_{a_1 b_1} B^P_{a_1 b_1} \\ &- 2S^{b_1}_{a_2}(2\tilde{a}^P B^P_{a_2 b_1} - A^P_{a_1 a_2} B^P_{a_1 b_1}) \\ &- 2S^{a_1}_{b_2}(2A^P_{a_1 b_2} \tilde{b}^P - A^P_{a_1 b_1} B^P_{b_1 b_2}) \\ &+ 4X_{b_1 b_2} \tilde{a}^P B^P_{b_1 b_2} + 4X_{a_1 a_2} A^P_{a_1 a_2} \tilde{b}^P - 2S^{b_1}_{a_2} S^{a_1}_{b_2} A^P_{a_1 a_2} B^P_{b_1 b_2} \end{aligned} \tag{58}$$

This formulation of the $E^{(10)}_{exch}(S^2)$ term allows for a memory efficient implementation by blocking over the P index.

2.2.4 $E^{(10)}_{exch}$

The leading exchange term, can also be evaluated through infinite-order in S. This is achieved by completely orthogonalizing the molecular orbitals of monomer A and B.[109]

$$P = \begin{bmatrix} 1 & S^{b_1}_{a_1} \\ S^{a_1}_{b_1} & 1 \end{bmatrix}^{-1} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The exact $E^{(10)}_{exch}$ evaluation is more involved than it is under the S^2 approximation. However, the both exact and S^2 $E^{(10)}_{exch}$ are a negligible fraction of the total SAPT0 computation time.

$$\begin{aligned} E^{(10)}_{exch} =& -2\tilde{\nu}^{b_1 a_1}_{a_1 b_1} - 2P^{a_2}_{a_1}(2\tilde{\nu}^{a_2 b_1}_{a_1 b_1} - \tilde{\nu}^{b_1 a_2}_{a_1 b_1}) - 2P^{b_2}_{b_1}(2\tilde{\nu}^{a_1 b_2}_{a_1 b_1} - \tilde{\nu}^{b_1 a_1}_{a_1 b_2}) \\ &+ 2P^{b_1}_{a_1}(2\tilde{\nu}^{a_2 b_1}_{a_2 a_1} - \tilde{\nu}^{a_2 b_1}_{a_1 a_2}) + 2P^{b_1}_{a_1}(2\tilde{\nu}^{b_1 b_2}_{a_1 b_2} - \tilde{\nu}^{b_2 b_1}_{a_1 b_2}) \\ &+ 2P^{b_1}_{a_1} P^{b_3}_{b_2}(2\tilde{\nu}^{b_1 b_3}_{a_1 b_2} - \tilde{\nu}^{b_3 b_1}_{a_1 b_2}) + 2P^{a_3}_{a_2} P^{b_1}_{a_1}(2\tilde{\nu}^{a_3 b_1}_{a_2 a_1} - \tilde{\nu}^{a_3 b_1}_{a_1 a_2}) \\ &+ 2P^{a_2}_{a_1} P^{b_2}_{b_1}(2\tilde{\nu}^{a_2 b_2}_{a_1 b_1} - \tilde{\nu}^{b_2 a_2}_{a_1 b_1}) + 2P^{b_1}_{a_1} P^{b_2}_{a_2}(2\tilde{\nu}^{b_1 a_2}_{a_1 b_2} - \tilde{\nu}^{a_2 b_1}_{a_1 b_2}) \end{aligned} \tag{59}$$

In structure, the $E^{(10)}_{exch}$ terms are identical to the $E^{(10)}_{exch}(S^2)$ terms, therefore, the factorization of these terms follow directly from the $E^{(10)}_{exch}(S^2)$ correction.

2.2.5 $E^{(20)}_{ind,r}$

The leading induction term contains two contributions: the changes to the electron density of monomer A induced by monomer B's electrostatic potential and the changes to the electron density of monomer B induced by monomer A's electrostatic potential. Through second-order in V, these changes are not coupled to each other.[110]

$$\begin{aligned} E^{(20)}_{ind,r}(A \leftarrow B) &= 2s^{r_1}_{a_1}(\omega_B)^{r_1}_{a_1} \\ E^{(20)}_{ind,r}(B \leftarrow A) &= 2s^{s_1}_{b_1}(\omega_A)^{b_1}_{s_1} \end{aligned} \tag{60}$$

This term can be evaluated with or without orbital response; the difference is the origin of the s coefficients. Here, we will discuss only the more rigorous formulation that includes the effect of orbital response. In this case, $s^{r_1}_{a_1}$ and $s^{s_1}_{b_1}$ are solutions to the CPHF equations for their corresponding monomer. Obviously, the work associated with the evaluation of this correction is entirely in the solution of the CPHF equations,

$$s^{r_1}_{a_1} = \left([4\nu^{a_1 a_2}_{r_1 r_2} - \nu^{a_2 a_1}_{r_1 r_2} - \nu^{a_1 r_1}_{a_2 r_2}]s^{r_1}_{a_1} + (\omega_B)^{r_1}_{a_1}\right) / \left(\epsilon_{a_1} - \epsilon_{r_1}\right). \tag{61}$$

It should be noted that these equations can be solved exactly, rather than iteratively, however, this would scale $\mathcal{O}(o^3v^3)$. The iterative solution of these equations scales $\mathcal{O}(o^2v^2)$ (per iteration), although the formation of the required matrix will scale $\mathcal{O}(N^5)$ (the exact scaling is dependant on the algorithm; under the DF approximation, the scaling is $\mathcal{O}(o^2v^2N_{df})$). To avoid the $\mathcal{O}(N^5)$ step and its associated storage requirements, the DF representation of the two-electron integrals is introduced and the multiplication by s is distributed.

$$s_{a_1}^{r_1} = \left(4C_{a_1r_1}^P C_{a_2r_2}^P s_{a_2}^{r_2} - C_{a_2r_1}^P C_{a_1r_2}^P s_{a_2}^{r_2} - C_{a_1a_2}^P C_{r_1r_2}^P s_{a_2}^{r_2} + (\omega_B)_{a_1}^{r_1}\right) / \left(\epsilon_{a_1} - \epsilon_{r_1}\right) \tag{62}$$

This allows contributions from $\nu_{r_1r_2}^{a_1a_2} s_{a_1}^{r_1}$ to be evaluated as $\mathcal{O}(ovN_{df})$, $\nu_{r_1r_2}^{a_2a_1} s_{a_1}^{r_1}$ as $\mathcal{O}(o^2vN_{df})$, and $\nu_{a_2r_2}^{a_1r_1} s_{a_1}^{r_1}$ as $\mathcal{O}(ov^2N_{df})$. The bottleneck associated with this implementation is the disk I/O associated with the $C_{r_1r_2}^P s_{a_2}^{r_2}$ contraction. To improve the efficiency of this algorithm, the two-fold permutational symmetry of the $C_{r_1r_2}^P$ DF integrals is exploited in their storage.

Algorithm 2 Pseudocode to solve for the CPHF coefficients of monomer A.

while s_a^r NOT converged **do**
  for $P_{block} <$ #Pblocks **do**
    Read C_{ar}^P block
    for $P < \#P_{block}$ **do**
      $X^P \leftarrow C_{a_2r_2}^P s_{a_2}^{r_2}$
      $Z_{a_1}^{r_1} \leftarrow 4C_{a_1r_1}^P X^P$
      $X_{a_1a_2} \leftarrow C_{a_1r_2}^P s_{a_2}^{r_2}$
      $Z_{a_1}^{r_1} \leftarrow -C_{a_2r_1}^P X_{a_1a_2}$
    end for
  end for
  for $P_{block} <$ #Pblocks **do**
    Read C_{aa}^P and C_{rr}^P blocks
    for $P < \#P_{block}$ **do**
      Unpack C_{rr}^P permutational symmetry
      $X_{a_2r_1} \leftarrow C_{r_1r_2}^P s_{a_2}^{r_2}$
      $Z_{a_1}^{r_1} \leftarrow -C_{a_1a_2}^P X_{a_2r_1}$
    end for
  end for
  $s_{a_1}^{r_1} = \left(Z_{a_1}^{r_1} + (\omega_B)_{a_1}^{r_1}\right) / \left(\epsilon_{a_1} - \epsilon_{r_1}\right)$
end while

2.2.6 $E^{(20)}_{exch-ind,r}(S^2)$

The $E^{(20)}_{ind,r}$ term is evaluated while maintaining only the antisymmetry of the monomer wavefunctions, not the antisymmetry of the monomer A wavefunction with the monomer B wavefunction. This additional exchange interaction is computed under the S^2 approximation with the CPHF coefficients used to evaluate $E^{(20)}_{ind,r}$.[110]

$$\begin{aligned} E^{(20)}_{exch-ind,r}(A \leftarrow B) = &- 2s^{a_1}_{r_1}\tilde{\nu}^{b_1 r_1}_{a_1 b_1} \\ &- 4s^{a_2}_{r_1}S^{b_1}_{a_2}\tilde{\nu}^{a_1 r_1}_{a_1 b_1} + 2s^{a_1}_{r_1}S^{b_1}_{a_2}\tilde{\nu}^{a_2 r_1}_{a_1 b_1} \\ &- 4s^{a_1}_{r_1}S^{b_1}_{a_2}\tilde{\nu}^{r_1 a_2}_{a_1 b_1} + 2s^{a_2}_{r_1}S^{b_1}_{a_2}\tilde{\nu}^{r_1 a_1}_{a_1 b_1} \\ &- 4s^{a_1}_{r_1}S^{r_1}_{b_2}\tilde{\nu}^{b_2 b_1}_{a_1 b_1} + 2s^{a_1}_{r_1}S^{r_1}_{b_2}\tilde{\nu}^{b_1 b_2}_{a_1 b_1} \\ &+ 4s^{a_2}_{r_1}S^{b_1}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{a_1 b_2}_{a_1 b_1} + 4s^{a_1}_{r_1}S^{b_2}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{a_2 b_1}_{a_1 b_1} - 2s^{a_1}_{r_1}S^{b_1}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{a_2 b_2}_{a_1 b_1} \\ &+ 4s^{a_1}_{r_1}S^{b_1}_{a_2}S^{a_2}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_1 b_1} + 4s^{a_2}_{r_1}S^{b_2}_{a_2}S^{a_1}_{b_2}\tilde{\nu}^{r_1 b_1}_{a_1 b_1} - 2s^{a_2}_{r_1}S^{b_1}_{a_2}S^{a_1}_{b_2}\tilde{\nu}^{r_1 b_2}_{a_1 b_1} \end{aligned} \tag{63}$$

This term is factorized in a similar manner to the $E^{(10)}_{exch}$ term; the difference here is that virtual orbitals and CPHF coefficients appear in the expression. This term is also negligible compared to the rest of a SAPT0 computation, however, for very large systems, the use of generalized DF integrals is necessary to avoid the formation of four-index integrals, which will be too large to store in memory. By blocking the evaluation of $E^{(20)}_{ind,r}$ over the P index, this term can be easily evaluated with a minimal memory requirement.

2.2.7 $E^{(20)}_{disp}$

When the interaction of nonpolar monomers is considered, the $E^{(20)}_{disp}$ correction provides the leading contribution to the interaction energy. This term has a very simple form,[110]

$$E^{(20)}_{disp} = 4t^{r_1 s_1}_{a_1 b_1}\nu^{a_1 b_1}_{r_1 s_1}, \tag{64}$$

however, this formulation requires the explicit formation of the t^{rs}_{ab} amplitudes resulting in a scaling of $\mathcal{O}(o^2 v^2 N_{df})$ if the DF approximation is applied. To clarify, this is a result of the energy denominator that appears in the expression,

$$E^{(20)}_{disp} = 4\nu^{r_1 s_1}_{a_1 b_1}\nu^{a_1 b_1}_{r_1 s_1}\Delta^{r_1 s_1}_{a_1 b_1}. \tag{65}$$

If we write the denominator,

$$\Delta_{a_1 b_1}^{r_1 s_1} = \frac{1}{\epsilon_{a_1} + \epsilon_{b_1} - \epsilon_{r_1} - \epsilon_{s_1}}, \tag{66}$$

and then take the Laplace transform of the energy denominator.[5,80]

$$\frac{1}{x} = \int_0^\infty e^{-xt} dt \tag{67}$$

$$\int_0^\infty e^{-(\epsilon_{a_1} + \epsilon_{b_1} - \epsilon_{r_1} - \epsilon_{s_1})t} dt = \int_0^\infty e^{(\epsilon_{r_1} - \epsilon_{a_1})t} e^{(\epsilon_{s_1} - \epsilon_{b_1})t} dt = \Delta_{a_1 b_1}^{r_1 s_1} \tag{68}$$

In practice, we will evaluate this integral numerically,[22,216] which leads to:

$$\Delta_{a_1 b_1}^{r_1 s_1} \approx d_{a_1 r_1}^L d_{b_1 s_1}^L, \tag{69}$$

where L indexes the integration points. Now, the $E_{disp}^{(20)}$ correction can be written with approximate, Laplace transformed denominators as,

$$E_{disp}^{(20)} = 4 \nu_{a_1 b_1}^{r_1 s_1} \nu_{r_1 s_1}^{a_1 b_1} d_{a_1 r_1}^L d_{b_1 s_1}^L, \tag{70}$$

the utility of which becomes obvious when we also invoke the DF representation of the two electron integrals:

$$E_{disp}^{(20)} = 4 C_{a_1 r_1}^P C_{b_1 s_1}^P C_{a_1 r_1}^Q C_{b_1 s_1}^Q d_{a_1 r_1}^L d_{b_1 s_1}^L. \tag{71}$$

Although it may not appear that the application of a Laplace transform leads to a simpler formulation of $E_{disp}^{(20)}$, the scaling of this term is reduced from $\mathcal{O}(N^5)$ to, per integration point, $\mathcal{O}(N^4)$. In practice, only 6-10 integration points are required for sufficient accuracy in typical systems.

2.2.8 $E_{exch-disp}^{(20)}(S^2)$

The $E_{exch-disp}^{(20)}$ evaluation is the most computationally demanding portion of a SAPT0 computation. Under the approximations considered, it is the only SAPT0 term that unavoidably

Algorithm 3 Pseudocode to evaluate $E^{(20)}_{disp}$ with Laplace transformed energy denominators.

for $P_{block} <$ #Pblocks **do**
  Read C^P_{ar} and C^P_{bs} blocks
  for $Q_{block} <$ #Qblocks **do**
    Read C^Q_{ar} and C^Q_{bs} blocks
    for $L <$ #IntegrationPoints **do**
      $T^Q_{a_1r_1} \leftarrow C^Q_{a_1r_1} d^L_{a_1r_1}$
      $T^Q_{b_1s_1} \leftarrow C^Q_{b_1s_1} d^L_{b_1s_1}$
      $X_{PQ} \leftarrow C^P_{a_1r_1} T^Q_{a_1r_1}$
      $Y_{PQ} \leftarrow C^P_{b_1s_1} T^Q_{b_1s_1}$
      $E^{(20)}_{disp} = -4 X_{PQ} Y_{PQ}$
    end for
  end for
end for

scales $\mathcal{O}(N^5)$. First, we will consider the canonical form of this correction:[110]

$$\begin{aligned}
E^{(20)}_{exch-disp}(S^2) = &- 2 t^{a_1 b_1}_{r_1 s_1} \Big[\tilde{\nu}^{s_1 r_1}_{a_1 b_1} \\
&+ S^{s_1}_{a_1} (2\tilde{\nu}^{a_2 r_1}_{a_2 b_1} - \tilde{\nu}^{r_1 a_2}_{a_2 b_1}) + S^{s_1}_{a_2} (2\tilde{\nu}^{r_1 a_2}_{a_1 b_1} - \tilde{\nu}^{a_2 r_1}_{a_1 b_1}) \\
&+ S^{r_1}_{b_1} (2\tilde{\nu}^{s_1 b_2}_{a_1 b_2} - \tilde{\nu}^{b_2 s_1}_{a_1 b_2}) + S^{r_1}_{b_2} (2\tilde{\nu}^{b_2 s_1}_{a_1 b_1} - \tilde{\nu}^{s_1 b_2}_{a_1 b_1}) \\
&+ S^{r_1}_{b_1} S^{b_2}_{a_2} \tilde{\nu}^{a_2 s_1}_{a_1 b_2} - 2 S^{r_1}_{b_2} S^{b_2}_{a_2} \tilde{\nu}^{a_2 s_1}_{a_1 b_1} - 2 S^{r_1}_{b_1} S^{b_2}_{a_1} \tilde{\nu}^{a_2 s_1}_{a_2 b_2} + 4 S^{r_1}_{b_2} S^{b_2}_{a_1} \tilde{\nu}^{a_2 s_1}_{a_2 b_1} \\
&+ S^{s_1}_{a_1} S^{a_2}_{b_2} \tilde{\nu}^{r_1 b_2}_{a_2 b_1} - 2 S^{s_1}_{a_2} S^{a_2}_{b_2} \tilde{\nu}^{r_1 b_2}_{a_1 b_1} - 2 S^{s_1}_{a_1} S^{a_2}_{b_1} \tilde{\nu}^{r_1 b_2}_{a_2 b_2} + 4 S^{s_1}_{a_2} S^{a_2}_{b_1} \tilde{\nu}^{r_1 b_2}_{a_1 b_2} \\
&+ S^{r_1}_{b_2} S^{s_1}_{a_2} \tilde{\nu}^{a_2 b_2}_{a_1 b_1} - 2 S^{r_1}_{b_1} S^{s_1}_{a_2} \tilde{\nu}^{a_2 b_2}_{a_1 b_2} - 2 S^{r_1}_{b_2} S^{s_1}_{a_1} \tilde{\nu}^{a_2 b_2}_{a_2 b_1} \\
&+ S^{a_2}_{b_1} S^{b_2}_{a_1} \tilde{\nu}^{r_1 s_1}_{a_2 b_2} - 2 S^{a_2}_{b_1} S^{b_2}_{a_2} \tilde{\nu}^{r_1 s_1}_{a_1 b_2} - 2 S^{a_2}_{b_2} S^{b_2}_{a_1} \tilde{\nu}^{r_1 s_1}_{a_2 b_1} \Big]. \qquad (72)
\end{aligned}$$

Note that this formula contains errors as given in Ref. 110. As written above, this term can be implemented with $\mathcal{O}(o^3v^2)$ scaling. Heßelmann and Jansen propose a $\mathcal{O}(o^2v^2N_{df})$ factorization of the $t^{a_1b_1}_{r_1s_1} S^{a_2}_{b_1} S^{b_2}_{a_1} \tilde{\nu}^{r_1s_1}_{a_2b_2}$ term that uses DF integrals. It is also possible to use a $\mathcal{O}(o^4v^2)$ algorithm to evaluate some of these terms; for small systems, the $\mathcal{O}(N^6)$ algorithm is surprisingly competitive with the others.

To find the optimal factorization of the $E^{(20)}_{exch-disp}$ term, we will examine each term separately and group similar terms together. First, we will look at the terms that unavoidably

scale, with DF integrals and Laplace transformed energy denominators, $\mathcal{O}(N^5)$.

$$E^{(20)}_{exch-disp}(S^2)(N^5) = -2t^{a_1b_1}_{r_1s_1}\Big[\tilde{\nu}^{s_1r_1}_{a_1b_1} - S^{s_1}_{a_2}\tilde{\nu}^{a_2r_1}_{a_1b_1} - S^{r_1}_{b_2}\tilde{\nu}^{s_1b_2}_{a_1b_1} + S^{r_1}_{b_2}S^{s_1}_{a_2}\tilde{\nu}^{a_2b_2}_{a_1b_1} + S^{a_2}_{b_1}S^{b_2}_{a_1}\tilde{\nu}^{r_1s_1}_{a_2b_2}\Big] \tag{73}$$

Due to the exchange integrals that appear in this expression, the use of Laplace transformed energy denominators does not lead to any computational savings. The introduction of DF integrals, however, is helpful. First we will define a few simple intermediates:

$$\begin{aligned} X^P_{a_1s_1} &= S^{s_1}_{a_2}A^P_{a_1a_2} \\ X^P_{b_1r_1} &= S^{r_1}_{b_2}B^P_{b_1b_2} \\ Y^P_{a_1s_1} &= S^{b_2}_{a_1}C^P_{b_2s_1} \\ Y^P_{b_1r_1} &= S^{a_2}_{b_1}C^P_{a_2r_1}. \end{aligned} \tag{74}$$

These intermediates allow us to rewrite these $\mathcal{O}(N^5)$ terms in a particularly simple form.

$$E^{(20)}_{exch-disp}(S^2)(N^5) = -2t^{a_1b_1}_{r_1s_1}(A^P_{a_1s_1} - X^P_{a_1s_1})(B^P_{b_1r_1} - X^P_{b_1r_1}) - 2t^{a_1b_1}_{r_1s_1}Y^P_{a_1s_1}Y^P_{b_1r_1} \tag{75}$$

With this factorization, the contributions from these five terms can be evaluated with three $\mathcal{O}(o^2v^2N_{df})$ multiplications. The dispersion amplitudes do need to be formed, but they do not need to be stored. As a byproduct of this procedure, the $E^{(20)}_{disp}$ term can be obtained. Up to this point, it has been convenient to store the DF integrals with the auxiliary index running slowest, now, since two auxiliary indices appear in each term, it is better to transpose the DF integrals. This is also an opportunity to remove frozen occupied orbitals from the DF integral arrays.

While the Laplace transformed energy denominators does not help with all of the $E^{(20)}_{exch-disp}$ terms, it can reduce the scaling of some of the terms.

$$E^{(20)}_{exch-disp}(S^2)(\mathcal{L}) = -2t^{a_1b_1}_{r_1s_1}\Big[2S^{s_1}_{a_2}\tilde{\nu}^{r_1a_2}_{a_1b_1} - 2S^{s_1}_{a_2}S^{a_2}_{b_2}\tilde{\nu}^{r_1b_2}_{a_1b_1} - 2S^{a_2}_{b_1}S^{b_2}_{a_2}\tilde{\nu}^{r_1s_1}_{a_1b_2}\Big] - 2t^{a_1b_1}_{r_1s_1}\Big[2S^{r_1}_{b_2}\tilde{\nu}^{b_2s_1}_{a_1b_1} - 2S^{r_1}_{b_2}S^{b_2}_{a_2}\tilde{\nu}^{a_2s_1}_{a_1b_1} - 2S^{a_2}_{b_2}S^{b_2}_{a_1}\tilde{\nu}^{r_1s_1}_{a_2b_1}\Big] \tag{76}$$

Algorithm 4 Pseudocode to evaluate the $E^{(20)}_{exch-disp}(S^2)(N^5)$ contributions.

for $P_{block} <$ #Pblocks **do**
  Read active C^P_{ar} block
  Transpose and write C^P_{ar} block to disk
end for
for $P_{block} <$ #Pblocks **do**
  Read active C^P_{bs} block
  Transpose and write C^P_{bs} block to disk
end for
for $P_{block} <$ #Pblocks **do**
  Read A^P_{aa} and A^P_{as} blocks
  $X^P_{a_1 s_1} \leftarrow A^P_{a_1 s_1} - S^{s_1}_{a_2} A^P_{a_1 a_2}$
  Transpose and write X^P_{as} block to disk
end for
for $P_{block} <$ #Pblocks **do**
  Read B^P_{bb} and B^P_{rb} blocks
  $X^P_{b_1 r_1} \leftarrow B^P_{r_1 b_1} - S^{r_1}_{b_2} B^P_{b_1 b_2}$
  Transpose and write X^P_{br} block to disk
end for
for $P_{block} <$ #Pblocks **do**
  Read C^P_{ar} block
  $Y^P_{b_1 r_1} \leftarrow S^{a_2}_{b_1} C^P_{a_2 r_1}$
  Transpose and write Y^P_{br} block to disk
end for
for $P_{block} <$ #Pblocks **do**
  Read C^P_{bs} block
  $Y^P_{a_1 s_1} \leftarrow S^{b_2}_{a_1} C^P_{b_2 s_1}$
  Transpose and write Y^P_{as} block to disk
end for
for $A_{block} <$ #Ablocks **do**
  Read C^P_{ar}, X^P_{as}, and Y^P_{as} blocks
  for $B_{block} <$ #Bblocks **do**
    Read C^P_{bs}, X^P_{br}, and Y^P_{br} blocks
    for $a < \#A_{block}$ **do**
      for $b < \#B_{block}$ **do**
        $t^{r_1 s_1}_{a_1 b_1} \leftarrow C^P_{a_1 r_1} C^P_{b_1 s_1} \Delta^{r_1 s_1}_{a_1 b_1}$
        $E^{(20)}_{exch-disp} \leftarrow -2 t^{r_1 s_1}_{a_1 b_1} X^P_{a_1 s_1} X^P_{b_1 r_1}$
        $E^{(20)}_{exch-disp} \leftarrow -2 t^{r_1 s_1}_{a_1 b_1} Y^P_{a_1 s_1} Y^P_{b_1 r_1}$
      end for
    end for
  end for
end for

To explicitly show the origin of this savings, we first must recognize two useful intermediates that could be formed.

$$T^P_{a_1 r_1} = t^{a_1 b_1}_{r_1 s_1} B^P_{b_1 s_1} \approx -C^Q_{a_1 r_1} C^Q_{b_1 s_1} B^P_{b_1 s_1} d^L_{a_1 r_1} d^L_{b_1 s_1}$$
$$T^P_{b_1 s_1} = t^{a_1 b_1}_{r_1 s_1} A^P_{a_1 r_1} \approx -C^Q_{a_1 r_1} C^Q_{b_1 s_1} A^P_{a_1 r_1} d^L_{a_1 r_1} d^L_{b_1 s_1} \tag{77}$$

This allows us to write:

$$\begin{aligned} E^{(20)}_{exch-disp}(S^2)(\mathcal{L}) = &- 2T^P_{b_1 s_1} \Big[2S^{s_1}_{a_2} B^P_{a_2 b_1} - 2S^{s_1}_{a_2} S^{a_2}_{b_2} B^P_{b_1 b_2} - 2S^{a_2}_{b_1} S^{b_2}_{a_2} B^P_{b_2 s_1} \Big] \\ &- 2T^P_{a_1 r_1} \Big[2S^{r_1}_{b_2} A^P_{a_1 b_2} - 2S^{r_1}_{b_2} S^{b_2}_{a_2} A^P_{a_1 a_2} - 2S^{a_2}_{b_2} S^{b_2}_{a_1} A^P_{a_2 r_1} \Big]. \end{aligned} \tag{78}$$

Using canonical energy denominators, the formation of these intermediates would scale $\mathcal{O}(o^2 v^2 N_{df})$; if the Laplace transform is applied to the energy denominator, this scaling can be reduced to $\mathcal{O}(ovN^2_{df})$. Once these intermediates are available, their contribution to $E^{(20)}_{exch-disp}$ can be evaluated simply by contracting the overlap integrals with the DF integrals that appear in a certain term and then multiplying by the appropriate T type intermediate. The remaining terms can be reduced to a dispersion amplitude and two two-index arrays. Structurally, there are two types of these terms, $t^{a_1 b_1}_{r_1 s_1} X_{a_1 r_1} Y_{b_1 s_1}$ and $t^{a_1 b_1}_{r_1 s_1} X_{a_1 s_1} Y_{b_1 r_1}$. When the dispersion amplitudes are written with Laplace transformed energy denominators and DF integrals, the former can be evaluated with $\mathcal{O}(ovN_{df})$ work, whereas the latter can be evaluated with $\mathcal{O}(o^2 v N_{df})$ work.

$$\begin{aligned} E^{(20)}_{exch-disp}(S^2)(N^3, N^4) = &- 2t^{a_1 b_1}_{r_1 s_1} \Big[S^{s_1}_{a_1} (2\tilde{\nu}^{a_2 r_1}_{a_2 b_1} - \tilde{\nu}^{r_1 a_2}_{a_2 b_1}) + S^{r_1}_{b_1} (2\tilde{\nu}^{s_1 b_2}_{a_1 b_2} - \tilde{\nu}^{b_2 s_1}_{a_1 b_2}) \\ &+ S^{r_1}_{b_1} S^{b_2}_{a_2} \tilde{\nu}^{a_2 s_1}_{a_1 b_2} - 2S^{r_1}_{b_1} S^{b_2}_{a_1} \tilde{\nu}^{a_2 s_1}_{a_2 b_2} + 4S^{r_1}_{b_2} S^{b_2}_{a_1} \tilde{\nu}^{a_2 s_1}_{a_2 b_1} \\ &+ S^{s_1}_{a_1} S^{a_2}_{b_2} \tilde{\nu}^{r_1 b_2}_{a_2 b_1} - 2S^{s_1}_{a_1} S^{a_2}_{b_1} \tilde{\nu}^{r_1 b_2}_{a_2 b_2} + 4S^{s_1}_{a_2} S^{a_2}_{b_1} \tilde{\nu}^{r_1 b_2}_{a_1 b_2} \\ &- 2S^{r_1}_{b_1} S^{s_1}_{a_2} \tilde{\nu}^{a_2 b_2}_{a_1 b_2} - 2S^{r_1}_{b_2} S^{s_1}_{a_1} \tilde{\nu}^{a_2 b_2}_{a_2 b_1} \Big] \end{aligned} \tag{79}$$

2.3 Intramonomer Correlation Corrections

2.3.1 $E^{(12)}_{elst,r}$

The $E^{(12)}_{elst,r}$ correction attempts to correct the electrostatic energy by using second-order correlation corrections to the density to improve the electrostatic energies computed with

Algorithm 5 Pseudocode to form T^P_{ar} intermediates with Laplace transformed energy denominators.

for $P_{block} <$ #Pblocks **do**
  Read B^P_{bs} block
  for $Q_{block} <$ #Qblocks **do**
    Read C^Q_{bs} block
    for $L <$ #IntegrationPoints **do**
      $D^Q_{b_1 s_1} \leftarrow C^Q_{b_1 s_1} d^L_{b_1 s_1}$
      $X^L_{PQ} \leftarrow B^P_{b_1 s_1} D^Q_{b_1 s_1}$
    end for
  end for
end for
for $A_{block} <$ #Ablocks **do**
  Read C^P_{ar} block
  for $a < \#A_{block}$ **do**
    for $L <$ #IntegrationPoints **do**
      $D^P_{a_1 r_1} \leftarrow C^P_{a_1 r_1} d^L_{a_1 r_1}$
      $T^Q_{a_1 r_1} \leftarrow X^L_{PQ} D^P_{a_1 r_1}$
    end for
    Transpose and write T^Q_{ar} to disk
  end for
end for

Algorithm 6 Pseudocode to evaluate the types of terms that appear in $E^{(20)}_{exch-disp}(N^3, N^4)$.

for $P_{block} <$ #Pblocks **do**
  Read C^P_{ar} and C^P_{bs} blocks
  for $L <$ #IntegrationPoints **do**
    $D^P_{a_1 r_1} \leftarrow C^P_{a_1 r_1} d^L_{a_1 r_1}$
    $D^P_{b_1 s_1} \leftarrow C^P_{b_1 s_1} d^L_{b_1 s_1}$
    $X^P \leftarrow D^P_{a_1 r_1} X_{a_1 r_1}$
    $Y^P \leftarrow D^P_{b_1 s_1} Y_{b_1 s_1}$
    $E^{(20)}_{exch-disp} \leftarrow X^P Y^P$
    for $P < \#P_{block}$ **do**
      $X_{a_1 b_1} \leftarrow D^P_{a_1 r_1} X_{b_1 r_1}$
      $Y_{a_1 b_1} \leftarrow D^P_{b_1 s_1} Y_{a_1 s_1}$
      $E^{(20)}_{exch-disp} \leftarrow X_{a_1 b_1} Y_{a_1 b_1}$
    end for
  end for
end for

Hartree-Fock densities.[110]

$$E^{(120)}_{elst,resp} = 2\theta^{a_1a_2}_{r_1r_2} t^{r_1r_3}_{a_1a_2} (\omega_B)^{r_2}_{r_3} - 2\theta^{a_1a_2}_{r_1r_2} t^{r_1r_2}_{a_1a_3} (\omega_B)^{a_2}_{a_3} + 4Re \overset{(2)}{Y}{}^{a_1}_{r_1} s^{r_1}_{a_1} \tag{80}$$

Here, $s^{r_1}_{a_1}$ is a solution to the CPHF equations. The first two terms can be evaluated efficiently by realizing that the contraction of the amplitudes results in the *oo* and *vv* blocks of the unrelaxed MP2 one-particle density matrix (OPDM).

$$P^{a_2}_{a_3} = \theta^{a_1a_2}_{r_1r_2} t^{r_1r_2}_{a_1a_3} \tag{81}$$

$$P^{r_2}_{r_3} = \theta^{a_1a_2}_{r_1r_2} t^{r_1r_3}_{a_1a_2} \tag{82}$$

This results in a simpler form for the correction:

$$E^{(120)}_{elst,resp} = 2P^{r_2}_{r_1} (\omega_B)^{r_1}_{r_2} - 2P^{a_2}_{a_1} (\omega_B)^{a_1}_{a_2} + 4Re \overset{(2)}{Y}{}^{a_1}_{r_1} s^{r_1}_{a_1}. \tag{83}$$

The work in evaluating the $E^{(12)}_{elst,r}$ correction is in the formation of $\overset{(2)}{Y}$.

$$\begin{aligned} \overset{(2)}{Y}{}^{a_1}_{r_1} = {} & \theta^{a_4a_2}_{r_4r_2} t^{r_4r_3}_{a_4a_2} (2\nu^{a_1r_2}_{r_1r_3} - \nu^{r_2a_1}_{r_1r_3}) - \theta^{a_4a_2}_{r_4r_2} t^{r_4r_2}_{a_4a_3} (2\nu^{a_1a_3}_{r_1a_2} - \nu^{a_3a_1}_{r_1a_2}) \\ & + \theta^{a_3a_1}_{r_3r_2} \nu^{r_3r_2}_{a_3r_1} - \theta^{a_3a_2}_{r_3r_1} \nu^{r_3a_1}_{a_3a_2} \end{aligned} \tag{84}$$

This is the first place where the formation of the $\Theta^P_{a_1r_1}$ intermediates is useful.

$$\Theta^P_{a_1r_1} = \theta^{r_1r_2}_{a_1a_2} A^P_{a_2r_2} \tag{85}$$

The introduction of $\Theta^P_{a_1r_1}$ intermediates along with density fitting integrals and MP2 OPDM allows the $\overset{(2)}{Y}{}^{a_1}_{r_1}$ to be written as:

$$\begin{aligned} \overset{(2)}{Y}{}^{a_1}_{r_1} = {} & 2P^{r_2}_{r_3} C^P_{a_1r_1} C^P_{r_2r_3} - P^{r_2}_{r_3} C^P_{a_1r_3} C^P_{r_1r_2} - 2P^{a_2}_{a_3} C^P_{a_1r_1} C^P_{a_2a_3} + P^{a_2}_{a_3} C^P_{a_3r_1} C^P_{a_1a_2} \\ & + \Theta^P_{a_1r_2} C^P_{r_1r_2} - \Theta^P_{a_2r_1} C^P_{a_1a_2}. \end{aligned} \tag{86}$$

This factorization of the $E^{(12)}_{elst,r}$ correction does not reduce the scaling, $\mathcal{O}(N^5)$, but rather avoids the need to from and operate on the four-index *ovvv* integrals. By reformulating the equations in terms of three-index quantities, this potential I/O bottleneck is removed.

2.3.2 $E^{(13)}_{elst,r}$

The third-order correction to the electrostatic energy was a similar form to $E^{(12)}_{elst,r}$, however, the amplitudes required to evaluate this correction are more involved.[110]

$$E^{(130)}_{elst,resp} = 4Re(\overset{(2)}{\theta}{}^{a_1a_2}_{r_1r_2} t^{r_1r_3}_{a_1a_2}(\omega_B)^{r_2}_{r_3} - \overset{(2)}{\theta}{}^{a_1a_2}_{r_1r_2} t^{r_1r_2}_{a_1a_3}(\omega_B)^{a_2}_{a_3} + \overset{(3)}{Y}{}^{a_1}_{r_1} s^{r_1}_{a_1}) \tag{87}$$

The first two terms can be evaluated in a similar manner to the $E^{(12)}_{elst,r}$ correction with the following contractions:

$$Q^{a_2}_{a_3} = \overset{(2)}{\theta}{}^{a_1a_2}_{r_1r_2} t^{r_1r_2}_{a_1a_3} \tag{88}$$

$$Q^{r_2}_{r_3} = \overset{(2)}{\theta}{}^{a_1a_2}_{r_1r_2} t^{r_1r_3}_{a_1a_2}. \tag{89}$$

Again, the more involved step in the evaluation of this term is the formation of the $\overset{(3)}{Y}$,

$$\begin{aligned}
\overset{(3)}{Y}{}^{a_1}_{r_1} = \Big[& \overset{(2)}{\theta}{}^{a_3a_1}_{r_3r_2} \nu^{r_3r_2}_{a_3r_1} - \overset{(2)}{\theta}{}^{a_3a_2}_{r_3r_1} \nu^{r_3a_1}_{a_3a_2} \\
& + \overset{(2)}{\theta}{}^{a_4a_2}_{r_4r_2} t^{r_4r_3}_{a_4a_2}(2\nu^{a_1r_2}_{r_1r_3} - \nu^{r_2a_1}_{r_1r_3}) + \theta^{a_4a_2}_{r_4r_2} \overset{(2)}{t}{}^{r_4r_3}_{a_4a_2}(2\nu^{a_1r_2}_{r_1r_3} - \nu^{r_2a_1}_{r_1r_3}) \\
& - \overset{(2)}{\theta}{}^{a_4a_2}_{r_4r_2} t^{r_4r_2}_{a_4a_3}(2\nu^{a_1a_3}_{r_1a_2} - \nu^{a_3a_1}_{r_1a_2}) - \theta^{a_4a_2}_{r_4r_2} \overset{(2)}{t}{}^{r_4r_2}_{a_4a_3}(2\nu^{a_1a_3}_{r_1a_2} - \nu^{a_3a_1}_{r_1a_2}) \\
& + \overset{(2)}{\gamma}{}^{a_1a_2}_{a_4a_3}(2\nu^{a_4a_3}_{r_1a_2} - \nu^{a_3a_4}_{r_1a_2}) - \overset{(2)}{\gamma}{}^{r_4r_3}_{r_1r_2}(2\nu^{a_1r_2}_{r_4r_3} - \nu^{r_2a_1}_{r_4r_3}) \\
& + \overset{(2)}{\gamma}{}^{a_1r_3}_{r_2a_3}(2\nu^{a_3r_2}_{r_3r_1} - \nu^{r_2a_3}_{r_3r_1}) - \overset{(2)}{\bar{\gamma}}{}^{a_1r_3}_{r_2a_3}(2\nu^{r_2a_3}_{r_3r_1} - \nu^{a_3r_2}_{r_3r_1}) \\
& - \overset{(2)}{\gamma}{}^{a_2r_3}_{r_1a_3}(2\nu^{a_1a_3}_{a_2r_3} - \nu^{a_3a_1}_{a_2r_3}) + \overset{(2)}{\bar{\gamma}}{}^{a_2r_3}_{r_1a_3}(2\nu^{a_3a_1}_{a_2r_3} - \nu^{a_1a_3}_{a_2r_3}) \Big] \Delta^{r_1}_{a_1},
\end{aligned} \tag{90}$$

where the $\overset{(2)}{\gamma}$ quantities are defined as:

$$\overset{(2)}{\gamma}{}^{a_1a_2}_{a_4a_3} = t^{r_1r_2}_{a_4a_3} t^{a_1a_2}_{r_1r_2} \tag{91}$$

$$\overset{(2)}{\gamma}{}^{r_4r_3}_{r_1r_2} = t^{r_4r_3}_{a_1a_2} t^{a_1a_2}_{r_1r_2} \tag{92}$$

$$\overset{(2)}{\gamma}{}^{a_1r_2}_{r_1a_2} = \theta^{a_1a_3}_{r_1r_3} t^{a_3a_2}_{r_3r_2} - t^{a_1a_3}_{r_1r_3} t^{r_2r_3}_{a_3a_2} \tag{93}$$

$$\overset{(2)}{\bar{\gamma}}{}^{a_1r_2}_{r_1a_2} = t^{a_3a_1}_{r_1r_3} t^{r_3r_2}_{a_2a_3}. \tag{94}$$

The first six terms that contribute to $\overset{(3)}{Y}$ are structurally identical to the $\overset{(2)}{Y}$ and can be evaluated simply. This uses the Q intermediates defined above as well as:

$$\bar{Q}^{a_2}_{a_3} = \theta^{a_1a_2}_{r_1r_2} \overset{(2)}{t}{}^{r_1r_2}_{a_1a_3} \tag{95}$$

$$\bar{Q}^{r_2}_{r_3} = \theta^{a_1a_2}_{r_1r_2} \overset{(2)}{t}{}^{r_1r_3}_{a_1a_2}. \tag{96}$$

It is the $\overset{(2)}{\gamma}$ quantities that are problematic; in particular, the $\overset{(2)}{\gamma}{}^{r_4 r_3}_{r_1 r_2}$ term is especially difficult. To directly form this quantity would scale $\mathcal{O}(o^2v^4)$ and, therefore, must be avoided. The preferred method is described below.

$$X^{a_1 r_2}_{a_2 a_3} = t^{r_4 r_3}_{a_2 a_3}(2\nu^{a_1 r_2}_{r_4 r_3} - \nu^{r_2 a_1}_{r_4 r_3}) \tag{97}$$

$$\overset{(3)}{Y}{}^{a_1}_{r_1} \leftarrow X^{a_1 r_2}_{a_2 a_3} t^{a_2 a_3}_{r_1 r_2} \tag{98}$$

The formation of the *ovvv* integrals scales as $\mathcal{O}(ov^3 N_{df})$, followed by the contraction of these integrals with the doubles amplitudes, which scales as $\mathcal{O}(o^3v^3)$. The *ovvv* integrals can be formed in batches and do not need to be stored. The $\overset{(2)}{\gamma}{}^{a_1 a_2}_{a_4 a_3}$ term scales $\mathcal{O}(o^4v^2)$ to form followed by an $\mathcal{O}(o^4v)$ contraction to evaluate its contribution to $\overset{(3)}{Y}$. The $\overset{(2)}{\gamma}{}^{a_1 r_3}_{r_2 a_3}$ and $\overset{(2)}{\bar{\gamma}}{}^{a_1 r_3}_{r_2 a_3}$ quantities are also evaluated canonically (scaling $\mathcal{O}(o^3v^3)$), but DF integrals are introduced in order to evaluate their contribution to $\overset{(3)}{Y}$. By defining a few new intermediates,

$$Y^P_{a_1 r_2} = \overset{(2)}{\gamma}{}^{a_1 r_3}_{r_2 a_3} C^P_{a_3 r_3} \tag{99}$$

$$\bar{Y}^P_{a_1 r_2} = \overset{(2)}{\bar{\gamma}}{}^{a_1 r_3}_{r_2 a_3} C^P_{a_3 r_3} \tag{100}$$

$$Y^P_{a_1 a_3} = \overset{(2)}{\gamma}{}^{a_1 r_3}_{r_2 a_3} C^P_{r_2 r_3} \tag{101}$$

$$Y^P_{r_1 r_3} = \overset{(2)}{\gamma}{}^{a_2 r_3}_{r_1 a_3} C^P_{a_2 a_3} \tag{102}$$

$$\bar{Y}^P_{a_1 a_3} = \overset{(2)}{\bar{\gamma}}{}^{a_1 r_3}_{r_2 a_3} C^P_{r_2 r_3} \tag{103}$$

$$\bar{Y}^P_{r_1 r_3} = \overset{(2)}{\bar{\gamma}}{}^{a_2 r_3}_{r_1 a_3} C^P_{a_2 a_3}, \tag{104}$$

we can rewrite the entirety of the $E^{(130)}_{elst,r}$ correction in a form that describes its implementation.

$$E^{(130)}_{elst,resp} = 4Re(Q^{r_3}_{r_2}(\omega_B)^{r_2}_{r_3} - Q^{a_3}_{a_2}(\omega_B)^{a_2}_{a_3} + \overset{(3)}{Y}{}^{a_1}_{r_1} s^{r_1}_{a_1}) \tag{105}$$

$$
\begin{aligned}
\overset{(3)}{Y}{}^{a_1}_{r_1} = \Big[& \overset{(2)}{\Theta}{}^{P}_{a_1 r_2} C^{P}_{r_1 r_2} - \overset{(2)}{\Theta}{}^{P}_{a_2 r_1} C^{P}_{a_1 a_2} \\
& + Q^{r_2}_{r_3} (2\nu^{a_1 r_2}_{r_1 r_3} - \nu^{r_2 a_1}_{r_1 r_3}) + \bar{Q}^{r_2}_{r_3} (2\nu^{a_1 r_2}_{r_1 r_3} - \nu^{r_2 a_1}_{r_1 r_3}) \\
& - Q^{a_2}_{a_3} (2\nu^{a_1 a_3}_{r_1 a_2} - \nu^{a_3 a_1}_{r_1 a_2}) - \bar{Q}^{a_2}_{a_3} (2\nu^{a_1 a_3}_{r_1 a_2} - \nu^{a_3 a_1}_{r_1 a_2}) \\
& + \overset{(2)}{\gamma}{}^{a_1 a_2}_{a_4 a_3} (2\nu^{a_4 a_3}_{r_1 a_2} - \nu^{a_3 a_4}_{r_1 a_2}) - X^{a_1 r_2}_{a_2 a_3} t^{a_2 a_3}_{r_1 r_2} \\
& + 2Y^{P}_{a_1 r_2} C^{P}_{r_1 r_2} - Y^{P}_{a_1 a_3} C^{P}_{a_3 r_1} - 2\bar{Y}^{P}_{a_1 a_3} C^{P}_{a_3 r_1} + \bar{Y}^{P}_{a_1 r_2} C^{P}_{r_1 r_2} \\
& - 2Y^{P}_{a_2 r_1} C^{P}_{a_1 a_2} + Y^{P}_{r_1 r_3} C^{P}_{a_1 r_3} + 2\bar{Y}^{P}_{r_1 r_3} C^{P}_{a_1 r_3} - \bar{Y}^{P}_{a_2 r_1} C^{P}_{a_1 a_2} \Big] \Delta^{r_1}_{a_1}
\end{aligned} \tag{106}
$$

The equations for the $E^{(13)}_{elst,r}$ correction as given in Ref. 110 contain errors.

2.3.3 $E^{(11)}_{exch}$

The first-order intramonomer correction to exchange, $E^{(110)}_{exch}$, it the first example of generalized DF integrals used in higher-order SAPT. In our previous work, the second quantized form of correction was evaluated in order to exploit DF factorizations. Now, the more efficient interaction density matrix formulation of the term can be used. This correction is given as:[154]

$$
\begin{aligned}
E^{(110)}_{exch} = & -2\theta^{r_1 r_2}_{a_1 a_2} \tilde{\nu}^{a_1 a_2}_{r_1 b_1} S^{b_1}_{r_2} - 2\theta^{a_1 a_2}_{r_1 r_2} \tilde{\nu}^{r_1 r_2}_{a_1 b_1} S^{b_1}_{a_2} \\
& + 4Re\theta^{a_1 a_2}_{r_1 r_2} (\tilde{\nu}^{r_2 b_2}_{a_2 b_1} S^{b_1}_{a_1} S^{r_1}_{b_2} - 2\tilde{\nu}^{r_2 b_1}_{a_2 b_1} S^{b_2}_{a_1} S^{r_1}_{b_2}).
\end{aligned} \tag{107}
$$

Using intermediates that have been formed with the appropriate generalized DF integrals, we can rewrite $E^{(110)}_{exch}$ as:

$$
\begin{aligned}
E^{(110)}_{exch} = & -2\Theta^{P}_{a_2 r_2} B^{P}_{a_2 b_1} S^{b_1}_{r_2} - 2\Theta^{P}_{a_2 r_2} B^{P}_{r_2 b_1} S^{b_1}_{a_2} \\
& + 4\Theta^{P}_{a_1 r_1} B^{P}_{b_1 b_2} S^{b_1}_{a_1} S^{r_1}_{b_2} - 8\Theta^{P}_{a_1 r_1} B^{P}_{b_1 b_1} S^{b_2}_{a_1} S^{r_1}_{b_2}.
\end{aligned} \tag{108}
$$

This allows the $E^{(110)}_{exch}$ correction to be evaluated in terms of dressed DF integrals, amplitudes, and overlap integrals. Ignoring the formation of Θ, the scaling of the $E^{(110)}_{exch}$ evaluation is changed from $\mathcal{O}(o^3v^2)$ to $\mathcal{O}(o^2vN_{df})$; in practice, this is almost always a reduction. As written in Ref. 110, this formula (and subsequently, the formula for $E^{(12)}_{exch}K^{u}_{2}$) contain errors.

2.3.4 $E_{exch}^{(111)}$

The bilinear exchange term is given as:[154]

$$E_{exch}^{(111)} = -4Re\theta_{a_1a_2}^{r_1r_2}\theta_{b_1b_2}^{s_1s_2}\tilde{\nu}_{r_1s_1}^{a_1b_1}S_{r_2}^{b_2}S_{s_2}^{a_2} - 4Re\theta_{r_1r_2}^{a_1a_2}\theta_{s_1s_2}^{b_1b_2}\tilde{\nu}_{a_1s_1}^{r_1b_1}S_{s_2}^{r_2}S_{a_2}^{b_2}. \tag{109}$$

For this correction, particularly great savings can be realized, since the canonical evaluation of the term is $\mathcal{O}(o^3v^3)$. The scaling of this term can be reduced to $\mathcal{O}(o^2v^2N_{df})$ if appropriate intermediates are formed.

$$E_{exch}^{(111)} = -4Re\Theta_{a_1r_1}^{P}\Theta_{b_1s_1}^{P}S_{r_1}^{b_1}S_{s_1}^{a_1} - 4Re\Theta_{a_1r_1}^{P}\Theta_{b_1s_1}^{P}S_{s_1}^{r_1}S_{a_1}^{b_1} \tag{110}$$

This particularly convenient form of $E_{exch}^{(111)}$ justifies the use of the Θ type intermediates in other terms (*e.g.* $E_{exch}^{(110)}$) that scale better than $\mathcal{O}(o^2v^2N_{df})$ canonically.

2.3.5 $E_{exch}^{(12)}K_2^u$

This second-order intramonomer correlation correction is identical to the $E_{exch}^{(11)}$ correction with the exception that the doubles amplitudes that appear in this correction are the second-order doubles amplitudes, $\overset{(2)}{t}{}_{a_1a_2}^{r_1r_2}$. As such, it can be evaluated in an identical manner using slightly different intermediates.[110]

$$\begin{aligned} E_{exch}^{(120)}K_2^u = &- 2\overset{(2)}{\Theta}{}_{a_2r_2}^{P}B_{a_2b_1}^{P}S_{r_2}^{b_1} - 2\overset{(2)}{\Theta}{}_{a_2r_2}^{P}B_{r_2b_1}^{P}S_{a_2}^{b_1} \\ &+ 4\overset{(2)}{\Theta}{}_{a_1r_1}^{P}B_{b_1b_2}^{P}S_{a_1}^{b_1}S_{b_2}^{r_1} - 8\overset{(2)}{\Theta}{}_{a_1r_1}^{P}B_{b_1b_1}^{P}S_{a_1}^{b_2}S_{b_2}^{r_1}. \end{aligned} \tag{111}$$

2.3.6 $E_{exch}^{(12)}K_2^f$

Another of the second-order intramonomer correlation correction to exchange is given by:[110]

$$\begin{aligned} E_{exch}^{(120)}K_2^f = &- 2t_{a_1}^{r_1}\tilde{\nu}_{r_1b_1}^{b_1a_1} - 2t_{r_1}^{a_1}\tilde{\nu}_{a_1b_1}^{b_1r_1} - 4t_{r_1}^{a_2}S_{a_2}^{b_1}\tilde{\nu}_{a_1b_1}^{a_1r_1} - 4t_{a_1}^{r_1}S_{a_2}^{b_1}\tilde{\nu}_{r_1b_1}^{a_1a_2} \\ &- 4t_{r_1}^{a_1}S_{a_2}^{b_1}\tilde{\nu}_{a_1b_1}^{r_1a_2} - 4t_{a_2}^{r_1}S_{r_1}^{b_1}\tilde{\nu}_{a_1b_1}^{a_1a_2} + 2t_{a_2}^{r_1}S_{r_1}^{b_1}\tilde{\nu}_{a_1b_1}^{a_2a_1} + 2t_{r_1}^{a_2}S_{a_2}^{b_1}\tilde{\nu}_{a_1b_1}^{r_1a_1} \\ &+ 2t_{a_1}^{r_1}S_{a_2}^{b_1}\tilde{\nu}_{r_1b_1}^{a_2a_1} + 2t_{r_1}^{a_1}S_{a_2}^{b_1}\tilde{\nu}_{a_1b_1}^{a_2r_1} - 4t_{r_1}^{a_1}S_{b_2}^{r_1}\tilde{\nu}_{a_1b_1}^{b_2b_1} - 4t_{a_1}^{r_1}S_{b_2}^{a_1}\tilde{\nu}_{r_1b_1}^{b_2b_1} \\ &+ 2t_{r_1}^{a_1}S_{b_2}^{r_1}\tilde{\nu}_{a_1b_1}^{b_1b_2} + 2t_{a_1}^{r_1}S_{b_2}^{a_1}\tilde{\nu}_{r_1b_1}^{b_1b_2} + 8Ret_{r_1}^{a_2}S_{a_2}^{b_1}S_{b_2}^{r_1}\tilde{\nu}_{a_1b_1}^{a_1b_2} \\ &+ 8Ret_{r_1}^{a_1}S_{a_2}^{b_1}S_{b_2}^{a_2}\tilde{\nu}_{a_1b_1}^{r_1b_2} + 8Ret_{r_1}^{a_1}S_{a_2}^{b_2}S_{b_2}^{r_1}\tilde{\nu}_{a_1b_1}^{a_2b_1} + 8Ret_{r_1}^{a_2}S_{a_2}^{b_2}S_{b_2}^{a_1}\tilde{\nu}_{a_1b_1}^{r_1b_1} \\ &- 4Ret_{r_1}^{a_1}S_{a_2}^{b_1}S_{b_2}^{r_1}\tilde{\nu}_{a_1b_1}^{a_2b_2} - 4Ret_{r_1}^{a_2}S_{a_2}^{b_1}S_{b_2}^{a_1}\tilde{\nu}_{a_1b_1}^{r_1b_2}. \end{aligned} \tag{112}$$

This corrections can be rewritten with the $\tilde{V}_{r_1}^{a_1}$ exchange-induction integrals to simplify the expression (remembering that certain terms contributing to $\tilde{V}_{r_1}^{a_1}$ must be scaled differently in this case).

$$\begin{aligned} E_{exch}^{(120)} K_2^f = &- 2t_{a_1}^{r_1} \tilde{V}_{r_1}^{a_1} - 2t_{a_1}^{r_1} A_{r_1 b_1}^P B_{a_1 b_1}^P - 4t_{a_2}^{r_1} S_{r_1}^{b_1} \tilde{a}^P B_{a_2 b_1}^P \\ &+ 2t_{a_2}^{r_1} S_{r_1}^{b_1} A_{a_1 a_2}^P B_{a_1 b_1}^P + 2t_{a_1}^{r_1} S_{a_2}^{b_1} A_{a_2 r_1}^P B_{a_1 b_2}^P \\ &- 4t_{a_1}^{r_1} S_{b_2}^{a_1} A_{r_1 b_2}^P \tilde{b}^P + 2t_{a_1}^{r_1} S_{b_2}^{a_1} A_{r_1 b_1}^P B_{b_1 b_2}^P \end{aligned} \tag{113}$$

The $E_{exch}^{(12)} K_2^f$ term can be evaluated more efficiently than most of the other correlation corrections; the use of generalized DF integrals makes this term especially easy to evaluate.

2.3.7 $E_{exch}^{(12)} K_{11}^u$

The most involved of the second-order intramonomer correlation corrections to exchange is the $E_{exch}^{(12)} K_{11}^u$ term. Given as:[110]

$$\begin{aligned} E_{exch}^{(120)} K_{11}^u = &-\theta_{a_1 a_2}^{r_1 r_2} t_{r_3 r_2}^{a_1 a_2} \Big(2\tilde{\nu}_{r_1 b_1}^{b_1 r_3} + 4S_{a_3}^{b_1} \tilde{\nu}_{r_1 b_1}^{r_3 a_3} - 2S_{a_3}^{b_1} \tilde{\nu}_{r_1 b_1}^{a_3 r_3} + 4S_{r_1}^{b_1} \tilde{\nu}_{a_3 b_1}^{a_3 r_3} \\ &- 2S_{r_1}^{b_1} \tilde{\nu}_{a_3 b_1}^{r_3 a_3} + 4S_{b_2}^{r_3} \tilde{\nu}_{r_1 b_1}^{b_2 b_1} - 2S_{b_2}^{r_3} \tilde{\nu}_{r_1 b_1}^{b_1 b_2} - 4S_{a_3}^{b_1} S_{b_2}^{a_3} \tilde{\nu}_{r_1 b_1}^{r_3 b_2} \\ &- 4S_{r_1}^{b_1} S_{b_2}^{r_3} \tilde{\nu}_{a_3 b_1}^{a_3 b_2} - 8S_{r_1}^{b_1} S_{b_1}^{a_3} \tilde{\nu}_{a_3 b_2}^{r_3 b_2} + 4S_{r_1}^{b_1} S_{b_2}^{a_3} \tilde{\nu}_{a_3 b_1}^{r_3 b_2} \Big) \\ &+ \theta_{a_1 a_2}^{r_1 r_2} t_{r_1 r_2}^{a_3 a_2} \Big(2\tilde{\nu}_{a_3 b_1}^{b_1 a_1} + 4S_{a_4}^{b_1} \tilde{\nu}_{a_3 b_1}^{a_1 a_4} - 2S_{a_4}^{b_1} \tilde{\nu}_{a_3 b_1}^{a_4 a_1} + 4S_{a_3}^{b_1} \tilde{\nu}_{a_4 b_1}^{a_4 a_1} \\ &- 2S_{a_3}^{b_1} \tilde{\nu}_{a_4 b_1}^{a_1 a_4} + 4S_{b_2}^{a_1} \tilde{\nu}_{a_3 b_1}^{b_2 b_1} - 2S_{b_2}^{a_1} \tilde{\nu}_{a_3 b_1}^{b_1 b_2} - 4S_{a_4}^{b_1} S_{b_2}^{a_4} \tilde{\nu}_{a_3 b_1}^{a_1 b_2} \\ &- 4S_{a_3}^{b_1} S_{b_2}^{a_1} \tilde{\nu}_{a_4 b_1}^{a_4 b_2} - 8S_{a_3}^{b_1} S_{b_1}^{a_4} \tilde{\nu}_{a_4 b_2}^{a_1 b_2} + 4S_{a_3}^{b_1} S_{b_2}^{a_4} \tilde{\nu}_{a_4 b_1}^{a_1 b_2} \Big) \\ &- \theta_{a_1 a_2}^{r_1 r_2} t_{r_3 r_4}^{a_1 a_2} \Big(2S_{r_2}^{b_1} \tilde{\nu}_{r_1 b_1}^{r_3 r_4} + 4S_{r_2}^{b_2} S_{b_2}^{r_4} \tilde{\nu}_{r_1 b_1}^{r_3 b_1} - 2S_{r_2}^{b_1} S_{b_2}^{r_4} \tilde{\nu}_{r_1 b_1}^{r_3 b_2} \Big) \\ &- \theta_{a_1 a_2}^{r_1 r_2} t_{r_1 r_2}^{a_3 a_4} \Big(2S_{a_4}^{b_1} \tilde{\nu}_{a_3 b_1}^{a_1 a_2} + 4S_{a_4}^{b_2} S_{b_2}^{a_2} \tilde{\nu}_{a_3 b_1}^{a_1 b_1} - 2S_{a_4}^{b_1} S_{b_2}^{a_2} \tilde{\nu}_{a_3 b_1}^{a_1 b_2} \Big) \\ &- 2\theta_{a_1 a_3}^{r_1 r_3} \theta_{r_2 r_1}^{a_2 a_1} \Big(S_{a_2}^{b_1} \tilde{\nu}_{r_3 b_1}^{a_3 r_2} + S_{r_3}^{b_1} \tilde{\nu}_{a_2 b_1}^{r_2 a_3} - 2S_{r_3}^{b_1} S_{b_2}^{a_3} \tilde{\nu}_{a_2 b_1}^{r_2 b_2} + 4S_{r_3}^{b_2} S_{b_2}^{a_3} \tilde{\nu}_{a_2 b_1}^{r_2 b_1} \Big) \\ &- \Big(3t_{r_1 r_2}^{a_1 a_2} t_{a_1 a_3}^{r_1 r_3} + \theta_{r_1 r_2}^{a_2 a_1} \theta_{a_3 a_1}^{r_1 r_3} \Big) \Big(-S_{a_2}^{b_1} \tilde{\nu}_{r_3 b_1}^{r_2 a_3} - S_{r_3}^{b_1} \tilde{\nu}_{a_2 b_1}^{a_3 r_2} \\ &- 2S_{r_3}^{b_2} S_{b_2}^{r_2} \tilde{\nu}_{a_2 b_1}^{a_3 b_1} - 2S_{a_2}^{b_2} S_{b_2}^{a_3} \tilde{\nu}_{r_3 b_1}^{r_2 b_1} + S_{a_2}^{b_1} S_{b_2}^{a_3} \tilde{\nu}_{r_3 b_1}^{r_2 b_2} + S_{r_3}^{b_1} S_{b_2}^{r_2} \tilde{\nu}_{a_2 b_1}^{a_3 b_2} \Big), \end{aligned} \tag{114}$$

this term contains some particularly expensive contractions if they are not optimally implemented. Note that this formula contains errors as given in Ref. 110. First, we will consider the first two groups of terms that appear in this correction. If the θ and t amplitudes are

contracted directly, the result ($P^{r_3}_{r_1}$ and $P^{a_3}_{a_1}$, respectively) is identical to the two P quantities that appear in the $E^{(12)}_{elst,r}$ and $E^{(13)}_{elst,r}$ corrections. The introduction of these intermediates allows these two groups of terms to be easily evaluated after generalized DF integrals are introduced. The third group of terms is the most expensive, therefore, we will address it in detail.

$$E^{(120)}_{exch} K^u_{11} \leftarrow -\theta^{r_1 r_2}_{a_1 a_2} t^{a_1 a_2}_{r_3 r_4} \left(2 S^{b_1}_{r_2} \tilde{\nu}^{r_3 r_4}_{r_1 b_1} + 4 S^{b_2}_{r_2} S^{r_4}_{b_2} \tilde{\nu}^{r_3 b_1}_{r_1 b_1} - 2 S^{b_1}_{r_2} S^{r_4}_{b_2} \tilde{\nu}^{r_3 b_2}_{r_1 b_1} \right) \tag{115}$$

Direct contraction of the θ and t amplitudes, in this case, would scale as $\mathcal{O}(o^2 v^4)$ and, therefore, should be avoided. However, if the amplitudes are first contracted with the overlap integrals, the resultants can be directly contracted with, at worst, $\mathcal{O}(o^3 v^3)$ cost. In the case of the second term, the two-electron integrals can be internally contracted to a two-index quantity, which can also be contracted with the amplitudes.

$$\begin{aligned}
U^{r_1 b_1}_{a_1 a_2} &= \theta^{r_1 r_2}_{a_1 a_2} S^{b_1}_{r_2} \\
W^{r_1 b_1}_{r_3 r_4} &= U^{r_1 b_1}_{a_1 a_2} t^{a_1 a_2}_{r_3 r_4} \\
X^{r_1 r_2}_{a_1 a_2} &= t^{a_1 a_2}_{r_3 r_4} S^{b_2}_{r_2} S^{r_4}_{b_2} A^P_{r_1 r_3} \tilde{b}^P \\
Y^{a_1 a_2}_{r_3 b_2} &= t^{a_1 a_2}_{r_3 r_4} S^{b_2}_{r_4} \\
Z^{r_1 b_1}_{r_3 b_2} &= U^{r_1 b_1}_{a_1 a_2} Y^{a_1 a_2}_{r_3 b_2}
\end{aligned} \tag{116}$$

With these new intermediates, the third group of terms in $E^{(120)}_{exch} K^u_{11}$ can be evaluated with $\mathcal{O}(o^3 v^3)$ effort.

$$E^{(120)}_{exch} K^u_{11} \leftarrow -2 W^{r_1 b_1}_{r_3 r_4} \tilde{\nu}^{r_3 r_4}_{r_1 b_1} - 4 X^{r_1 r_2}_{a_1 a_2} \theta^{r_1 r_2}_{a_1 a_2} + 2 Z^{r_1 b_1}_{r_3 b_2} \tilde{\nu}^{r_3 b_2}_{r_1 b_1} \tag{117}$$

The fourth group of terms is much easier to handle; the amplitudes can be directly contracted, removing all the virtual indices. This contraction scales as $\mathcal{O}(o^4 v^2)$ and the resultant is an *oooo* quantity. The fourth and fifth terms are evaluated similarly, the amplitudes are directly contracted, scaling $\mathcal{O}(o^3 v^3)$, and the resultant is contracted with the overlap integrals and two-electron integrals to evaluate the contribution to $E^{(120)}_{exch} K^u_{11}$.

2.3.8 $^t E^{(22)}_{ind}$

The equation for the second-order correction to induction is taken from reference 212, however, the formula presented in that work contains errors. This correction has two parts,

${}^{t}E^{(220)}_{ind}(A \leftarrow B)$ and ${}^{t}E^{(220)}_{ind}(A \rightarrow B)$. Here, $(A \leftarrow B)$ refers to the interaction of monomer A with the electric field of monomer B and $(A \rightarrow B)$ refers to the interaction of monomer B with the electric field of monomer A. Both of these corrections can be evaluated more efficiently through the introduction of DF intermediates. The ${}^{t}E^{(220)}_{ind}(A \leftarrow B)$ correction is given by

$$\begin{aligned}
{}^{t}E^{(220)}_{ind}(A \leftarrow B) = & (2X^{r_1r_2}_{a_1a_2} - X^{r_2r_1}_{a_1a_2})X^{a_1a_2}_{r_1r_2}\Delta^{r_1r_2}_{a_1a_2} + 4t^{r_1}_{a_1}z^{a_1}_{r_1} \\
& - 2\theta^{a_1a_2}_{r_1r_2}I^{a_3}_{r_3}((\omega_B)^{r_2}_{a_3}t^{r_3r_1}_{a_2a_1} + (\omega_B)^{r_3}_{a_2}t^{r_2r_1}_{a_3a_1}) \\
& - 2\theta^{r_1r_2}_{a_1a_2}I^{r_3}_{a_3}(I^{a_2}_{r_3}\nu^{a_3a_1}_{r_2r_1} + I^{a_3}_{r_2}\nu^{a_2a_1}_{r_3r_1}) \\
& + 2I^{r_1}_{a_1}I^{r_2}_{a_2}\overset{(2)}{\theta}{}^{a_1a_2}_{r_1r_2}/\Delta^{r_1r_2}_{a_1a_2} - 4I^{r_1}_{a_1}I^{r_2}_{a_2}(2\nu^{a_1r_3}_{r_1a_3} - \nu^{r_3a_1}_{r_1a_3})\theta^{a_3a_2}_{r_3r_2},
\end{aligned} \tag{118}$$

where

$$I^{a_1}_{r_1} = (\omega_B)^{a_1}_{r_1}\Delta^{r_1}_{a_1} \tag{119}$$

and

$$z^{a_1}_{r_1} = (\omega_B)^{r_2}_{r_1}I^{a_1}_{r_2} - (\omega_B)^{a_1}_{a_2}I^{a_2}_{r_1}. \tag{120}$$

The $X^{a_1a_2}_{r_1r_2}$ quantity is given as

$$\begin{aligned}
X^{a_1a_2}_{r_1r_2} = & (\omega_B)^{r_3}_{r_1}t^{a_1a_2}_{r_3r_2} + (\omega_B)^{r_3}_{r_2}t^{a_1a_2}_{r_1r_3} - (\omega_B)^{a_1}_{a_3}t^{a_3a_2}_{r_1r_2} - (\omega_B)^{a_2}_{a_3}t^{a_1a_3}_{r_1r_2} \\
& \nu^{a_1r_3}_{r_1r_2}I^{a_2}_{r_3} + \nu^{r_3a_2}_{r_1r_2}I^{a_1}_{r_3} - \nu^{a_1a_2}_{r_1a_3}I^{a_3}_{r_2} - \nu^{a_1a_2}_{a_3r_2}I^{a_3}_{r_1}.
\end{aligned} \tag{121}$$

DF intermediates can only be used in the first, fourth and sixth terms of ${}^{t}E^{(220)}_{ind}(A \leftarrow B)$. First, we will examine the formation of $X^{a_1a_2}_{r_1r_2}$. The first four terms of $X^{a_1a_2}_{r_1r_2}$ cannot use DF intermediates. Including the contributions from the last two terms conventionally scales as $\mathcal{O}(o^3v^2)$. Using DF intermediates would lead to a scaling of $\mathcal{O}(o^2v^2n_{ri})$, therefore we will not use DF intermediates to evaluate this term. The fifth and sixth terms depend on the ov^3 integrals and can be evaluated by defining an intermediate as

$$D^{P}_{a_1r_1} = C^{P}_{r_3r_1}I^{a_1}_{r_3}. \tag{122}$$

This contribution to $X^{a_1a_2}_{r_1r_2}$ can be evaluated by forming another quantity,

$$Y^{a_1a_2}_{r_1r_2} = D^{P}_{a_1r_1}C^{P}_{a_2r_2}, \tag{123}$$

and adding

$$X_{r_1 r_2}^{a_1 a_2} = Y_{r_1 r_2}^{a_1 a_2} + Y_{r_2 r_1}^{a_2 a_1}. \tag{124}$$

To evaluate the fourth ${}^{t}E_{ind}^{(220)}(A \leftarrow B)$ term, the Θ intermediates can be reused along with two new quantities:

$$E_{a_2 r_2}^{P} = I_{a_3}^{r_3} I_{r_3}^{a_2} C_{a_3 r_2}^{P} \tag{125}$$

and

$$F_{a_2 r_2}^{P} = I_{a_3}^{r_3} I_{r_2}^{a_3} C_{a_2 r_3}^{P}. \tag{126}$$

With these quantities, this contribution to ${}^{t}E_{ind}^{(220)}(A \leftarrow B)$ can be evaluated as

$$-2\Theta_{a_2 r_2}^{P}(E_{a_2 r_2}^{P} + F_{a_2 r_2}^{P}) \tag{127}$$

The sixth term contributing to ${}^{t}E_{ind}^{(220)}(A \leftarrow B)$ can be evaluated conventionally as $\mathcal{O}(o^2 v^2)$. The introduction of DF intermediates leads to a scaling of $\mathcal{O}(ov^2 n_{ri})$, so DF intermediates may not be optimal in this case.

A conventional evaluation of ${}^{t}E_{ind}^{(220)}(A \leftarrow B)$ will scale $\mathcal{O}(o^2 v^3)$. Specifically, the first, third and fourth terms exhibit this scaling, while the fifth and sixth terms will scale $\mathcal{O}(o^2 v^2)$. The purpose of introducing DF intermediates into the ${}^{t}E_{ind}^{(220)}(A \leftarrow B)$ evaluation was to remove the need to deal with a type of ov^3 integrals explicitly. The two terms where DF intermediates were introduced scale $\mathcal{O}(o^2 v^2 n_{ri})$.

The ${}^{t}E_{ind}^{(220)}(A \rightarrow B)$ contributions to ${}^{t}E_{ind}^{(220)}$ can be written as

$${}^{t}E_{ind}^{(220)}(A \rightarrow B) = -8\theta_{r_1 r_2}^{a_3 a_2} t_{a_3 a_1}^{r_1 r_2} \nu_{a_2 b_1}^{a_1 s_1} I_{s_1}^{b_1} + 8\theta_{r_3 r_2}^{a_1 a_2} t_{a_1 a_2}^{r_3 r_1} \nu_{r_2 b_1}^{r_1 s_1} I_{s_1}^{b_1} + 16 t_{r_1}^{a_1} \nu_{r_1 b_1}^{a_1 s_1} I_{s_1}^{b_1}. \tag{128}$$

This can be evaluated by substituting some of the intermediates (including the P quantities used in several places) defined earlier:

$$\begin{aligned} {}^{t}E_{ind}^{(220)}(A \rightarrow B) = &- 8(P_{a_2}^{a_1} C_{a_1 a_2}^{P})(C_{b_1 s_1}^{P} I_{s_1}^{b_1}) + 8(P_{r_2}^{r_1} C_{r_1 r_2}^{P})(C_{b_1 s_1}^{P} I_{s_1}^{b_1}) \\ &+ 16(t_{r_1}^{a_1} C_{a_1 r_1}^{P})(C_{b_1 s_1}^{P} I_{s_1}^{b_1}). \end{aligned} \tag{129}$$

The quantities inside of the parenthesis should be fully contracted, then this term can be evaluated as a series of dot products. Conventionally, this term scales $\mathcal{O}(ov^3)$; the

DF implementation scales $\mathcal{O}(v^2 n_{ri})$ and avoids the r^2bs integrals. By introducing DF intermediates into the ${}^tE_{ind}^{(220)}$ evaluation, the computation of two types of ov^3 integrals has been avoided.

2.3.9 $E_{disp}^{(21)}$

The first-order intramonomer correlation correction to dispersion is given as:[110]

$$E_{disp}^{(210)} = 4t_{r_1 s_1}^{a_1 b_1} t_{a_2 b_1}^{r_2 s_1} (2\nu_{a_1 r_2}^{r_1 a_2} - \nu_{a_2 r_2}^{a_1 r_1}) + 8Ret_{r_1 s_1}^{a_1 b_1} \nu_{r_2 s_1}^{a_2 b_1} \theta_{a_1 a_2}^{r_1 r_2}. \tag{130}$$

Canonically, this term involves two $\mathcal{O}(o^3 v^3)$ contractions. At first glance, it is obvious that DF intermediates can be applied to the second term, but it does not appear that the first term will benefit from the introduction of DF integrals. Indeed, DF integrals alone cannot improve the scaling of the first term, however, if Laplace transformed energy denominators are also introduced, the scaling can be reduced. We will consider two paths to the contraction,

$$Z_{r_1 r_2}^{a_1 a_2} = t_{r_1 s_1}^{a_1 b_1} t_{a_2 b_1}^{r_2 s_1} \tag{131}$$

through the introduction of DF integrals and Laplace transformed energy denominators. First, we will apply Laplace transformed energy denominators to one of the amplitudes and leave the other in its canonical form.

$$W_{a_1 r_1}^{P,L} = t_{r_1 s_1}^{a_1 b_1} C_{b_1 s_1}^{P} d_{b_1 s_1}^{L} \tag{132}$$

$$Z_{r_1 r_2}^{a_1 a_2} = W_{a_1 r_1}^{P,L} C_{a_2 r_2}^{P} d_{a_2 r_2}^{L} \tag{133}$$

Here, there are two steps, each of which scale $\mathcal{O}(o^2 v^2 N_{df}) N_\delta$ This will out perform a canonical implementation for realistic systems (*i.e.* $o >> 1$). Next, we will apply Laplace transforms and DF integrals to both of the amplitudes.

$$X_{PQ}^{LM} = C_{b_1 s_1}^{P} C_{b_1 s_1}^{Q} d_{b_1 s_1}^{L} d_{b_1 s_1}^{M} \tag{134}$$

$$Y_{a_1 r_1}^{P,M} = C_{a_1 r_1}^{Q} X_{PQ}^{LM} d_{a_1 r_1}^{L} \tag{135}$$

$$Z_{r_1 r_2}^{a_1 a_2} = Y_{a_1 r_1}^{P,M} C_{a_2 r_2}^{P} d_{a_2 r_2}^{M} \tag{136}$$

Here there are two steps that scale as $\mathcal{O}(ovN_{df}^2) N_\delta^2$ and a final step that scales $\mathcal{O}(o^2 v^2 N_{df}) N_\delta$. It is uncertain which of these factorizations will be faster in general. In theory, the second

factorization can be no more than 2 times faster than the first factorization, however, it also could be infinitely slower. The factorization of the $t^{a_1 b_1}_{r_1 s_1} t^{r_2 s_1}_{a_2 b_1}$ contraction was shown for illustrative purposes. In practice, it is better to contract one of the dispersion amplitudes with the antisymmetrized two-electron integrals. This contraction also appears in the $E^{(211)}_{disp}$ and $E^{(22)}_{disp}$(D) corrections, therefore, it can be reused. As such, the final form for the $E^{(21)}_{disp}$ correction is:

$$E^{(210)}_{disp} = 4\,{}^{A}G^{a_1 b_1}_{r_1 s_1} t^{r_1 s_1}_{a_1 b_1} + 8ReT^{P}_{a_1 r_1} \Theta^{P}_{a_1 r_1}. \tag{137}$$

2.3.10 $E^{(211)}_{disp}$

Much like for the exchange corrections, the second-order intramonomer correlations corrections to dispersion contain a "bilinear term." Given as:[110]

$$E^{(211)}_{disp} = 8ReX^{a_1 b_1}_{r_1 s_1} Y^{r_1 s_1}_{a_1 b_1} \Delta^{a_1 b_1}_{r_1 s_1} + 8Re\theta^{a_1 a_2}_{r_1 r_2} \theta^{b_1 b_2}_{s_1 s_2} \nu^{r_2 s_1}_{a_2 b_1} t^{a_1 b_2}_{r_1 s_2}, \tag{138}$$

where

$$X^{a_1 b_1}_{r_1 s_1} = \theta^{b_1 b_2}_{s_1 s_2} \nu^{a_1 s_2}_{r_1 b_2} + t^{a_1 b_2}_{r_1 s_2} (2\nu^{b_1 b_2}_{s_1 s_2} - \nu^{b_1 s_1}_{b_2 s_2}) \tag{139}$$

$$Y^{a_1 b_1}_{r_1 s_1} = \theta^{a_1 a_2}_{r_1 r_2} \nu^{b_1 r_2}_{s_1 a_2} + t^{a_2 b_1}_{r_2 s_1} (2\nu^{a_1 a_2}_{r_1 r_2} - \nu^{a_1 r_1}_{a_2 r_2}), \tag{140}$$

canonically, this correction scales $\mathcal{O}(o^3 v^3)$. The second term in $E^{(211)}_{disp}$ is immediately amenable to the introduction of DF intermediates; the scaling of this term is reduced to $\mathcal{O}(o^2 v^2 N_{df})$. Likewise, the first term that appears in the X and Y quantities can benefit from the introduction of DF intermediates. The scaling of the second term that appears in these quantities can also be reduced if Laplace transformed energy denominators are applied. With the ${}^{A}G$ and ${}^{B}G$ intermediates, the X and Y quantities can be rewritten in a manner that illustrates a more efficient implementation.

$$X^{a_1 b_1}_{r_1 s_1} = C^{P}_{a_1 r_1} \Theta^{P}_{b_1 s_1} + {}^{B}G^{a_1 b_1}_{r_1 s_1} \tag{141}$$

$$Y^{a_1 b_1}_{r_1 s_1} = \Theta^{P}_{a_1 r_1} C^{P}_{b_1 s_1} + {}^{A}G^{a_1 b_1}_{r_1 s_1} \tag{142}$$

This allows the $E^{(211)}_{disp}$ correction to be written in a form that will scale $\mathcal{O}(N^5)$ instead of $\mathcal{O}(N^6)$.

$$E^{(211)}_{disp} = 8ReX^{a_1 b_1}_{r_1 s_1} Y^{r_1 s_1}_{a_1 b_1} \Delta^{a_1 b_1}_{r_1 s_1} + 8Re\Theta^{P}_{a_1 r_1} \Theta^{P}_{b_1 s_1} t^{a_1 b_1}_{r_1 s_1} \tag{143}$$

2.3.11 $E_{disp}^{(22)}$(S)

The singles contribution to $E_{disp}^{(22)}$ has a particularly simple form:[110]

$$E_{disp}^{(22)}(\mathrm{S}) = 8Re t_{r_1}^{a_1} X_{a_1}^{r_1} \Delta_{r_1}^{a_1}, \tag{144}$$

where

$$X_{a_1}^{r_1} = \nu_{r_2 s_1}^{r_1 b_1} t_{a_1 b_1}^{r_2 s_1} - \nu_{a_1 s_1}^{a_2 b_1} t_{a_2 b_1}^{r_1 s_1}. \tag{145}$$

The X quantity can be rewritten entirely in terms of three-index quantities.

$$X_{a_1}^{r_1} = T_{a_1 r_2}^{P} C_{r_1 r_2}^{P} - T_{a_2 r_1}^{P} C_{a_1 a_2}^{P} \tag{146}$$

This does not lower the scaling of the $E_{disp}^{(22)}$(S) correction, since, canonically it is already $\mathcal{O}(N^5)$. The advantage of the DF factorization is that the *ovvv* integrals do not need to be formed.

2.3.12 $E_{disp}^{(22)}$(D)

The doubles contribution to $E_{disp}^{(22)}$ is given as:[110]

$$E_{disp}^{(22)}(\mathrm{D}) = 4Re(2X_{r_1 r_2}^{a_1 a_2} - X_{r_2 r_1}^{a_1 a_2}) \overset{(2)}{t}{}_{a_1 a_2}^{r_1 r_2} + 4Y_{a_1 b_1}^{r_1 s_1} Y_{r_1 s_1}^{a_1 b_1} \Delta_{a_1 b_1}^{r_1 s_1}, \tag{147}$$

where

$$X_{r_1 r_2}^{a_1 a_2} = \nu_{r_1 b_1}^{a_1 s_1} t_{r_2 s_1}^{a_2 b_1} + \nu_{r_2 b_1}^{a_2 s_1} t_{r_1 s_1}^{a_1 b_1} \tag{148}$$

$$Y_{r_1 s_1}^{a_1 b_1} = \nu_{r_2 b_1}^{a_2 s_1} \theta_{r_1 r_2}^{a_1 a_2} + (2\nu_{r_1 a_2}^{a_1 r_2} - \nu_{a_2 r_2}^{a_1 r_1}) t_{r_2 s_1}^{a_2 b_1} \tag{149}$$

Note that the definition of U in Ref. 110 is equivalent to the $\overset{(2)}{t}$ amplitudes; the different multiplicative factors are a result of a slightly different definition for θ. The X quantity can be formed with existing DF intermediates. Additionally, the first term contributing to Y can use existing DF intermediates. The second term can be factored in $\mathcal{O}(N^5)$ if Laplace transformed energy denominators are applied. This gives a more convenient form for the X and Y quantities.

$$X_{r_1 r_2}^{a_1 a_2} = C_{a_1 r_1}^{P} T_{a_2 r_2}^{P} + T_{a_1 r_1}^{P} C_{a_2 r_2}^{P} \tag{150}$$

$$Y_{r_1 s_1}^{a_1 b_1} = \Theta_{a_1 r_1}^{P} C_{b_1 s_1}^{P} + {}^{A}G_{r_1 s_1}^{a_1 b_1} \tag{151}$$

The quantities required to evaluate the $E^{(22)}_{disp}$(D) correction can now be formed with $\mathcal{O}(N^5)$ effort. This neglects, obviously, the formation of the second-order doubles amplitudes.

2.3.13 $E^{(22)}_{disp}$(T)

The most computationally intensive intramonomer dispersion correction is the triples contribution to $E^{(22)}_{disp}$. This term is essential to obtain accurate dispersion energies from wavefunction-based SAPT, unfortunately, the canonical scaling of this correction is $\mathcal{O}(o^3v^4)$. This correction is given as[110]

$$E^{(220)}_{disp}(\mathrm{T}) = (4W^{a_1a_2b_1}_{r_1r_2s_1} - 2W^{a_2a_1b_1}_{r_1r_2s_1})W^{r_1r_2s_1}_{a_1a_2b_1}\Delta^{r_1r_2s_1}_{a_1a_2b_1}, \tag{152}$$

where the triples amplitudes are defined as

$$\begin{aligned} W^{a_1a_2b_1}_{r_1r_2s_1} = {} & \nu^{r_3b_1}_{r_2s_1}t^{a_1a_2}_{r_1r_3} + \nu^{r_3b_1}_{r_1s_1}t^{a_1a_2}_{r_3r_2} - \nu^{a_2b_1}_{a_3s_1}t^{a_1a_3}_{r_1r_2} - \nu^{a_1b_1}_{a_3s_1}t^{a_3a_2}_{r_1r_2} \\ & + \nu^{r_3a_2}_{r_1r_2}t^{a_1b_1}_{r_3s_1} + \nu^{a_1r_3}_{r_1r_2}t^{a_2b_1}_{r_3s_1} - \nu^{a_1a_2}_{a_3r_2}t^{a_3b_1}_{r_1s_1} - \nu^{a_1a_2}_{r_1a_3}t^{a_3b_1}_{r_2s_1}. \end{aligned} \tag{153}$$

In practice, the four symmetry unique terms that appear in Equation 153 are evaluated and the resulting amplitudes are symmetrized. The energy contribution from these amplitudes is then evaluated with Equation 152. Since the triples amplitudes must be formed in blocks, existing implementations of this term use one of two possible loop structures to evaluate $E^{(22)}_{disp}$ using Equations 153 and 152. Blocking over a_1, a_2, and b_1, which is how the term is implemented in SAPT2008,[25] leads to an algorithm with $\mathcal{O}(o^3v^3)$ disk I/O. Alternatively, the loop can be blocked over b_1 and s_1; this is used in the parallel implementation of SAPT[26] and in the present context. While canonically inferior, due to $\mathcal{O}(o^2v^4)$ disk I/O, this loop structure is more amenable to the introduction of DF integrals.

The large amount of disk I/O required to evaluate this correction is a result of the ov^3 integrals, which cannot be held in memory for most computations. With the b_1s_1 loop structure, all of the $(ar|rr)$ integrals are required for each b_1 and s_1. To avoid this bottleneck, the DF representation of these integrals is introduced.

$$D^P_{a_1r_1} = C^P_{r_3r_1}t^{a_1b_1}_{r_3s_1} \tag{154}$$

$$\nu^{r_3a_2}_{r_1r_2}t^{a_1b_1}_{r_3s_1} \approx D^P_{a_1r_1}C^P_{a_2r_2} \tag{155}$$

The refactored equations now scale slightly worse at $\mathcal{O}(o^3v^3N_{df})$, however, the redundant disk I/O has been removed. The memory requirement for this algorithm is modest considering the $\mathcal{O}(N^7)$ scaling of the term. It is worthy of note that this correction could be evaluated by inserting Equations 153 into Equation 152 and expanding the 128 possible terms, many of which will be equivalent by symmetry. This approach has not been implemented, but would likely result in an algorithm scaling $\mathcal{O}(N^6)$ if Laplace transformed energy denominators were applied.

Algorithm 7 Pseudocode to evaluate the $E_{disp}^{(220)}(\mathrm{T})$ correction with DF integrals.

```
Read/Form t^{a1a2}_{r1r2}
Read/Form ν^{a1a2}_{a3r1}
Read C^P_{a1r1}
Read C^P_{r1r2}
for b1 < #OccupiedB do
  for s1 < #VirtualB do
    Read/Form t^{a1b1}_{r1s1}
    Read/Form ν^{r1b1}_{r2s1}
    Read/Form ν^{a1b1}_{a2s1}
    V^{a1a2}_{r1r2} ← ν^{r3b1}_{r2s1} t^{a1a2}_{r1r3}
    V^{a1a2}_{r1r2} ← ν^{a2b1}_{a3s1} t^{a1a3}_{r1r2}
    V^{a1a2}_{r1r2} ← ν^{a1a2}_{a3r2} t^{a3b1}_{r1s1}
    D^P_{a1r1} ← C^P_{r3r1} t^{a1b1}_{r3s1}
    V^{a1a2}_{r1r2} ← D^P_{a1r1} C^P_{a2r2}
    E^{(220)}_{disp}(T ← (4W^{a1a2}_{r1r2} − 2W^{a2a1}_{r1r2}) W^{r1r2}_{a1a2} Δ^{r1r2s1}_{a1a2b1}
  end for
end for
```

2.3.14 $E_{disp}^{(22)}(\mathrm{Q})$

The final second-order intramonomer correlation correction to dispersion is the quadruples contribution to $E_{disp}^{(22)}$ given as:[110]

$$E_{disp}^{(22)}(\mathrm{Q}) = 4\Big(\nu^{r_3s_1}_{a_3b_1}\theta^{r_1r_2}_{a_1a_2}\theta^{a_2a_3}_{r_2r_3}t^{a_1b_1}_{r_1s_1} - t^{a_1a_2}_{r_1r_3}\theta^{r_1r_2}_{a_1a_2}\nu^{a_3b_1}_{r_2s_1}t^{a_3b_1}_{r_3s_1} - t^{a_1a_3}_{r_1r_2}\theta^{r_1r_2}_{a_1a_2}\nu^{a_2b_1}_{r_3s_1}t^{a_3b_1}_{r_3s_1}$$
$$+t^{r_3s_1}_{a_3b_1}g^{r_1r_2}_{a_1a_2}\theta^{a_2a_3}_{r_2r_3}t^{a_1b_1}_{r_1s_1} - t^{r_1r_3}_{a_1a_2}g^{a_1a_2}_{r_1r_2}t^{a_3b_1}_{r_3s_1}t^{r_2s_1}_{a_3b_1} - t^{r_1r_2}_{a_1a_3}g^{a_1a_2}_{r_1r_2}t^{a_3b_1}_{r_3s_1}t^{r_3s_1}_{a_2b_1}\Big) \qquad (156)$$

The first term can be easily factored with DF intermediates that have been previously defined. The second and third terms can be efficiently evaluated canonically if the P type

intermediates are introduced as well as:

$$Q_{r_2}^{r_3} = C_{a_3 r_2}^{P} T_{a_3 r_3}^{P} \tag{157}$$

$$Q_{a_2}^{a_3} = C_{a_2 r_3}^{P} T_{a_3 r_3}^{P}. \tag{158}$$

The fourth term requires the introduction of DF integrals and Laplace transformed energy denominators in order to achieve the optimal factorization.

$$W_{a_2 r_2}^{P,L} = g_{a_1 a_2}^{r_1 r_2} C_{a_1 r_1}^{P} d_{a_1 r_1}^{L} \tag{159}$$

$$X_{r_2 s_1}^{a_2 b_1} = W_{a_2 r_2}^{P,L} C_{b_1 s_1}^{P} d_{b_1 s_1}^{L} \tag{160}$$

$$Y_{a_2 r_2}^{P,L} = X_{r_2 s_1}^{a_2 b_1} C_{b_1 s_1}^{P} d_{b_1 s_1}^{L} \tag{161}$$

$$Z_{r_2 r_3}^{a_2 a_3} = Y_{a_2 r_2}^{P,L} C_{a_3 r_3}^{P} d_{a_3 r_3}^{L} \tag{162}$$

The fifth and sixth terms can be evaluated canonically after performing a few familiar contractions.

$$R_{r_2}^{r_3} = t_{a_1 a_2}^{r_1 r_3} g_{r_1 r_2}^{a_1 a_2} \tag{163}$$

$$R_{a_2}^{a_3} = t_{a_1 a_3}^{r_1 r_2} g_{r_1 r_2}^{a_1 a_2} \tag{164}$$

$$S_{r_2}^{r_3} = t_{r_3 s_1}^{a_3 b_1} t_{a_3 b_1}^{r_2 s_1} \tag{165}$$

$$S_{a_2}^{a_3} = t_{r_3 s_1}^{a_3 b_1} t_{a_2 b_1}^{r_3 s_1} \tag{166}$$

Now the quadruples contribution to $E_{disp}^{(22)}$ can be written in a much more compact form:

$$\begin{aligned} E_{disp}^{(22)}(\mathrm{Q}) = 4\Big(& T_{a_1 r_1}^{P} \Theta_{a_2 r_2}^{P} \theta_{a_1 a_2}^{r_1 r_2} - P_{r_2}^{r_1} Q_{r_1}^{r_2} - P_{a_2}^{a_1} Q_{a_1}^{a_2} \\ & + Z_{r_1 r_2}^{a_1 a_2} \theta_{r_1 r_2}^{a_1 a_2} - R_{r_2}^{r_1} S_{r_1}^{r_2} - R_{a_2}^{a_1} S_{a_1}^{a_2} \Big) \end{aligned} \tag{167}$$

This factorization reduces the scaling of $E_{disp}^{(22)}(\mathrm{Q})$ from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$.

2.4 Third-order Corrections

2.4.1 $E_{ind}^{(30)}$

The third-order induction energy begins to included induced multipole-induced multipole interactions. The form of this correction is given by:[162]

$$\begin{aligned} E_{ind}^{(30)} = & 2 {}^{(0)}s_{a_1}^{r_1} (\omega_B)_{r_1}^{r_2} {}^{(0)}s_{r_2}^{a_1} - 2 {}^{(0)}s_{a_1}^{r_1} (\omega_B)_{a_2}^{a_1} {}^{(0)}s_{r_2}^{a_1} + 2 {}^{(0)}s_{b_1}^{s_1} (\omega_B)_{s_1}^{s_2} {}^{(0)}s_{s_2}^{b_1} \\ & - 2 {}^{(0)}s_{b_1}^{s_1} (\omega_B)_{b_2}^{b_1} {}^{(0)}s_{s_2}^{b_1} + 16 {}^{(0)}s_{a_1}^{r_1} \nu_{r_1 s_1}^{a_1 b_1} {}^{(0)}s_{b_1}^{s_1}. \end{aligned} \tag{168}$$

In practice, the $E_{ind}^{(30)}$ should not be included in the interaction energy without it exchange counterpart, $E_{exch-ind}^{(30)}$. The $E_{exch-ind}^{(30)}$ correction is written, in part, in terms of third-order induction amplitudes. Therefore, it is useful to rewrite $E_{ind}^{(30)}$ in terms of third-order induction amplitudes, since they will be needed elsewhere.

$$(u_{ind})_{r_1}^{a_1} = \Big[(\omega_B)_{r_1}^{r_2} {}^{(0)}s_{r_2}^{a_1} - (\omega_B)_{a_2}^{a_1} {}^{(0)}s_{r_1}^{a_2} + 2\nu_{r_1 s_1}^{a_1 b_1} {}^{(0)}s_{s_1}^{b_1} + 2(\omega_A)_{b_1}^{s_1} t_{r_1 s_1}^{a_1 b_1}\Big]\Delta_{a_1}^{r_1} \tag{169}$$

$$(u_{ind})_{s_1}^{b_1} = \Big[(\omega_A)_{s_1}^{s_2} {}^{(0)}s_{s_2}^{b_1} - (\omega_A)_{b_2}^{b_1} {}^{(0)}s_{s_1}^{b_2} + 2\nu_{r_1 s_1}^{a_1 b_1} {}^{(0)}s_{r_1}^{a_1} + 2(\omega_B)_{a_1}^{r_1} t_{r_1 s_1}^{a_1 b_1}\Big]\Delta_{b_1}^{s_1} \tag{170}$$

This allows the energy to be written in terms of induction amplitudes and ω integrals.

$$E_{ind}^{(30)} = 2(\omega_B)_{a_1}^{r_1}(u_{ind})_{r_1}^{a_1} + 2(\omega_A)_{b_1}^{s_1}(u_{ind})_{s_1}^{b_1} \tag{171}$$

Since this correction can be evaluated very easily, no further elaboration is required.

2.4.2 $E_{exch-ind}^{(30)}$

The third-order exchange-induction is extremely important if third-order induction effects are to be included. The induction series tends to diverge; exchange effects are required to remove this divergence. The $E_{exch-ind}^{(30)}$ term contains five separate contributions; conveniently, three of these contributions can be written in terms of the $\tilde{V}$ collection of exchange integrals that was introduced earlier. Using these definitions, the form of the $E_{exch-ind}^{(30)}$ correction is given by:[162]

$$\begin{aligned}
E_{exch-ind}^{(30)} = -2\Big[&(u_{ind})_{r_1}^{a_1}\tilde{V}_{a_1}^{r_1} + (u_{ind})_{s_1}^{b_1}\tilde{V}_{b_1}^{s_1} \\
&+ {}^{(0)}s_{r_1}^{a_1} {}^{(0)}s_{s_1}^{b_1}\Big(\tilde{V}_{a_1 b_1}^{r_1 s_1} - 4S_{a_2}^{s_1}S_{b_1}^{a_2}\tilde{\nu}_{a_1 b_2}^{r_1 b_2} - 4S_{a_1}^{b_2}S_{b_2}^{r_1}\tilde{\nu}_{a_2 b_1}^{a_2 s_1}\Big)\Big] \\
&+2\,{}^{(0)}s_{r_1}^{a_1} {}^{(0)}s_{r_2}^{a_2}\Big(S_{a_2}^{b_1}\tilde{\nu}_{a_1 b_1}^{r_2 r_1} - 2S_{a_1}^{b_1}\tilde{\nu}_{a_2 b_1}^{r_2 r_1} + 2S_{a_2}^{b_1}S_{b_1}^{r_1}\tilde{\nu}_{a_1 b_2}^{r_2 b_2} \\
&- S_{a_2}^{b_1}S_{b_2}^{r_1}\tilde{\nu}_{a_1 b_1}^{r_2 b_2} + 2S_{a_1}^{b_1}S_{b_2}^{r_1}\tilde{\nu}_{a_2 b_1}^{r_2 b_2}\Big) \\
&+2\,{}^{(0)}s_{s_1}^{b_1} {}^{(0)}s_{s_2}^{b_2}\Big(S_{b_2}^{a_1}\tilde{\nu}_{b_1 a_1}^{s_2 s_1} - 2S_{b_1}^{a_1}\tilde{\nu}_{b_2 a_1}^{s_2 s_1} + 2S_{b_2}^{a_1}S_{a_1}^{s_1}\tilde{\nu}_{b_1 a_2}^{s_2 a_2} \\
&- S_{b_2}^{a_1}S_{a_2}^{s_1}\tilde{\nu}_{b_1 a_1}^{s_2 a_2} + 2S_{b_1}^{a_1}S_{a_2}^{s_1}\tilde{\nu}_{b_2 a_1}^{s_2 a_2}\Big).
\end{aligned} \tag{172}$$

Much like $E_{ind}^{(30)}$, its exchange counterpart is fairly straightforward to implement. The remaining terms can be easily evaluated if the generalized DF integrals are introduced.

2.4.3 $E^{(30)}_{ind-disp}$

The induction-dispersion term includes the coupling between induction and dispersion interactions; *i.e.* how does the induced polarization affect dispersion. The $E^{(30)}_{ind-disp}$ term is given as:[162]

$$\begin{aligned} E^{(30)}_{ind-disp} =& 8\,{}^{(0)}s^{r_1}_{a_1}\nu^{r_2 s_1}_{r_1 b_1}t^{a_1 b_1}_{r_2 s_1} - 8\,{}^{(0)}s^{r_1}_{a_1}\nu^{a_1 s_1}_{a_2 b_1}t^{a_2 b_1}_{r_1 s_1} \\ &+8\,{}^{(0)}s^{s_1}_{b_1}\nu^{r_1 s_2}_{a_1 s_1}t^{a_1 b_1}_{r_1 s_2} - 8\,{}^{(0)}s^{s_1}_{b_1}\nu^{r_1 b_1}_{a_1 b_2}t^{a_1 b_2}_{r_1 s_1} \\ &+4t^{r_1 s_1}_{a_1 b_1}(\omega_B)^{r_2}_{r_1}t^{a_1 b_1}_{r_2 s_1} - 4t^{r_1 s_1}_{a_1 b_1}(\omega_B)^{a_1}_{a_2}t^{a_2 b_1}_{r_1 s_1} \\ &+4t^{r_1 s_1}_{a_1 b_1}(\omega_A)^{s_2}_{s_1}t^{a_1 b_1}_{r_1 s_2} - 4t^{r_1 s_1}_{a_1 b_1}(\omega_A)^{b_1}_{b_2}t^{a_1 b_2}_{r_1 s_1}. \end{aligned} \tag{173}$$

As was the case for $E^{(30)}_{ind}$, in practice, this correction is evaluated through the formation of induction-dispersion amplitudes so that they can be reused during the evaluation of $E^{(30)}_{exch-ind-disp}$. The relevant amplitudes are given as:

$$(u_{ind-disp})^{a_1}_{r_1} = \left[2\nu^{r_2 s_1}_{r_1 b_1}t^{a_1 b_1}_{r_2 s_1} - 2\nu^{a_1 s_1}_{a_2 b_1}t^{a_2 b_1}_{r_1 s_1}\right]\Delta^{r_1}_{a_1} \tag{174}$$

$$(u_{ind-disp})^{b_1}_{s_1} = \left[2\nu^{r_1 s_2}_{a_1 s_1}t^{a_1 b_1}_{r_1 s_2} - 2\nu^{r_1 b_1}_{a_1 b_2}t^{a_1 b_2}_{r_1 s_1}\right]\Delta^{s_1}_{b_1} \tag{175}$$

$$\begin{aligned} (u_{ind-disp})^{a_1 b_1}_{r_1 s_1} = \Big[&\nu^{r_2 s_1}_{r_1 b_1}\,{}^{(0)}s^{a_1}_{r_2} - \nu^{a_1 s_1}_{a_2 b_1}\,{}^{(0)}s^{a_2}_{r_1} + \nu^{r_1 s_2}_{a_1 s_1}\,{}^{(0)}s^{b_1}_{s_2} - \nu^{r_1 b_1}_{a_1 b_2}\,{}^{(0)}s^{b_2}_{s_1} \\ &+(\omega_B)^{r_2}_{r_1}t^{a_1 b_1}_{r_2 s_1} - (\omega_B)^{a_1}_{a_2}t^{a_2 b_1}_{r_1 s_1} + (\omega_A)^{s_2}_{s_1}t^{a_1 b_1}_{r_1 s_2} - (\omega_A)^{b_1}_{b_2}t^{a_1 b_2}_{r_1 s_1}\Big]\Delta^{r_1 s_1}_{a_1 b_1}. \end{aligned} \tag{176}$$

DF integrals should be introduced into these amplitude equations in order to avoid the need for *ovvv* integrals. The T intermediates as well as a few new intermediates should be applied.

$$X^P_{a_1 r_1} = C^P_{r_1 r_2}\,{}^{(0)}s^{a_1}_{r_2} \tag{177}$$

$$Y^P_{a_1 r_1} = C^P_{a_1 a_2}\,{}^{(0)}s^{a_2}_{r_1} \tag{178}$$

$$X^P_{b_1 s_1} = C^P_{s_1 s_2}\,{}^{(0)}s^{b_1}_{s_2} \tag{179}$$

$$Y^P_{b_1 s_1} = C^P_{b_1 b_2}\,{}^{(0)}s^{b_2}_{s_1} \tag{180}$$

This allows the amplitude equations to be rewritten:

$$(u_{ind-disp})^{a_1}_{r_1} = \Big[2T^P_{a_1r_2}C^P_{r_1r_2} - 2T^P_{a_2r_1}C^P_{a_1a_2}\Big]\Delta^{r_1}_{a_1} \tag{181}$$

$$(u_{ind-disp})^{b_1}_{s_1} = \Big[2T^P_{b_1s_2}C^P_{s_1s_2} - 2T^P_{b_2s_1}C^P_{b_1b_2}\Big]\Delta^{s_1}_{b_1} \tag{182}$$

$$\begin{aligned}(u_{ind-disp})^{a_1b_1}_{r_1s_1} = \Big[&X^P_{a_1r_1}C^P_{b_1s_1} - Y^P_{a_1r_1}C^P_{b_1s_1} + C^P_{a_1r_1}X^P_{b_1s_1} - C^P_{a_1r_1}Y^P_{b_1s_1} \\ &+(\omega_B)^{r_2}_{r_1}t^{a_1b_1}_{r_2s_1} - (\omega_B)^{a_1}_{a_2}t^{a_2b_1}_{r_1s_1} + (\omega_A)^{s_2}_{s_1}t^{a_1b_1}_{r_1s_2} - (\omega_A)^{b_1}_{b_2}t^{a_1b_2}_{r_1s_1}\Big]\Delta^{r_1s_1}_{a_1b_1}.\end{aligned} \tag{183}$$

Once the amplitudes are available, the energy can be easily computed.

$$E^{(30)}_{ind-disp} = 2(u_{ind-disp})^{a_1}_{r_1}(\omega_B)^{r_1}_{a_1} + 2(u_{ind-disp})^{b_1}_{s_1}(\omega_A)^{s_1}_{b_1} + 4(u_{ind-disp})^{a_1b_1}_{r_1s_1}\nu^{r_1s_1}_{a_1b_1} \tag{184}$$

As was the case with $E^{(30)}_{ind}$, the $E^{(30)}_{ind-disp}$ should never be included in the interaction energy without its exchange counterpart, $E^{(30)}_{exch-ind-disp}$.

2.4.4 $E^{(30)}_{exch-ind-disp}$

The evaluation of the $E^{(30)}_{exch-ind-disp}$ term is significantly more involved than the $E^{(30)}_{ind-disp}$ term. Following from Ref. 162, this exchange correction can be written as a sum of five components.

$$\begin{aligned}E^{(30)}_{exch-ind-disp} = &E^{(30)}_{exch-ind-disp}(10) + E^{(30)}_{exch-ind-disp}(01) + E^{(30)}_{exch-ind-disp}(11) \\ &+E^{(30)}_{exch-ind-disp}(21) + E^{(30)}_{exch-ind-disp}(12)\end{aligned} \tag{185}$$

Here, the first three terms are defined in terms of induction-dispersion amplitudes and the $\tilde{V}$ integrals.

$$E^{(30)}_{exch-ind-disp}(10) = -2(u_{ind-disp})^{a_1}_{r_1}\tilde{V}^{r_1}_{a_1} \tag{186}$$

$$E^{(30)}_{exch-ind-disp}(01) = -2(u_{ind-disp})^{b_1}_{s_1}\tilde{V}^{s_1}_{b_1} \tag{187}$$

$$E^{(30)}_{exch-ind-disp}(11) = -2(u_{ind-disp})^{a_1b_1}_{r_1s_1}\tilde{V}^{r_1s_1}_{a_1b_1} \tag{188}$$

The remaining terms are significantly more complex. Only one of the corrections here, since the definition and factorization of $E^{(30)}_{exch-ind-disp}(12)$ can be inferred from $E^{(30)}_{exch-ind-disp}(21)$.

$$\begin{aligned} E^{(30)}_{exch-ind-disp}(21) =& 2\Big({}^{(0)}s^{a_1}_{r_1} t^{a_2 b_1}_{r_2 s_1} + {}^{(0)}s^{a_2}_{r_2} t^{a_1 b_1}_{r_1 s_1} \Big) \Big(S^{s_1}_{a_1} \tilde{\nu}^{r_1 r_2}_{a_2 b_1} - 2 S^{s_1}_{a_2} \tilde{\nu}^{r_1 r_2}_{a_1 b_1} - S^{b_2}_{a_2} S^{r_1}_{b_1} \tilde{\nu}^{r_2 s_1}_{a_1 b_2} \\ &+ 2 S^{b_2}_{a_2} S^{r_1}_{b_2} \tilde{\nu}^{r_2 s_1}_{a_1 b_1} + 2 S^{b_2}_{a_1} S^{r_1}_{b_1} \tilde{\nu}^{r_2 s_1}_{a_2 b_2} - S^{s_1}_{a_2} S^{r_1}_{b_2} \tilde{\nu}^{r_2 b_2}_{a_1 b_1} + 2 S^{s_1}_{a_1} S^{r_1}_{b_2} \tilde{\nu}^{r_2 b_2}_{a_2 b_1} \\ &+ 2 S^{s_1}_{a_2} S^{r_1}_{b_1} \tilde{\nu}^{r_2 b_2}_{a_1 b_2} \Big) - 8\, {}^{(0)}s^{a_2}_{r_2} t^{a_1 b_1}_{r_1 s_1} S^{b_2}_{a_1} S^{r_1}_{b_2} \tilde{\nu}^{r_2 s_1}_{a_2 b_1} - 8\, {}^{(0)}s^{a_1}_{r_1} t^{a_2 b_1}_{r_2 s_1} S^{s_1}_{a_1} S^{r_1}_{b_1} \tilde{\nu}^{r_2 b_2}_{a_2 b_2} \end{aligned} \quad (189)$$

As written above, some of the complexity is hidden; the amplitudes involved in this expression cannot be joined into a six-index quantity in an efficient implementation. This requires that the multiplication by the amplitudes is distributed, doubling the number of terms present. A few of these terms can benefit from the introduction of T intermediates, but perhaps the most straightforward way to implement this correction is to introduce generalized DF integrals and immediately contract the overlap integrals, ${}^{(0)}s$ amplitudes, and two-electron integrals. This will eventually reduce each term to a dispersion amplitude and either one four-index quantity or two two-index quantities. This will allow each of the terms to be evaluated in $\mathcal{O}(N^4)$ or $\mathcal{O}(N^5)$.

2.4.5 $E^{(30)}_{disp}$

The third-order dispersion term accounts for Axilrod-Teller-Muto-like terms within a dimer.[11,156] This correction has a simple form, but is quite expensive to evaluate:[162]

$$E^{(30)}_{disp} = 4\Big(t^{r_1 s_1}_{a_1 b_1} \nu^{r_2 s_2}_{r_1 s_1} t^{a_1 b_1}_{r_2 s_2} - t^{r_1 s_1}_{a_1 b_1} \nu^{r_2 b_1}_{r_1 b_2} t^{a_1 b_2}_{r_2 s_1} - t^{r_1 s_1}_{a_1 b_1} \nu^{a_1 s_2}_{a_2 s_1} t^{a_2 b_1}_{r_1 s_2} + t^{r_1 s_1}_{a_1 b_1} \nu^{a_1 b_1}_{a_2 b_2} t^{a_2 b_2}_{r_1 s_1} \Big). \quad (190)$$

The term in this correction that involves the $vvvv$ integrals scales as $\mathcal{O}(o^2 v^4)$. Unfortunately, the introduction of DF integrals does not improve the scaling of any of the four terms (all of which scale $\mathcal{O}(N^6)$). Henceforth, we will limit our discussion to the term involving the $vvvv$ integrals. This term is unique in that the implementation of it depends on what level of SAPT is going to be computed. At SAPT2+(3), the $E^{(30)}_{exch-disp}$ correction will be neglected and the $E^{(30)}_{disp}$ energy can be evaluated as shown above. At SAPT2+3 and beyond, the $E^{(30)}_{exch-disp}$ correction will be included and the third-order dispersion amplitudes are required. First, we will address the evaluation of this correction through the formation of

third-order dispersion amplitudes, then we will discuss a more efficient evaluation of the $E^{(30)}_{disp}$ energy directly. The third-order dispersion amplitudes are given as:

$$(u_{disp})^{a_1 b_1}_{r_1 s_1} = \left[\nu^{r_2 s_2}_{r_1 s_1} t^{a_1 b_1}_{r_2 s_2} - \nu^{r_2 b_1}_{r_1 b_2} t^{a_1 b_2}_{r_2 s_1} - \nu^{a_1 s_2}_{a_2 s_1} t^{a_2 b_1}_{r_1 s_2} + \nu^{a_1 b_1}_{a_2 b_2} t^{a_2 b_2}_{r_1 s_1}\right] \Delta^{r_1 s_1}_{a_1 b_1}. \tag{191}$$

There is little that can be done to improve the efficiency of the formation of these amplitudes. In principle, the full, four-fold permutational symmetry of the $\nu^{r_2 s_2}_{r_1 s_1}$ integrals could be exploited during their formation if sufficient reordering was done to allow their contraction with $t^{a_1 b_1}_{r_2 s_2}$. The $E^{(30)}_{disp}$ energy can be evaluated easily once the third-order dispersion amplitudes have been obtained.

$$E^{(30)}_{disp} = 4 (u_{disp})^{a_1 b_1}_{r_1 s_1} \nu^{r_1 s_1}_{a_1 b_1} \tag{192}$$

If the amplitudes are not required, it is possible to evaluate the $E^{(30)}_{disp}$ energy more efficiently by contracting the dispersion amplitudes and subsequently multiplying by the *vvvv* integrals. It is possible to exploit the four-fold symmetry of the *vvvv* integrals in their formation, which scales $\mathcal{O}(v^4 N_{df})$, and a two-fold symmetry in the contraction of the dispersion amplitudes, which scales $\mathcal{O}(o^2 v^4)$. It should be noted that the resultant of the contraction of the dispersion amplitudes has four-fold symmetry, but that this cannot be exploited if the contraction is to be cast as a matrix multiply. As the evaluation of the $E^{(30)}_{disp}$ term in SAPT2+(3) computation can take a sizable fraction of the total time, pseudocode is presented to elaborate on an optimal implementation. The bulk of the work is spent in two matrix multiplies, each called v times. The restriction that $r2 <= r1$ can be placed explicitly in the matrix multiply calls.

2.4.6 $E^{(30)}_{exch-disp}$

The $E^{(30)}_{exch-disp}$ is not as crucial as the $E^{(30)}_{exch-ind}$ or $E^{(30)}_{exch-ind-disp}$ due to the generally small magnitude of $E^{(30)}_{disp}$. The $E^{(30)}_{exch-disp}$ correction contains terms that scale $\mathcal{O}(N^7)$ that we will neglect (following from Ref. 162), otherwise $E^{(30)}_{exch-disp}$ is relatively inexpensive to evaluate once the third-order dispersion amplitudes have been formed. The portions of $E^{(30)}_{exch-disp}$ that we will consider are:[162]

$$E^{(30)}_{exch-disp} = E^{(30)}_{exch-disp}(11) + E^{(30)}_{exch-disp}(20) + E^{(30)}_{exch-disp}(02) + E^{(30)}_{exch-disp}(22), \tag{193}$$

Algorithm 8 Pseudocode to evaluate the $E_{disp}^{(30)}$ correction.

Read/Form $t_{r_1 s_1}^{a_1 b_1}$ with $a_1 b_1$ as the fast running index
Read $C_{r_1 r_2}^{P}$ with permutational symmetry
Read $C_{s_1 s_2}^{P}$ with permutational symmetry
for $P <$ #DF Functions **do**
  for $r_1 <$ #VirtualA **do**
    for $r_2 < r_1$ **do**
      $C_{r_1 r_2}^{P} \leftarrow 2 C_{r_1 r_2}^{P}$
    end for
  end for
end for
for $P <$ #DF Functions **do**
  for $s_1 <$ #VirtualB **do**
    for $s_2 < s_1$ **do**
      $C_{s_1 s_2}^{P} \leftarrow 2 C_{s_1 s_2}^{P}$
    end for
  end for
end for
for $r_1 <$ #VirtualA **do**
  for $r_2 <= r_1$ **do**
    $X_{s_1 s_2} \leftarrow t_{r_1 s_1}^{a_1 b_1} t_{a_1 b_1}^{r_2 s_2}$
    $Y_{s_1 s_2} \leftarrow C_{r_1 r_2}^{P} C_{s_1 s_2}^{P}$
    for $s_1 <$ #VirtualB **do**
      for $s_2 <= s_1$ **do**
        $Z_{s_1 s_2} \leftarrow X_{s_1 s_2} + X_{s_2 s_1}$
      end for
    end for
    $E_{disp}^{(30)} \leftarrow 2 Y_{s_1 s_2} Z_{s_1 s_2}$
  end for
end for

where

$$E^{(30)}_{exch-disp}(11) = -2(u_{disp})^{a_1b_1}_{r_1s_1}\tilde{V}^{r_1s_1}_{a_1b_1} \tag{194}$$

$$\begin{aligned} E^{(30)}_{exch-disp}(20) =& 4\Bigg(\Big[\nu^{a_1s_1}_{r_1b_1}t^{a_2b_1}_{r_2s_1} + \nu^{a_2s_1}_{r_2b_1}t^{a_1b_1}_{r_1s_1}\Big]\Delta^{r_1r_2}_{a_1a_2}\Bigg)\Bigg(S^{b_1}_{a_2}\tilde{\nu}^{r_2r_1}_{a_1b_1} - 2S^{b_1}_{a_1}\tilde{\nu}^{r_2r_1}_{a_2b_1} \\ &+2S^{b_1}_{a_2}S^{r_1}_{b_1}\tilde{\nu}^{r_2b_2}_{a_1b_2} - 4S^{b_1}_{a_1}S^{r_1}_{b_1}\tilde{\nu}^{r_2b_2}_{a_2b_2} - S^{b_1}_{a_2}S^{r_1}_{b_2}\tilde{\nu}^{r_2b_2}_{a_1b_1} + 2S^{b_1}_{a_1}S^{r_1}_{b_2}\tilde{\nu}^{r_2b_2}_{a_2b_1}\Bigg) \end{aligned} \tag{195}$$

$$\begin{aligned} E^{(30)}_{exch-disp}(22) =& 2t^{a_1b_1}_{r_1s_1}t^{a_2b_2}_{r_2s_2}\Big(2S^{s_2}_{a_2}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_1}_{a_1b_2} + 2S^{s_1}_{a_2}S^{r_2}_{b_2}\tilde{\nu}^{r_1s_2}_{a_1b_1} - 4S^{s_1}_{a_2}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_2}_{a_1b_2} \\ &+2S^{s_2}_{a_1}S^{r_2}_{b_2}\tilde{\nu}^{r_1s_1}_{a_2b_1} - S^{s_2}_{a_1}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_1}_{a_2b_2} - S^{s_1}_{a_1}S^{r_2}_{b_2}\tilde{\nu}^{r_1s_2}_{a_2b_1} + 2S^{s_1}_{a_1}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_2}_{a_2b_2}\Big). \end{aligned} \tag{196}$$

Again, the (02) term can be inferred from the (20) term, so it won't be addressed explicitly. First, we will consider the implementation of the $E^{(30)}_{exch-disp}(20)$ term. Here, the most expensive portion is the $\mathcal{O}(o^3v^3)$ contraction of dispersion amplitudes and integrals. This can be avoided if the T intermediates are introduced.

$$\Big[\nu^{a_1s_1}_{r_1b_1}t^{a_2b_1}_{r_2s_1} + \nu^{a_2s_1}_{r_2b_1}t^{a_1b_1}_{r_1s_1}\Big]\Delta^{r_1r_2}_{a_1a_2} = \Big[C^P_{a_1r_1}T^P_{a_2r_2} + T^P_{a_1r_1}C^P_{a_2r_2}\Big]\Delta^{r_1r_2}_{a_1a_2} \tag{197}$$

This substitution reduces the scaling to $\mathcal{O}(o^2v^2N_{df})$. The remaining work can be handled easily if the DF integrals are introduced and the overlap integrals are contracted with them to form *ovov* quantities. These can then be multiplied by the amplitudes to evaluate their contribution to $E^{(30)}_{exch-disp}$. The $E^{(30)}_{exch-disp}(22)$ term is slightly trickier, but, again, the introduction of the T intermediates greatly improves the efficiency.

$$\begin{aligned} E^{(30)}_{exch-disp}(22) =& 4T^P_{b_1s_1}C^P_{b_2s_1}S^{s_2}_{a_2}S^{r_2}_{b_1}t^{a_2b_2}_{r_2s_2} + 4T^P_{b_1s_1}C^P_{b_1s_2}S^{s_1}_{a_2}S^{r_2}_{b_2}t^{a_2b_2}_{r_2s_2} - 8T^P_{a_2r_2}T^P_{b_1s_1}S^{s_1}_{a_2}S^{r_2}_{b_1} \\ &+4T^P_{a_1r_1}C^P_{a_2r_1}S^{s_2}_{a_1}S^{r_2}_{b_2}t^{a_2b_2}_{r_2s_2} - 2t^{a_1b_1}_{r_1s_1}t^{a_2b_2}_{r_2s_2}S^{s_2}_{a_1}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_1}_{a_2b_2} \\ &-2t^{a_1b_1}_{r_1s_1}t^{a_2b_2}_{r_2s_2}S^{s_1}_{a_1}S^{r_2}_{b_2}\tilde{\nu}^{r_1s_2}_{a_2b_1} + 4T^P_{a_2r_2}C^P_{a_2r_1}S^{s_1}_{a_1}S^{r_2}_{b_1}t^{a_1b_1}_{r_1s_1} \end{aligned} \tag{198}$$

Of the two terms that cannot use the T intermediates, the second is very easily evaluated (contraction of the S integrals with the dispersion amplitudes leads to two two-index quantities); evaluation of the other term scales $\mathcal{O}(o^4v^2)$ if it is properly factored.

$$X^{a_2b_2}_{b_1a_1} = t^{a_2b_2}_{r_2s_2}S^{s_2}_{a_1}S^{r_2}_{b_1} \tag{199}$$

$$Y^{a_2b_2}_{r_1s_1} = t^{a_1b_1}_{r_1s_1}X^{a_2b_2}_{b_1a_1} \tag{200}$$

$$-2t^{a_1b_1}_{r_1s_1}t^{a_2b_2}_{r_2s_2}S^{s_2}_{a_1}S^{r_2}_{b_1}\tilde{\nu}^{r_1s_1}_{a_2b_2} = -2Y^{a_2b_2}_{r_1s_1}\tilde{\nu}^{r_1s_1}_{a_2b_2} \tag{201}$$

With these factorizations, all but one contraction (above) in $E^{(30)}_{exch-disp}$ can be evaluated in $\mathcal{O}(N^5)$. Assuming that the third-order dispersion amplitudes have been formed, the evaluation of their exchange counterpart is trivial.

2.5 Natural Orbitals in SAPT

2.5.1 Triples Correction to Dispersion

The following is adapted from Ref. 95.

We use MP2 naturals orbitals (NOs)[42,137] instead of HF molecular orbitals (MOs) to evaluate the triples correction; this allows an appreciable fraction of the virtual orbitals to be removed from the computation without significant loss of accuracy. Natural orbitals are those orbitals which diagonalize the one-particle density matrix (OPDM).[137] For a two-electron system, natural orbitals comprise the basis which requires the fewest configurations to reach a given accuracy in the energy,[138] and in general, the natural orbitals tend to concentrate the electron correlation energy into the those virtual NOs with the largest occupation numbers (one-particle density matrix eigenvalues). Conversely, the virtual NOs with the smallest eigenvalues contribute very little to the correlation energy and may be neglected. Natural orbitals have been used to select active spaces or as guess orbitals in multi-configurational self-consistent-field (MCSCF) computations,[108,175,187] or as replacements for fully-optimized MCSCF orbitals.[2,17] They have also been used in highly-correlated configuration interaction computations[1,30–33,66,196] and coupled-cluster computations.[124,131,219] The optimized virtual orbital subspace (OVOS) approach of Adamowicz and Bartlett[3] is an alternative technique with the same goal of limiting the number of virtual orbitals for highly-correlated computations; this approach has been reformulated by Urban and co-workers[159] and used to reduce the cost of CCSD(T) computations,[166] including an impressive recent study of the benzene dimer.[168] A related approach to truncate the virtual space using pair natural orbitals (PNOs)[50,143,144] has been recently explored by Neese *et. al.* in the context of the coupled-electron pair approximation (CEPA), CCSD, and quadratic configuration interaction with single and double excitations (QCISD).[157,158] Although shown to have many promising applications, these methods may not be well suited

for the study of weakly interacting systems.[157]

Reducing the number of virtual orbitals in the evaluation of $E_{disp}^{(22)}$(T) is extremely beneficial due to its $\mathcal{O}(o^3v^4)$ scaling. However, only a small number of virtual HF MOs can be removed before the accuracy of the computation is severely impacted. To avoid this problem, we will use MP2 NOs in place of the HF MOs. The (unrelaxed) MP2 OPDM is given as:

$$P_{ij} = -2\frac{[2(ia|kb) - (ib|ka)](ja|kb)}{D_{ik}^{ab}D_{jk}^{ab}}, \tag{202}$$

$$P_{ab} = 2\frac{[2(ia|jc) - (ic|ja)](ib|jc)}{D_{ij}^{ac}D_{ij}^{bc}}, \tag{203}$$

where i, j, k correspond to occupied orbitals and a, b, c correspond to unoccupied orbitals and $D_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$. Summation over repeated indices is implied. In an SAPT computation, the MP2 OPDM is used to compute the $E_{elst}^{(12)}$ term, so, it will already be available. In this work, we will correlate all electrons in the MP2 OPDM formation regardless of whether or not they are correlated in the $E_{disp}^{(22)}$(T) evaluation.

The equation for $E_{disp}^{(22)}$T) presented above assumes that the Fock matrix for each monomer is diagonal. MP2 NOs do not diagonalize the Fock matrix, so they must be modified before they can be used to evaluate $E_{disp}^{(22)}$(T). Our procedure for generating a set of usable MP2 NOs is summarized below:

1. Form the MP2 OPDM in the HF MO basis.
2. Diagonalize the MP2 OPDM to obtain MP2 NOs (in the HF MO basis).
3. Truncate the MP2 NO virtual space.
4. Transform the MO based Fock matrix into the truncated MP2 NO basis.
5. Diagonalize the NO based Fock matrix to obtain semicanonical MP2 NOs (in the MP2 NO basis) and orbital energies.
6. Express the semicanonical MP2 NOs in terms of the AOs.

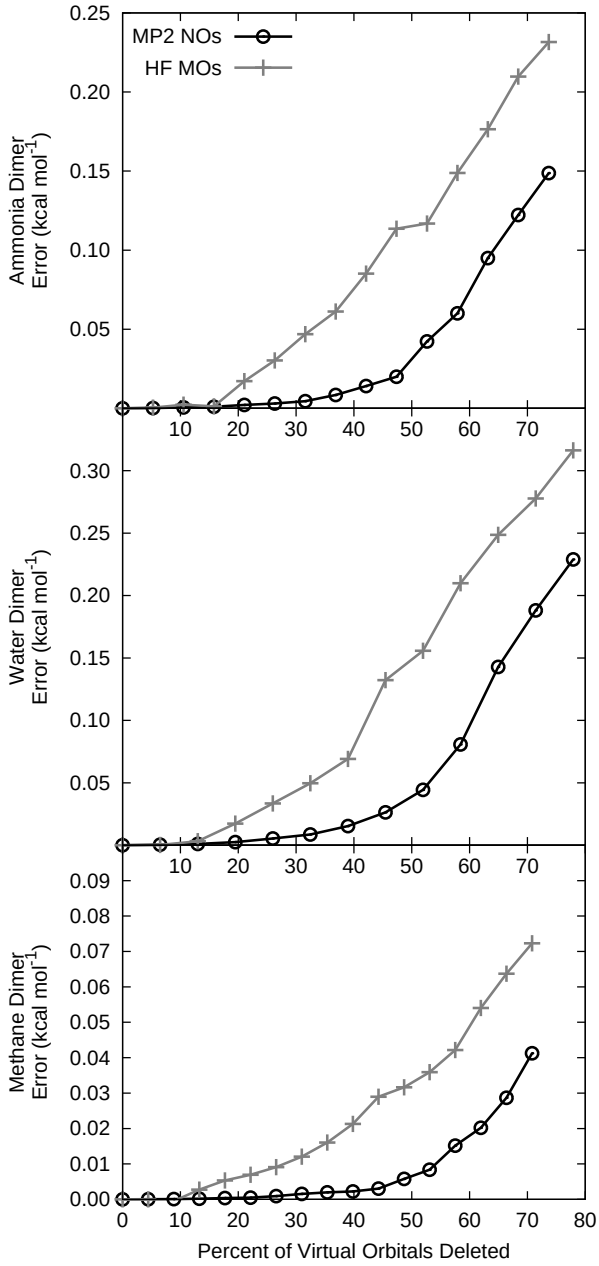
7. Using the above NO/AO transformation matrix, transform the integrals needed to evaluate $E_{disp}^{(22)}$(T) from the AO basis into the semicanonical MP2 NO basis and evaluate this term as usual.

The eigenvalues of the MP2 OPDM are occupation numbers that represent the number of electrons in each NO. These eigenvalues are a convenient metric for removing virtual orbitals from the computation. Since the unrelaxed MP2 OPDM is used, the occupied HF orbitals are recovered (the NO based Fock matrix is block diagonal); this is often called the "frozen natural orbital" procedure. If no virtual orbitals are removed, the virtual HF MOs are also recovered in the semicanonicalization process.

The use of MP2 NOs is based on the experience that NOs are more efficient than HF MOs for capturing electron correlation in the low-lying (or most heavily occupied) orbitals. This can be demonstrated in the present case for the evaluation of $E_{disp}^{(22)}$(T) by removing virtual HF MOs and MP2 NOs at the same rate and determining how much error arises from the reduced virtual space. The results of such a test are shown in Figure 2 for the ammonia dimer (95 virtual orbitals), water dimer (77 virtual orbitals), and methane dimer (113 virtual orbitals). The MP2 NOs are clearly superior to HF MOs for reproducing the $E_{disp}^{(22)}$(T) correction with a smaller virtual space. However, the error associated with removing virtual MP2 NOs still increases too rapidly to remove more than roughly one third of the virtual orbitals if the error is to be kept negligible (greater fractions may be removed if small to modest errors may be tolerated). This is a great improvement over the HF MOs; only the few most high lying virtual HF MOs can be removed before significant errors begin to accrue.

As previously mentioned, the scaling of the $E_{disp}^{(22)}$(T) correction is $\mathcal{O}(o^3v^4)$. Assuming ideal behavior, removing one third of the virtual orbitals would lead to roughly a 5X speedup. If one half of the virtual orbitals could be removed, it would result in a 16X speedup. In Figure 2, the error created by removing half of the NOs is below 0.05 kcal mol^{-1} (or less than 15% of the total $E_{disp}^{(22)}$(T) contribution), which is probably acceptable in many or most applications. However, here we wish to explore ways in which we may achieve this level of computational savings while allowing truly negligible errors. To improve

Figure 2: Errors (in kcal mol^{-1}) of the $E_{disp}^{(22)}$(T) correction evaluated with the aug-cc-pVDZ basis set as virtual orbitals (HF MO or MP2 NO) are removed from the computation. The total $E_{disp}^{(22)}$(T) correction for these test cases are -0.281 (ammonia dimer), -0.344 (water dimer), and -0.102 kcal mol^{-1} (methane dimer).



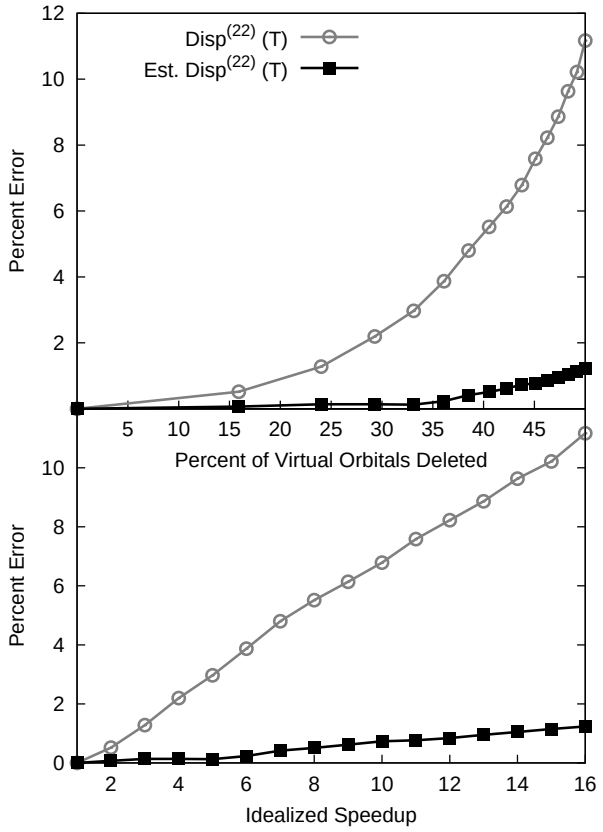
this approximation, we will assume that the magnitude of the $E_{disp}^{(22)}(\mathrm{T})$ correction changes at the same rate as $E_{disp}^{(20)}$ when virtual orbitals are removed, *viz.*:

$$\frac{E_{disp}^{(22)}(\mathrm{T})_{exact}}{E_{disp}^{(22)}(\mathrm{T})_{approx}} \approx \frac{E_{disp,exact}^{(20)}}{E_{disp,approx}^{(20)}}. \tag{204}$$

The $E_{disp}^{(20)}$ term scales as $\mathcal{O}(o^2v^2)$, so, for a given system, if it is possible to evaluate the $E_{disp}^{(22)}(\mathrm{T})$ correction, it is trivial to evaluate $E_{disp}^{(20)}$. We will denote the $E_{disp}^{(22)}(\mathrm{T})$ correction evaluated using this approximation as $Est.\,E_{disp}^{(22)}(\mathrm{T})$ to signify that the value is estimated, not computed explicitly. In order to determine whether or not this approximation is valid, $E_{disp}^{(22)}(\mathrm{T})$ and $Est.E_{disp}^{(22)}(\mathrm{T})$ are computed for the 11 smallest complexes from the S22 test set. Virtual orbitals are removed to determine how effective the removal of virtual orbitals will be for increasing computational efficiency. The results of this analysis are shown in Figure 3. The $Est.\,E_{disp}^{(22)}(\mathrm{T})$ correction performs much better than the unscaled correction. Simply removing a certain fraction of the virtual orbitals (as is the case in Figure 3) can provide a 16X speedup with only 1.25% error. The unscaled $E_{disp}^{(22)}(\mathrm{T})$ can only provide a 3X speedup with the same accuracy. Clearly, the approximation of Equation 204 significantly increases the number of virtual orbitals that can be removed while keeping the error negligible.

To this point, the number of virtual orbitals removed was not physically motivated. Each natural orbital has occupation associated with it (eigenvalues of the MP2 OPDM). We will use these values as a metric to determine which virtual orbitals can be removed. In Figure 4, again, $E_{disp}^{(22)}(\mathrm{T})$ and $Est.\,E_{disp}^{(22)}(\mathrm{T})$ are computed for the smallest 11 complexes in the S22 test set with different cutoffs based on the number of electrons in an orbital. Once again, the $Est.\,E_{disp}^{(22)}(\mathrm{T})$ correction is far superior to the unscaled correction. With this scaling, a cutoff of 10^{-6} electrons creates less than 1% error. In addition to the reduced virtual space, the computations shown in Figure 4 were performed under the frozen core approximation. It should be noted that $E_{disp,approx}^{(20)}$ in Equation 204 should include all relevant approximations (*i.e.* in this case it was computed with the core electrons frozen). This leads to our recommendation for evaluating the $E_{disp}^{(22)}(\mathrm{T})$ correction: virtual orbitals with less than 10^{-6} electrons should be removed, core electrons should be frozen, and the result should be scaled according to Equation 204. This procedure introduces only negligible

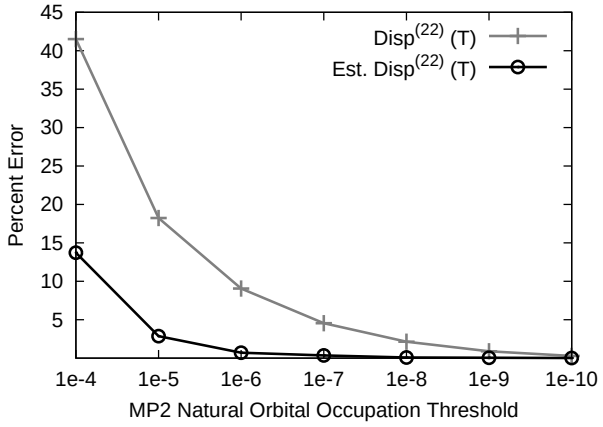
Figure 3: Percent errors for the $E_{disp}^{(22)}$(T) correction (scaled and unscaled) averaged over the 11 smallest complexes from the S22 test set[117] as virtual MP2 NOs are removed from the computation to achieve a certain percentage of deleted virtual orbitals (top panel) or to obtain a specified idealized speedup (bottom panel).



errors but greatly reduces computational costs. Of course, in very large molecules, one may wish to employ somewhat more aggressive truncations.

The accuracy of the approximations outlined above is assessed in Table 6 for the entire S22 test set. The $E_{disp}^{(22)}$(T) values in this table were all computed using the DF approximation of the two-electron integrals; errors due to the DF approximation have been considered previously[93,94] and are generally in the hundredths of one kcal mol^{-1} or less. The errors reported in Table 6 reflect the removal of virtual orbitals, the frozen core approximation, and the scaling shown in Equation 204. These approximations introduce only modest errors into the computation; the largest error is only 0.02 kcal mol^{-1} (appearing twice in the table, for

Figure 4: Percent errors for the $E_{disp}^{(22)}$(T) correction (scaled and unscaled) as virtual orbitals with less than a specified number of electrons are removed from the computation for the 11 smallest complexes from the S22 test set.[117]



molecules with total $E_{disp}^{(22)}$(T) contributions of -2.765 and -4.598 kcal mol^{-1}); such a small error seems acceptable for nearly any application. For these cases and the aug-cc-pVDZ basis set, a threshold of 10^{-6} electrons for the removal of virtual orbitals typically removes about half of the virtual orbitals. However, for cases where one monomer is much larger than the other (*e.g.*, benzene-methane), the majority of the smaller monomer's virtual orbitals can be removed. Note that in this implementation of SAPT, all computations are performed in the dimer basis. Because of this, some of the virtual orbitals will be composed mainly of basis functions centered on the other monomer. The procedure of removing MP2 NOs is a way of removing these extraneous virtual orbitals while retaining the important orbitals. It follows from this consideration that the hydrogen bonded and mixed complexes in the S22 will benefit the most from the removal of virtual orbitals, since they will have more spatially distant basis functions.

As defined above, the truncation of the virtual space using MP2 NO occupations will not result in a continuous potential energy surface. At two adjacent points, it is possible that different numbers of virtual orbitals will be removed. Due to the excellent performance of this approximation, however, discontinuities are likely to be unnoticeably small. Additionally, derivatives of SAPT energies are not typically computed. If continuity becomes

Table 6: The effect of the MP2 NO and frozen core approximations on the $E_{disp}^{(22)}$(T) correction evaluated with the aug-cc-pVDZ basis set for the complexes of the S22 test set (Ref. 117).[a]

	$E_{disp}^{(22)}(\mathrm{T})_{exact}$	Error[b]	HF MOs[c]		MP2 NOs[d]	
			A	B	A	B
H-Bonded Complexes						
$(NH_3)_2$	-0.281	0.003	95	95	46	46
$(H_2O)_2$	-0.344	0.003	77	77	36	39
Formic Acid Dimer	-1.767	0.008	162	162	83	83
Formamide Dimer	-1.389	0.010	180	180	90	90
Uracil Dimer	-1.838	0.005	411	411	194	194
2-Pyridoxine·2-Aminopyridine	-2.098	0.006	396	396	183	191
Adenine·Thymine WC	-2.153	0.004	501	503	241	229
Dispersion Dominated Complexes						
$(CH_4)_2$	-0.102	0.002	113	113	53	53
$(C_2H_4)_2$	-0.342	0.005	156	156	75	75
Benzene·CH_4	-0.514	0.006	230	246	165	55
PD Benzene Dimer	-2.396	0.008	363	363	174	174
Pyrazine Dimer	-2.601	0.004	327	327	159	159
Uracil Dimer	-2.765	0.020	411	411	231	231
Stacked Indole·Benzene	-3.708	0.004	441	431	179	242
Stacked Adenine·Thymine	-4.598	0.020	501	503	255	241
Mixed Complexes						
Ethene·Ethine	-0.232	0.001	138	139	74	54
Benzene·H_2O	-0.568	0.008	212	228	164	40
Benzene·NH_3	-0.552	0.007	221	237	165	48
Benzene·HCN	-0.736	0.005	226	240	167	52
T-shaped Benzene Dimer	-1.119	0.007	363	363	168	172
T-shaped Indole Benzene	-1.624	0.009	441	431	174	230
Phenol Dimer	-1.445	0.005	405	405	190	190

[a]MP2 NOs with occupancies less than 10^{-6} electrons are removed. Errors given in kcal mol^{-1}. [b]Signed error computed as $Est.\,E_{disp}^{(22)}(\mathrm{T})_{approx} - E_{disp}^{(22)}(\mathrm{T})_{exact}$. [c]Number of unoccupied HF MOs for monomer A and B. [d]Number of virtual MP2 NOs remaining after orbitals are removed.

an issue, we recommend determining a reasonable number of virtual orbitals to remove and fixing that number for the entire surface.

2.5.2 *vvvv* Two-Electron Integrals

Similar DF and NO based approximations can be applied[100] to the CCD+ST(CCD) dispersion approach of Williams *et al.*[247] (*vide infra*). This method has been found to be extremely accurate; however, it has been all but abandoned due to computational expense. We will refer to this approach as $\epsilon^{(2)}_{disp}$(CCD) throughout the present work. Asymptotically, the expense of this method is no different than any of the other higher-order SAPT methods; it also scales $\mathcal{O}(o^3v^4)$. The difference is that the evaluation of $\epsilon^{(2)}_{disp}$(CCD) requires the iterative solution of four sets of monomer CCD amplitudes, each of which scales as $\mathcal{O}(o^2v^4)$. Additionally, there are dispersion amplitudes that require iterative $\mathcal{O}(o^3v^3)$ work to solve. As is the case with any CCD or CCSD implementation, the majority of the work is the evaluation of the term involving v^4 integrals,

$$t^{a_1a_2}_{r_1r_2} \leftarrow \nu^{r_1r_2}_{r_3r_4} t^{a_1a_2}_{r_3r_4}. \tag{205}$$

In an effort to improve the efficiency of this term, we apply the same NO approximations that were successful for the $E^{(22)}_{disp}$(T) evaluation. First, we transform the virtual orbitals to the NO basis (we will denote the MO to NO transformation matrix V and label virtual NO indices with ρ).

$$t^{a_1a_2}_{\rho_1\rho_2} = V^{r_1}_{\rho_1} t^{a_1a_2}_{r_1r_2} V^{r_2}_{\rho_2} \tag{206}$$

Once the amplitudes are transformed to the NO basis, the term is evaluated as usual with v^4 integrals also in the NO basis.

$$t^{a_1a_2}_{\rho_1\rho_2} \leftarrow \nu^{\rho_1\rho_2}_{\rho_3\rho_4} t^{a_1a_2}_{\rho_3\rho_4} \tag{207}$$

Finally, this result is backtransformed to the MO basis and added to the CCD amplitudes.

$$t^{a_1a_2}_{r_1r_2} = V^{r_1}_{\rho_1} t^{a_1a_2}_{\rho_1\rho_2} V^{r_2}_{\rho_2} \tag{208}$$

The remainder of the CCD terms are evaluated in the MO basis. In practice, we have found that this approximation leads to significant speedups and introduces negligible errors. The

triples contribution to $\epsilon^{(2)}_{disp}$(CCD) is also evaluated in the NO basis. This is accomplished by transforming the CCD amplitudes to the NO basis and proceeding identically to the $E^{(22)}_{disp}$(T) evaluation.[100] This approach immediately extends to the second-order doubles amplitudes that appear in SAPT2 (and higher) corrections. Unfortunately, the $E^{(30)}_{disp}$ correction, which contains $(rr|ss)$ integrals, cannot be accurately computed with a reduced virtual space.

2.6 Density Fitting Errors in SAPT

2.6.1 DF/CD-SAPT0

The following was adapted from Ref. 93.

The approximate SAPT0 methods were assessed on the basis of their ability to reproduce the necessary MO integrals and on the deviation of the SAPT0 components. In this work, we will utilize the naming convention for MO indices common in the SAPT literature. The occupied and virtual orbitals of monomer A will be labeled a and r, respectively. The occupied and virtual orbitals of monomer B will be labeled b and s, respectively. It is important to note that the Cholesky decomposition (CD) guarantees the AO integrals to a specified accuracy; however, this error bound does not apply to the transformed MO integrals. For the purpose of computing SAPT0 energies, it was found that a CD threshold of 10^{-3}-10^{-4} E_h was reasonable. For a given AO basis, the former value tends to create a slightly smaller Cholesky basis than the corresponding DF basis, while the latter value will create a significantly larger Cholesky basis. This is illustrated in Table 7, which compares fitting basis size and errors in the SAPT0 energy components for the water, ammonia, and methane dimers.

Through a comparison of the approximate DF and CD MO integrals with the exact integrals, it was determined that the largest errors occur for the integrals centered entirely on one monomer in both cases. This includes the *aarr*, *arar*, *bbss*, and *bsbs* classes of integrals. Because these integrals are greater in magnitude than those which span both monomers, this result is not surprising. The largest errors for these integrals tend to be on the order of 10^{-1} E_h. In SAPT0, these integrals are used only to compute the orbital response coefficients that are involved in the evaluation of the $E^{(20)}_{ind,resp}$ and $E^{(20)}_{exch-ind,resp}$

Table 7: Deviation (in microhartree) from conventional SAPT0/aug-cc-pVDZ for water dimer, ammonia dimer, and methane dimer.[a]

	DF-SAPT	CD-SAPT		1C-CD-SAPT	
		10^{-3}	10^{-4}	10^{-3}	10^{-4}
Water Dimer					
N[b]	236	221	389	213	316
$E^{(10)}_{elst}$	1.52	-2.16	0.04	0.08	0.08
$E^{(10)}_{exch}$	4.63	6.43	0.55	5.10	0.69
$E^{(20)}_{ind,resp}$	1.28	-0.03	0.05	-0.71	0.05
$E^{(20)}_{exch-ind,resp}$	-0.24	-0.04	-0.01	0.09	0.01
$E^{(20)}_{disp}$	0.49	-0.25	-0.09	-1.35	-0.04
$E^{(20)}_{exch-disp}$	-1.76	-2.75	-0.13	-0.95	-0.22
E_{SAPT0}	5.93	1.20	0.41	2.25	0.57
Ammonia Dimer					
N[b]	282	247	412	244	355
$E^{(10)}_{elst}$	-1.07	-0.86	-0.12	-1.24	0.09
$E^{(10)}_{exch}$	5.15	8.37	1.13	7.14	2.11
$E^{(20)}_{ind,resp}$	0.03	-0.05	0.00	0.31	-0.10
$E^{(20)}_{exch-ind,resp}$	-0.03	0.00	-0.02	-0.29	-0.01
$E^{(20)}_{disp}$	-0.35	-0.19	-0.12	-1.66	0.04
$E^{(20)}_{exch-disp}$	-1.24	-2.58	-0.27	-1.78	-0.55
E_{SAPT0}	2.50	4.69	0.59	2.46	1.58
Methane Dimer					
N[b]	328	278	450	260	392
$E^{(10)}_{elst}$	-4.35	-0.06	-0.01	0.07	-0.10
$E^{(10)}_{exch}$	2.53	2.70	0.64	2.81	0.67
$E^{(20)}_{ind,resp}$	-0.02	0.01	0.00	0.01	0.00
$E^{(20)}_{exch-ind,resp}$	0.01	-0.01	0.00	-0.01	0.00
$E^{(20)}_{disp}$	-0.38	-0.25	0.01	-0.47	-0.06
$E^{(20)}_{exch-disp}$	-0.68	-1.13	-0.29	-0.97	-0.27
E_{SAPT0}	-2.90	1.27	0.34	1.43	0.24

[a]The geometries are taken from the S22 test set (Ref. 117). [b]Number of functions in the DF or CD basis.

terms. These integrals do not directly contribute to the SAPT0 energy, but they affect the energy indirectly when the response coefficients are contracted against the *arbb* or *aabs* type integrals. As shown in Tables 7 and 8, the error incurred in the evaluation of the $E^{(20)}_{ind,resp}$ and $E^{(20)}_{exch-ind,resp}$ terms with approximate integrals is negligible.

Table 8: Errors (in kcal mol^{-1}) of DF-SAPT and 1C-CD-SAPT for the S22 test set (Ref. 117) relative to conventional SAPT0/aug-cc-pVDZ.

	MAX[a]	MSE[b]	MUE[c]	RMS[d]
DF-SAPT				
$E^{(10)}_{elst}$	0.003	0.000	0.001	0.001
$E^{(10)}_{exch}$	0.014	0.005	0.005	0.007
$E^{(20)}_{ind,resp}$	0.002	0.000	0.000	0.001
$E^{(20)}_{exch-ind,resp}$	0.000	0.000	0.000	0.000
$E^{(20)}_{disp}$	0.002	0.000	0.001	0.001
$E^{(20)}_{exch-disp}$	0.007	-0.003	0.003	0.003
E_{SAPT0}	0.006	0.002	0.003	0.003
1C-CD-SAPT[e]				
$E^{(10)}_{elst}$	0.006	0.000	0.001	0.002
$E^{(10)}_{exch}$	0.057	0.011	0.011	0.017
$E^{(20)}_{ind,resp}$	0.006	-0.001	0.001	0.002
$E^{(20)}_{exch-ind,resp}$	0.002	0.000	0.000	0.001
$E^{(20)}_{disp}$	0.003	-0.001	0.001	0.001
$E^{(20)}_{exch-disp}$	0.038	-0.007	0.007	0.011
E_{SAPT0}	0.023	0.002	0.003	0.006

[a]Maximum absolute error. [b]Mean signed error. [c]Mean unsigned error. [d]Root mean square error. [e]AO integrals computed with a tolerance of 10^{-3} E_h.

As indicated in Table 7, the 1C-CD-SAPT results are very similar to those for CD-SAPT and for low tolerances, the size of the Cholesky basis is not reduced significantly. However, the automatic exclusion of two-center terms from the Cholesky basis makes the 1C-CD algorithm more efficient than the full CD algorithm. It is apparent that for the three small test systems in Table 7, a CD tolerance of 10^{-3} is acceptable for obtaining very accurate SAPT0 energy components, with errors of less than 0.01 millihartree (0.006 kcal mol^{-1}). Of the CD methods shown in Table 7, the 1C-CD-SAPT with a tolerance of 10^{-3} appears to have the most promise for general applicability.

At this point, it should be noted that, in practice, the error associated with the three-index integral approximations for the SAPT0 total interaction energy may only depend on the $E^{(20)}_{disp}$ and $E^{(20)}_{exch-disp}$ terms (assuming the integral approximations are not also used

for the Hartree-Fock computations). Often, a $\delta E^{(HF)}_{ind,resp}$ term is computed from the HF interaction energy as:

$$\delta E^{(HF)}_{ind,resp} = E^{HF}_{int} - \left(E^{(10)}_{elst} + E^{(10)}_{exch} + E^{(20)}_{ind,resp} + E^{(20)}_{exch-ind,resp} \right). \tag{209}$$

This term captures induced-multipole induced-multipole interactions that are not described by the $E^{(20)}_{ind,resp}$ and $E^{(20)}_{exch-ind,resp}$ terms. In CD- or DF-SAPT0, adding this term also has the effect of removing some fitting error from the total SAPT0 energy, since our SAPT implementation uses exact HF computations as a starting point. It is also important to note that SAPT computations are generally used to get a qualitative understanding of the fundamental physics of nonbonded interactions. A deviation of a few hundredths of one kcal mol^{-1} does not affect the SAPT results qualitatively. It should also be noted that SAPT is a perturbational method of computing interaction energies directly; as a consequence, the fitting error that occurs with SAPT is much smaller than the fitting error that occurs in an MP2 total energy, for example. This is somewhat similar to the observation by Böstrom et al.[19,20] that the CD threshold does not need to be as tight for excitation energy computations as it is for total energy computations.

To assess the performance of the 1C-CD-SAPT0 with a tolerance of 10^{-3}, this method and DF-SAPT0 were applied to the S22 test set. These results are shown in Table 8. As mentioned previously, the errors for the induction terms are negligible; for these cases, the error is always less than a hundredth of a kcal mol^{-1}. Perhaps surprisingly, the error in the electrostatic term is also very low. This term contains a small number of fairly large two-electron integrals; evidently, the *aabb* type integrals are approximated well by both methods. The largest errors occur in the evaluation of the exchange term; a large number of *oooo* type integrals (all occupied orbitals) are involved in the evaluation of this term. It is possible (and seems likely) that systematic errors accumulate during the evaluation of this term. The accuracy of the 1C-CD-SAPT0 with the chosen tolerance is not quite as good as that of DF-SAPT0, but the errors incurred by 1C-CD-SAPT0 are acceptable given the smaller CD basis.

The number of basis functions needed for each complex in the S22 test set is shown in

Table 9. For the larger complexes in the set, a CD basis needs to be only about 75-80% of the DF basis size. Once the three-index AO quantities (DF three-index integrals or CD vectors, which we will refer to generically as three-index integrals) have been computed, the first step of the transformation to the MO basis scales as $\mathcal{O}(N_{aux}N_{ao}^2N_{mo})$. By reducing the size of the auxiliary basis, the expense of this step as well as the storage requirements for the three-index integrals are reduced. The next step is the formation of the four-index MO integrals from the transformed three-index integrals. Assuming that all the MO integrals are needed (and ignoring any sparsity), this step scales as $\mathcal{O}(N_{aux}N_{mo}^4)$. The most computational savings from a smaller fitting basis is possible in this step. Due to the fact that the formation of the Cholesky basis is much more intensive than the formation of the DF integrals, to be competitive with density fitting the Cholesky basis needs to be smaller than a DF basis that performs with comparable accuracy. As systems become larger, the work associated with the formation of the three-index CD or DF integrals becomes negligible compared to the rest of the computation. At some point, the formation of the MO four-index integrals from the three-index integrals should become much more time consuming, and beyond this point 1C-CD-SAPT may become significantly more efficient than DF-SAPT.

2.6.2 DF-SAPT: Intramonomer Corrections

The following was adapted from Ref. 94.

In the previous section on DF-SAPT0, the errors introduced through the DF approximation of the two-electron integrals was negligible.[93] In that work, we report errors of, at most, about 0.01 kcal mol^{-1} for any of the zeroth-order components of the interaction energy. Additionally, previous works on DF-SAPT(DFT) report only negligible errors created by the DF integrals.[27,86,170] More generally, we are not aware of any case reported in the literature where DF approximations created problematic errors for *interaction energies*. Here, we will report the DF error associated with the second-order corrections to electrostatics, exchange, and induction for 15 complexes selected from the S22 test set. The accuracy of the density fitting is assessed against conventional SAPT energies computed with the

Table 9: Number of auxiliary basis functions required for each complex in the S22 test set.[117a]

	DF-SAPT[b]	1C-CD-SAPT[c]
H-Bonded Complexes		
$(NH_3)_2$	282	244
$(H_2O)_2$	236	213
Formic Acid Dimer	524	465
Formamide Dimer	570	488
Uracil Dimer	1336	1073
2-Pyridoxine·2-Aminopyridine	1261	970
Adenine·Thymine WC	1621	1270
Dispersion Dominated Complexes		
$(CH_4)_2$	328	260
$(C_2H_4)_2$	472	375
Benzene·CH_4	734	549
PD Benzene Dimer	1140	848
Pyrazine Dimer	1048	810
Uracil Dimer	1336	1080
Stacked Indole·Benzene	1379	1016
Stacked Adenine·Thymine	1621	1265
Mixed Complexes		
Ethene·Ethine	426	349
Benzene·H_2O	688	540
Benzene·NH_3	711	545
Benzene·HCN	737	572
T-shaped Benzene Dimer	1140	847
T-shaped Indole Benzene	1379	1030
Phenol Dimer	1284	978

[a]Computations performed with the aug-cc-pVDZ basis. [b]The aug-cc-pVDZ-RI basis was used. [c]AO integrals computed with a tolerance of 10^{-3} E_h.

SAPT2008 program.[25] The results of this analysis are shown in Table 10. It is evident that DF errors are negligibly small for the second-order corrections to SAPT. The max error for the cases considered was less than 0.01 kcal mol^{-1} and the average errors are on the order of a few thousandths of a kcal mol^{-1} at the most. Due to the excellent performance of these second-order terms and the expense associated with the conventional dispersion corrections, we will not present results for the second-order dispersion corrections.

Table 10: Errors (in kcal mol^{-1}) of DF-SAPT2/aug-cc-pVDZ for selected complexes from the S22 test set (Ref. 117) relative to conventional SAPT2/aug-cc-pVDZ.[a]

	MAX[a]	MSE[b]	MUE[c]	RMS[d]
$E^{(12)}_{elst,resp}$	0.009	0.002	0.002	0.003
$E^{(11)}_{exch} + E^{(12)}_{exch}$	-0.001	0.000	0.000	0.000
$E^{(22)}_{ind,resp}$	0.002	0.000	0.000	0.001

[a]SAPT2 energies for 15 of the 22 complexes were computed with SAPT2008.[25] Density fitting errors for the components of DF-SAPT0 have been analyzed in Ref. 93. [b]Maximum absolute error. [c]Mean signed error. [d]Mean unsigned error. [e]Root mean square error.

2.7 Performance of DF-SAPT

2.7.1 DF/CD-SAPT0

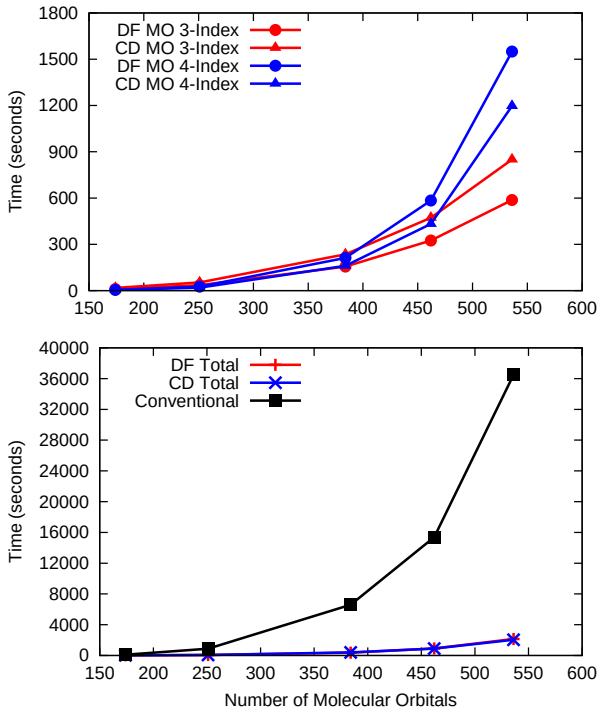
The following was adapted from Ref. 93.

We have performed timings of our SAPT program to understand the practical aspects of the DF and CD approximations in SAPT0. All the computations reported in this section were run on a quad-core Intel Xeon E5430 processor clocked at 2.66 GHz. We also compared the timings of our conventional integral transformation to the SAPT2008 program.[25] Both programs perform very similarly, so we will only show timings from our SAPT program.

DF- and CD-SAPT0 energies were computed for formic acid dimer, methane-benzene, T-shaped benzene dimer, T-shaped indole-benzene, and hydrogen bonded adenine-thymine. All of these complexes are taken from the S22 test set, and energies were computed with an aug-cc-pVDZ basis. The timings of the computation of the integrals required for the SAPT0 evaluation are shown in Figure 5. The "MO 3-index" timing refers to the formation of AO 3-index integrals (i.e., Equation 29 or 33-34) and their transformation to the MO basis. The "MO 4-index" timing refers to the formation of the approximate 4-index MO

integrals from the 3-index MO integrals. Performing a CD is more expensive than DF in the formation of AO 3-index quantities, but the reduction in size of the fitting basis recovers the overhead of the CD in the subsequent steps. The overall time for integral processing is nearly identical for the DF and CD approaches, with the CD approach becoming slightly more efficient beyond about 450 orbitals. Both CD and DF are much more efficient than the conventional integral transformation.

Figure 5: Timings of DF, CD, and conventional integral evaluation for SAPT0 computations on selected complexes from the S22 test set (Ref. 117) with an aug-cc-pVDZ basis. "4-index" timings refer to the construction of the 4-index integrals from the 3-index integrals.

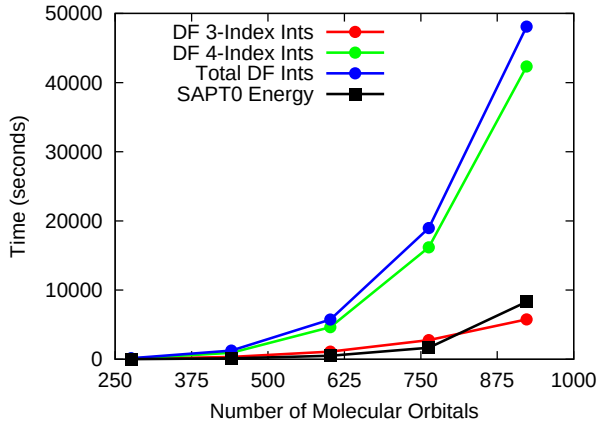


The tolerance chosen for the CD led to slightly larger errors in the SAPT energy components than in the DF approach; however, the errors remain no larger than a few hundredths of 1 kcal mol^{-1} (see Table 8). However, with the desire to study much larger systems and higher order corrections to the SAPT energy (which require additional groups of integrals such as the expensive *vvvv* type) the CD-SAPT approach could become significantly more efficient than DF-SAPT.

Figure 6 shows the timings for the DF-SAPT0 computations performed on the T-shaped

acene dimers. As the size of the DF-SAPT0 computations approaches 1000 MO's, formation of the four-index MO integrals dominates; the scaling of this step is $\mathcal{O}(N_{aux}o^2v^2)$. This step is far more costly than the energy evaluation. The most expensive portion of the SAPT0 energy evaluation is the $E^{(20)}_{exch-disp}$ term. This term conventionally scales as $\mathcal{O}(o^3v^2)$; Hesselmann and co-workers present equations in their DFT SAPT approach that use the DF representation of the two-electron integrals to evaluate the $E^{(20)}_{exch-disp}$ term in a manner that scales as $\mathcal{O}(N_{aux}o^2v^2)$.[86] Such a formulation of the $E^{(20)}_{exch-disp}$ term is more memory efficient, but slightly more costly than the conventional implementation. By reducing the size of N_{AUX} relative to DF, such terms will be more efficient using the CD approach. Our implementation of the $E^{(20)}_{exch-disp}$ term uses the conventional algorithm for smaller systems and Hesselmann's approach for larger systems. The remaining terms in SAPT0 scale as $\mathcal{O}(o^2v^2)$ or better.

Figure 6: Timings of DF-SAPT computations on T-shaped acene dimers with the aug-cc-pVDZ′ basis. "4-index" timings refer to the construction of the 4-index integrals from the 3-index integrals.



It is interesting to compare the scaling of SAPT0 to that of the supermolecular MP2 method. First, we will look at the scaling of SAPT0. As is evident from Figures 5 and 6, the construction of the 4-index MO integrals (whether by conventional transformation or by DF/CD approximations) is much more expensive than the energy evaluation (which scales as $\mathcal{O}(o^3v^2)$). There are several types of 4-index integrals required for SAPT0 that

are potentially similar in size: the *aarr*, *bbss*, *arar*, *bsbs* and *arbs* type integrals. It is possible to take advantage of the permutational symmetry of these integrals; the *arar* and *bsbs* type integrals have 2-fold symmetry and the *aarr* and *bbss* type integrals have 4-fold symmetry. The *arbs* type integrals do not have any permutational symmetry, which makes them potentially the most expensive to compute. It is important to remember that for SAPT computations, the occupied orbitals are divided into those from monomer A and those from monomer B. This makes a general comparison of the size of the various types of integrals impossible without knowing the relative sizes of the monomers. For simplicity, we will assume that the number of occupied orbitals on monomer A and monomer B are equal ($a = b$). Additionally we will assume that there are many more virtual orbitals than occupied orbital, since this is required for an accurate description of dispersion interactions ($a << r$ and $b << s$). With these assumptions, the asymptotic complexity of SAPT0 is $\mathcal{O}(aN_{ao}^4)$ for conventional SAPT0 and $\mathcal{O}(arbsN_{aux})$ for DF/CD SAPT0.

A counterpoise-corrected, supermolecular MP2 interaction energy requires three separate MP2 energy evaluations. The scaling of MP2 (much like SAPT0) is dominated by the formation of the 4-index MO integrals. Each MP2 computation requires the construction of an o^2v^2 group of integrals and then and energy evaluation, which scales as only $\mathcal{O}(o^2v^2)$. For the monomer computations, MP2 requires *arar* and *bsbs* integrals (and recall that counterpoise-corrected MP2 and SAPT0 both use the full dimer basis to describe each monomer). These integrals have a 2-fold symmetry and are identical to those found in SAPT0. The dimer MP2 computation uses a larger occupied space than any of the stages in SAPT0. Here, $o = a + b$ and $v = N_{MO} - a - b$. With this definition of the occupied and virtual space, the integrals needed for the dimer MP2 computation are o^2v^2 in size. This is noticeably larger than any of the types of integrals in SAPT0 despite the 2-fold symmetry of these integrals. Therefore, the asymptotic complexity of an MP2 interaction energy computation is $\mathcal{O}(oN_{ao}^4)$ for a conventional MP2 computation and $\mathcal{O}(o^2v^2N_{aux})$ for DF/CD MP2. Thus, the scaling of a supermolecular DF-MP2 interaction energy is worse than the scaling of DF/CD-SAPT0. In practice, however, each MO integral is only needed once to compute an MP2 energy, whereas each *ovov* integral is needed multiple times in SAPT0.

For large systems, where the *ovov* arrays do not fit into memory, it is likely that it would be faster to compute an MP2 interaction energy than to compute an SAPT0 interaction energy, despite the scalings.

2.7.2 Additional Improvements to DF-SAPT0

The following was adapted from Ref. 99.

Our previous implementation of DF-SAPT0 primarily used the DF approximation to avoid the two-electron integral AO to MO transformation.[93] Additionally, the DF integrals were used to factor one contribution to $E^{(20)}_{exch-disp}$ in which the *ovov* type integrals appear (this approach was described in Ref. 86). Despite this relatively simple approach, SAPT0 computations with as many as 116 atoms could be routinely performed with the program developed in Ref. 93. Unfortunately, this approach is not tractable for much larger systems. We will use the largest intercalator complex studied in this work as a concrete example of the deficiencies in our previous implementation. The solution of the CPHF equations in the MO basis involves iterative matrix-vector multiplies with a matrix that is $ov \times ov$. For the Pf·CGA complex, this matrix can become as large as 9.3 TB. The evaluation of the $E^{(20)}_{exch-disp}$ term could be written as a dot product between the dispersion amplitudes and a collection of integrals contracted with overlap integrals (see Equation 72). While this was done explicitly in our previous implementation, many of the contributions were evaluated as dot products between $ov \times ov$ matrices. The dispersion amplitudes can also become rather large, 650 GB or 1.3 TB for the Pf·CGA complex, depending on the use of the frozen core approximation. For the remaining exchange terms, even o^3v arrays can get as large as 250 GB.

Since our previous implementation was limited by the size of four-index arrays, the DF-SAPT0 algorithm developed in the present work minimizes the number of these arrays that must ever be formed. As a result, the new algorithm allows SAPT0 computations to be performed for much larger systems. Additionally, the factorization of the generalized two-electron integrals described above allows for increased efficiency with regard to memory usage; the evaluation of exchange terms can now be trivially blocked over the auxiliary

Figure 7: Timings of SAPT0/aug-cc-pVDZ′ computations on T-shaped acenes: benzene through pentacene. The "Old DF-SAPT0" implementation is described in Ref. 93; the "New DF-SAPT0" implementation is described in the present work. These computations were performed on dual quad-core Intel® Xeon® processors clocked at 2.66 GHz.

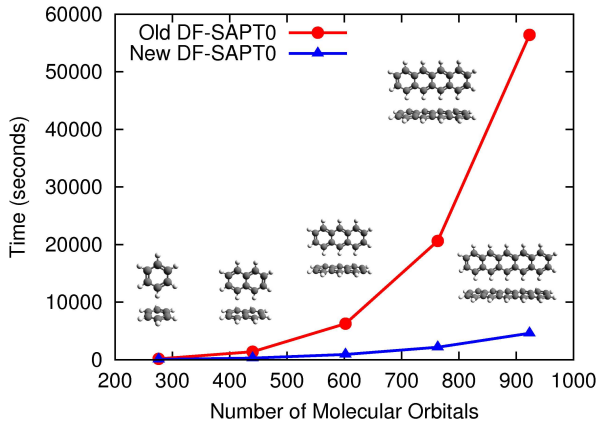
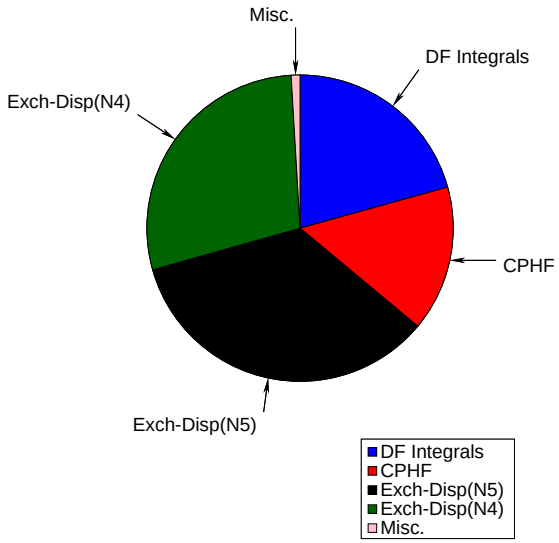


Figure 8: Timing of a SAPT0/aug-cc-pVDZ′ computation on the Pf·CGA complex with 220 atoms and 2800 basis functions. This computation took 61.7 hours running on dual quad-core Intel® Xeon® processors clocked at 2.66 GHz.



index. Possibly the greatest advantage of our new implementation of SAPT0 is that the improvements in terms of tractability and memory efficiency does not come at the cost of performance. Figure 7 shows the timings for a series of T-shaped acenes (described in

Ref. 93). For the largest system considered here, pentacene dimer, the new implementation of SAPT0 is over 12 times more efficient.

The advances described above have allowed the application of SAPT0 to large biological complexes. Here, we will highlight the Pf·CGA complex, which consists of 220 atoms and 2800 basis functions. This computation was performed using only modest computational resources: dual quad-core Intel® Xeon® processors clocked at 2.66, 40 GB of memory, and just over 2.5 days of wall time. Based on this performance, we estimate that our current DF-SAPT0 implementation should be scalable to 4000 basis functions. The timing of the computation on the Pf·CGA complex is shown in detail in Figure 8. The time of a SAPT0 computation is dominated by the evaluation of $E^{(20)}_{exch-disp}$, which takes nearly 2/3 of the total time. The evaluation of the DF integrals (including formation of the AO integrals, multiplication by $[J_{AB}]^{-1/2}$, and AO to MO transformation) takes roughly 1/5 of the time, with the bulk of the remainder being spent in the solution of the CPHF equations. To improve efficiency in terms with heavy disk I/O requirements (the DF integrals and CPHF evaluation), the disk I/O is done asynchronously. In the CPHF equations, for example, reading of the $C^P_{r_1 r_2}$ DF integrals is "hidden" under the contractions described in Equation 62. The remaining terms, $E^{(10)}_{elst,resp}$, $E^{(10)}_{exch}$, *etc.*, are trivial in comparison.

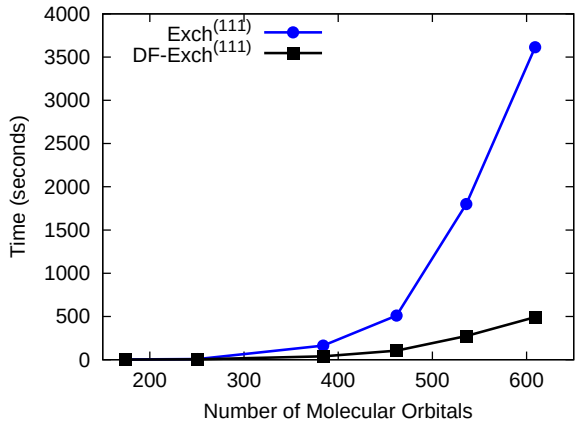
2.7.3 Higher-order SAPT

The following was adapted from Ref. 94.

We will begin by examining the improvement of the $E^{(111)}_{exch}$ evaluation due to the introduction of DF intermediates because it scales $\mathcal{O}(o^3v^3)$, which is as costly as any energy evaluation in SAPT2 (the formation of second-order double excitation amplitudes is slightly more expensive at $\mathcal{O}(o^2v^4)$). Additionally, due to the simple form of this correction, it is possible to compare the timings of the DF algorithm to a nearly optimal conventional algorithm. When DF intermediates are introduced, the scaling of the $E^{(111)}_{exch}$ term is reduced to $\mathcal{O}(o^2v^2N_{df})$ and depends on the formation of the Θ^P_{a1r1} and Θ^P_{b1s1} intermediates that exhibit $\mathcal{O}(o^2v^2N_{df})$ scaling. As illustrated by Figure 9, this is a significant improvement over the $\mathcal{O}(o^3v^3)$ conventional algorithm. These timings show that for systems with 600 molecular

orbitals, there is more than a factor of 7 speedup; this factor will continue to increase for larger systems.

Figure 9: Timings of the conventional and DF evaluation of $E_{exch}^{(111)}$ with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.



Our implementation of DF-SAPT has an advantage over conventional SAPT since it reduces the number of MO 4-index integrals that must be computed or stored. In addition to removing the need to store the v^4 integrals, there are contributions to the SAPT energy (evaluated as described above) from 6 types of ov^3 integrals. The $E_{elst,resp}^{(12)}$ term requires ar^3 and bs^3 type integrals; the $E_{exch}^{(12)}$ K_2^f term requires r^2bs and s^2ar type integrals. The ${}^tE_{ind}^{(22)}$ term and the singles contribution to $E_{disp}^{(22)}$ require all 4 of these ov^3 integrals. Through the use of DF intermediates these four types of ov^3 integrals do not need to be stored or even computed. Only two types of ov^3 integrals, which are needed for the $E_{exch}^{(12)}$ K_{11}^u term must be computed and stored on disk. Thus, an obvious advantage of a DF based algorithm is that only 2 of the 6 types of ov^3 integrals ever need to be computed. The $E_{elst,resp}^{(13)}$ term requires the ar^3 and bs^3 type integrals; while it is possible to avoid forming these integrals, it is not advantageous to do so. However, these ov^3 integrals are only needed once, so they do not need to be stored; they are computed from the three-index DF integrals (B^P). Another advantage of the DF based algorithm is that the 3 types of v^4 integrals do not need to be stored.

In order to understand how our DF-SAPT code performs relative to a conventional SAPT code, we compare the timings of our DF-SAPT code to the SAPT2008 program.[25] All of the timings reported in this work were run on a quad-core Intel® Xeon E5430 processor clocked at 2.66 GHz. Both programs were allocated a maximum of 16 GB of memory, which allows the conventional integral transformation in SAPT2008 to be performed "in core." The timings assume that the results of HF computations on the monomers are already available. In order to assure a fair comparison between the two programs, the following energy terms were included in the SAPT timing: $E_{elst}^{(10)}$, $E_{exch}^{(10)}$, $E_{ind,resp}^{(20)}$, $E_{exch-ind,resp}^{(20)}$, $E_{disp}^{(20)}$, $E_{exch-disp}^{(20)}$, $E_{elst,resp}^{(12)}$, $E_{exch}^{(11)}$, $E_{exch}^{(12)}$, ${}^{t}E_{ind}^{(22)}$, and $E_{disp}^{(21)}$. Additionally, the integral transformation is also included in the timing; for conventional SAPT this includes only the AO to MO transformation. For DF-SAPT this includes the formation of AO 3-index quantities, the AO to MO transformation of the 3-index quantities, and the formation of the necessary DF 4-index MO integrals from the 3-index MO integrals. SAPT2008 utilizes a CCSD program to compute the second-order double excitation amplitudes, while our program forms these quantities directly. The time spent in the CCSD program is included in the SAPT2008 timing. Although the CCSD amplitudes are not iterated until convergence, this may still be including some additional work in the SAPT2008 timing that is not included in the DF-SAPT timing; however, this time must be included in order to include the rate determining step into the SAPT timing.

At this point it should be noted that the timing differences between SAPT2008 and our DF-SAPT program are primarily a result of the introduction of DF integrals into the computation. Our conventional integral transformation and the energy terms which cannot benefit from DF perform similarly to those in SAPT2008. The timings of conventional SAPT and DF-SAPT are shown in Figure 10. At 350 orbitals, DF-SAPT is roughly a factor of 8 faster than the conventional SAPT. This speedup will grow for larger systems, as shown in Figure 9, since the overall scaling of certain terms is reduced. It is difficult to get meaningful timings past 350 orbitals because the conventional SAPT computations become I/O bound, therefore the timings would be strongly dependent on the available hardware.

Figure 10: Timings of the conventional and DF-SAPT2 computations of selected complexes from the S22 test set[117] with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.

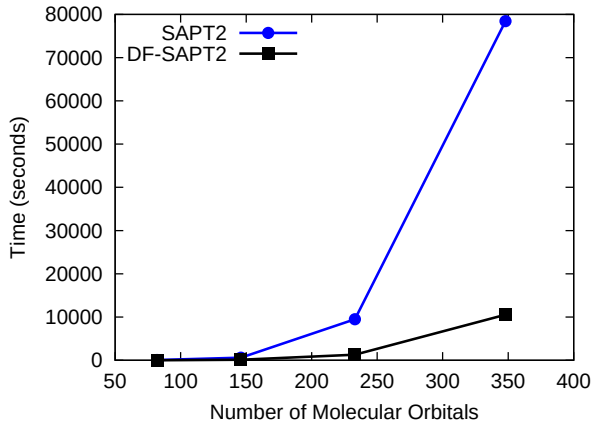
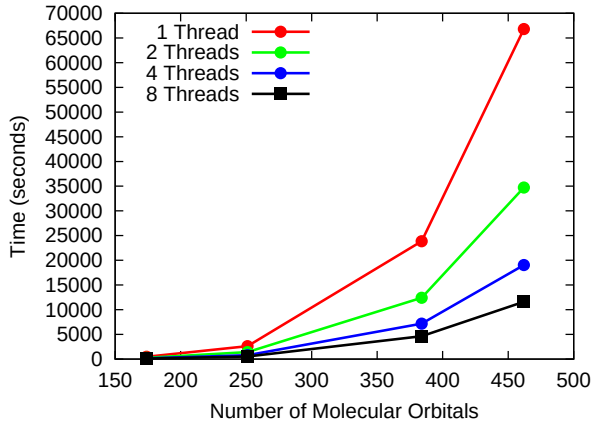


Figure 11: Timings of threaded DF-SAPT2 computations with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.



In practice, we get additional improvements due to the threading of the DF-SAPT energy evaluation. Since most modern computers are built with multi-core processors, our DF-SAPT code exploits the availability of these additional processors. The timings of our DF-SAPT with multiple threads are shown in Figure 11. Although to date we have only made trivial modifications to our code to allow for threading, even this minor effort has been useful for extending our code to larger systems. Since the most time consuming steps

in the DF-SAPT computation are cast as matrix multiplications, we use threaded Intel® MKL BLAS routines. Other parts of our program that could not be cast as BLAS routines, such as the 3-index integral evaluation, are threaded using OpenMP. The timings shown in Figure 11 were run on dual quad-core Intel® Xeon E5430 processors clocked at 2.66 GHz. For the T-shaped indole-benzene, the SAPT/aug-cc-pVDZ (462 orbitals) computations get a factor of 1.92 speedup from 1 to 2 threads. The efficiency degrades slightly from 2 to 4 threads with only an additional factor of 1.82 speedup. When the both processors are fully utilized, we only see a factor of 1.64 moving from 4 to 8 threads. Nevertheless, these speedups are encouraging given how straightforward they were to achieve. We intend to pursue larger-scale parallelization in future work.

2.7.4 Improvements from Natural Orbitals

The following was adapted from Ref. 95.

To evaluate this correction more efficiently, we use threaded Intel® MKL BLAS routines to form the triples amplitudes in Equation 153; the energy evaluation is threaded using OpenMP. The timings shown in Figure 12 were run on dual quad-core Intel® Xeon E5430 processors clocked at 2.66 GHz with 8 threads. The largest computation shown in this figure corresponds to hydrogen bonded adenine-thymine. Run with 1 thread, without any approximations (other than the DF integrals), this computation would take roughly 2 months. Our threaded code reduces this to 9 days, and with the frozen core approximation the computation takes a more manageable 3 days. When the virtual space is reduced, $E_{disp}^{(22)}(\mathrm{T})$ can be computed in less than 4 hours. For the systems considered, the combination of the frozen core and MP2 NO approximations result in a remarkable 50-60X speedup.

It is possible to realize even greater speedups for larger basis sets (to this point, we have only shown results for the modest aug-cc-pVDZ basis). For larger basis sets, truncation with a certain occupation threshold will remove a larger fraction of the virtual orbitals. The speedups for the $E_{disp}^{(22)}(\mathrm{T})$ evaluation with the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ bases are shown in Figure 13. Due to the expense of triples corrections in an aug-cc-pVQZ basis, we only show results for three small dimers from the S22 test set. The

Figure 12: Timings of the evaluation of $E_{disp}^{(22)}$(T) correction with the aug-cc-pVDZ basis set as various approximations are applied for (from left to right) formic acid dimer, methane-benzene, T-shaped benzene dimer, T-shaped indole-benzene, and hydrogen bonded adenine-thymine.

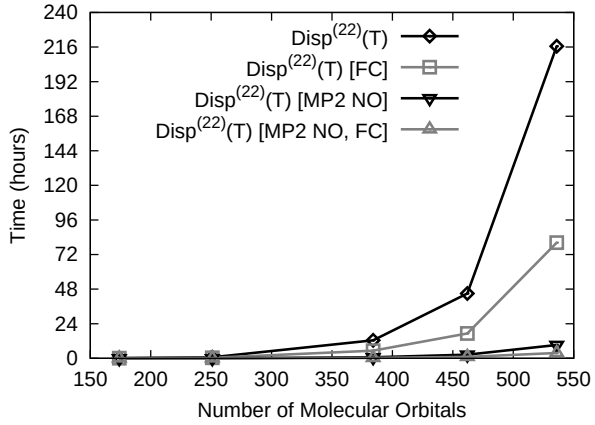
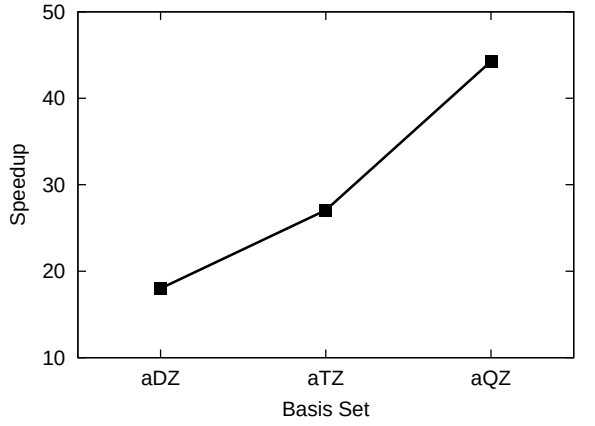


Figure 13: Average speedup for the $E_{disp}^{(22)}$(T) correction of water dimer, ammonia dimer, and methane dimer computed with various basis sets when virtual MP2 NOs are removed using a threshold of 10^{-6} electrons. An additional factor of ∼2X can be gained by the frozen core approximation.



average speedups that result from the truncation of MP2 NOs increase with the size of the basis set to an impressive 45X with the aug-cc-pVQZ basis. When the truncated virtual space is combined with the frozen core approximation, the overall speedup increases to 85X with the aug-cc-pVQZ basis. In our limited test cases, for a particular system, the amount

of error introduced by the deletion of NOs with occupation numbers smaller than 10^{-6} remains similar as one goes to larger basis sets.

2.8 Accuracy of SAPT

The following was adapted from Ref. 96.

The SAPT methods described above have been applied to the S22 test set of Hobza and co-workers[117] (as revised in the S22A benchmarks[213]) in order to gauge their accuracy and basis set requirements. The complexes of the S22 are grouped by interaction type: electrostatics-dominated, dispersion-dominated, and mixed-influence. These groupings were initially assigned intuitively; recently, they have been revised using SAPT(DFT) decompositions.[65] In this more rigorous approach, complexes are considered electrostatics-dominated if $E_{electrostatic}$ is more than twice $E_{dispersion}$ and dispersion-dominated if $E_{dispersion}$ is more than twice $E_{electrostatic}$. Mixed-influence complexes are those that are not either electrostatic- or dispersion-dominated. The original, intuitive groupings were found to be in good agreement with the SAPT(DFT) decompositions. The stacked uracil dimer and stacked adenine-thymine complex were found to be mixed-influence rather than dispersion-dominated; the T-shaped benzene dimer was found to be dispersion-dominated rather than mixed-influence. Our SAPT decompositions agree with the SAPT(DFT) decompositions regarding the classification of S22 complexes with the exception of benzene-HCN. Originally, this complex was considered to be of mixed-influence. The SAPT(DFT) decompositions suggest that it should be considered electrostatics-dominated; however, at the SAPT2+3/aug-cc-pVTZ level, $E_{electrostatic} \approx E_{dispersion}$ (-3.84 and -3.72 kcal mol^{-1}, respectively). Therefore, the original designation of the benzene-HCN interaction as mixed-influence is more appropriate.

In Tables 11-14, we report the performance of SAPT for the entire S22 test set and its subsets. The methods considered are defined by Equations 14-15 and 17-19. It has been suggested that explicit third-order induction terms are sufficient in certain complexes, making the inclusion of $\delta E_{HF}^{(3)}$ unnecessary.[162] To test this hypothesis, we report SAPT2+3

energies with and without the $\delta E^{(3)}_{HF}$ term. A truncated aug-cc-pVDZ basis, denoted aug-cc-pVDZ′, removes all diffuse functions from hydrogen atoms and the diffuse d functions on non-hydrogen atoms. This basis is used to encourage error cancellation for methods with a MP2-like account of dispersion, namely SAPT0 and SAPT2. It is not reasonable to pair higher-order SAPT methods with this basis set, so we do not report such results here.

Table 11: Accuracy of various SAPT methods for the entire S22 test set[117] relative to the estimated CCSD(T) CBS limit interaction energies of Reference 213.[a]

	MSE	MSRE	MUE	MURE	MAX	MAXR[b]
aug-cc-pVDZ′						
SAPT0	-0.19	5.10	0.47	11.74	-1.73	67.67
SAPT2	1.23	20.13	1.32	22.68	4.05	73.69
aug-cc-pVDZ						
SAPT0	-1.43	-20.50	1.45	22.03	-3.81	-69.19
SAPT2	0.06	-3.72	0.94	16.53	-2.62	-67.61
SAPT2+	0.03	1.41	0.32	6.05	-1.22	13.29
SAPT2+(3)	0.32	7.42	0.39	8.05	0.82	20.30
SAPT2+3 - $\delta E^{(3)}_{HF}$	1.19	16.42	1.24	16.90	4.61	29.90
SAPT2+3	0.04	2.96	0.27	5.77	-1.32	16.20
aug-cc-pVTZ[c]						
SAPT0	-1.74	-29.26	1.74	29.26	-5.12	-87.23
SAPT2	-0.63	-17.33	0.78	19.37	-3.43	-90.48
SAPT2+	-0.60	-10.71	0.60	10.71	-1.55	-27.71
SAPT2+(3)	-0.25	-2.58	0.28	3.48	-1.06	-10.13
SAPT2+3 - $\delta E^{(3)}_{HF}$	0.54	6.00	0.66	7.42	2.72	16.06
SAPT2+3	-0.55	-8.16	0.56	8.33	-1.86	-17.94

[a]Errors given in kcal mol^{-1}; relative errors given as percentages. [b]From left to right, mean signed error, mean signed relative error, mean unsigned error, mean unsigned relative error, max error, and max relative error. Signed errors are reported as: $E_{SAPT} - E_{ref}$. [c]The aug-cc-pVTZ computations exclude the two adenine-thymine dimers.

As we have previously documented,[93] the simplest SAPT method, SAPT0, performs remarkably well when paired with the aug-cc-pVDZ′ basis. For studies involving large systems or requiring large numbers of computations, SAPT0/aug-cc-pVDZ′ can be a practical solution. However, its performance depends entirely on error cancellation within the dispersion term. For most of the complexes in the S22 set, the SAPT0 treatment of the dispersion energy (in a large basis) is a poor approximation to the exact dispersion energy (overestimating it by as much as a factor of two in some cases). It is well known that dispersion terms converge slowly with respect to basis set size; by using a very small basis, one can hope to underestimate the complete basis SAPT0 dispersion energy by roughly the same amount that it overestimates the exact dispersion energy. Unfortunately, this error cancellation is not guaranteed to occur. For S22A, the most notable example is the methane dimer.

Table 12: Accuracy of various SAPT methods for the electrostatics-dominated subset of the S22 test set[117] relative to the estimated CCSD(T) CBS limit interaction energies of Reference 213.[a]

	MSE	MSRE	MUE	MURE	MAX	MAXR[b]
			aug-cc-pVDZ′			
SAPT0	-0.68	-2.43	0.77	5.44	-1.73	10.52
SAPT2	2.61	18.84	2.61	18.84	4.05	22.41
			aug-cc-pVDZ			
SAPT0	-1.98	-11.38	2.00	11.94	-3.81	-20.27
SAPT2	1.39	11.13	1.39	11.13	2.12	16.71
SAPT2+	0.41	4.02	0.42	4.12	0.86	9.17
SAPT2+(3)	0.67	6.74	0.67	6.74	0.82	13.83
SAPT2+3 - $\delta E^{(3)}_{HF}$	2.89	20.87	2.89	20.87	4.61	24.50
SAPT2+3	0.31	4.29	0.31	4.29	0.56	11.81
			aug-cc-pVTZ[c]			
SAPT0	-2.69	-16.87	2.69	16.87	-5.12	-27.22
SAPT2	0.19	2.20	0.30	2.84	0.62	7.87
SAPT2+	-0.77	-5.10	0.77	5.10	-1.52	-8.70
SAPT2+(3)	-0.42	-1.42	0.52	3.63	-1.06	-5.63
SAPT2+3 - $\delta E^{(3)}_{HF}$	1.73	12.36	1.73	12.36	2.72	14.44
SAPT2+3	-0.85	-4.53	0.88	5.08	-1.86	-9.88

[a]Errors given in kcal mol^{-1}; relative errors given as percentages. Here, we use the definition of electrostatics-dominated from Reference 117. [b]From left to right, mean signed error, mean signed relative error, mean unsigned error, mean unsigned relative error, max error, and max relative error. Signed errors are reported as: $E_{SAPT} - E_{ref}$. [c]The aug-cc-pVTZ computations exclude hydrogen bonded adenine-thymine.

Table 13: Accuracy of various SAPT methods for the dispersion-dominated subset of the S22 test set[117] relative to the estimated CCSD(T) CBS limit interaction energies of Reference 213.[a]

	MSE	MSRE	MUE	MURE	MAX	MAXR[b]
			aug-cc-pVDZ′			
SAPT0	0.03	15.59	0.38	24.71	0.71	67.67
SAPT2	0.15	21.79	0.44	29.80	0.91	73.69
			aug-cc-pVDZ			
SAPT0	-1.09	-30.62	1.12	34.88	-2.83	-69.19
SAPT2	-0.90	-21.88	1.01	32.80	-2.62	-67.61
SAPT2+	-0.13	-0.92	0.22	8.48	-0.59	13.29
SAPT2+(3)	0.18	9.69	0.18	9.69	0.30	20.30
SAPT2+3 - $\delta E^{(3)}_{HF}$	0.38	16.79	0.38	16.79	0.61	26.06
SAPT2+3	-0.01	3.44	0.14	6.97	-0.25	16.20
			aug-cc-pVTZ			
SAPT0	-1.43	-44.23	1.43	44.23	-3.49	-87.23
SAPT2	-1.37	-40.22	1.37	41.73	-3.43	-90.48
SAPT2+	-0.52	-17.20	0.52	17.20	-1.25	-27.71
SAPT2+(3)	-0.11	-2.86	0.11	3.28	-0.27	-7.33
SAPT2+3 - $\delta E^{(3)}_{HF}$	0.07	3.50	0.10	4.30	0.26	10.05
SAPT2+3	-0.35	-10.76	0.35	10.76	-0.76	-17.94

[a]Errors given in kcal mol^{-1}; relative errors given as percentages. Here, we use the definition of dispersion-dominated from Reference 65. [b]From left to right, mean signed error, mean signed relative error, mean unsigned error, mean unsigned relative error, max error, and max relative error. Signed errors are reported as: $E_{SAPT} - E_{ref}$.

Table 14: Accuracy of various SAPT methods for the mixed-character subset of the S22 test set[117] relative to the estimated CCSD(T) CBS limit interaction energies of Reference 213.[a]

	MSE	MSRE	MUE	MURE	MAX	MAXR[b]
			aug-cc-pVDZ′			
SAPT0	0.05	2.51	0.28	5.92	0.57	12.54
SAPT2	0.96	19.81	0.96	19.81	1.85	31.16
			aug-cc-pVDZ			
SAPT0	-1.26	-19.62	1.26	19.62	-3.71	-31.83
SAPT2	-0.26	-0.82	0.47	7.02	-1.85	-15.86
SAPT2+	-0.15	1.15	0.33	5.61	-1.22	-10.44
SAPT2+(3)	0.14	6.02	0.33	7.75	-0.56	14.53
SAPT2+3 - $\delta E_{HF}^{(3)}$	0.40	12.21	0.54	13.53	1.08	29.90
SAPT2+3	-0.14	1.37	0.36	6.02	-1.32	-11.29
			aug-cc-pVTZ[c]			
SAPT0	-1.23	-24.91	1.23	24.91	-2.66	-33.20
SAPT2	-0.60	-11.18	0.60	11.18	-1.52	-19.87
SAPT2+	-0.53	-9.03	0.53	9.03	-1.55	-15.92
SAPT2+(3)	-0.24	-3.29	0.24	3.54	-0.99	-10.13
SAPT2+3 - $\delta E_{HF}^{(3)}$	-0.01	3.03	0.31	6.30	-1.11	16.06
SAPT2+3	-0.50	-8.68	0.50	8.68	-1.57	-16.10

[a]Errors given in kcal mol^{-1}; relative errors given as percentages. Here, we use the definition of mixed-character from Reference 65 with the exception of the benzene-HCN complex, which we include here. [b]From left to right, mean signed error, mean signed relative error, mean unsigned error, mean unsigned relative error, max error, and max relative error. Signed errors are reported as: $E_{SAPT} - E_{ref}$. [c]The aug-cc-pVTZ computations exclude stacked adenine-thymine.

Here, SAPT0/aug-cc-pVDZ′ underestimates the (admittedly small) interaction energy by almost 70%; SAPT0/aug-cc-pVDZ underestimates the interaction energy by 12%. With the aug-cc-pVTZ basis, SAPT0 will accurately compute the interaction energy of methane dimer (0.4% error). These trends for methane dimer should generalize to any alkane-alkane interaction; therefore, SAPT0/aug-cc-pVDZ′ is a very poor choice to describe these interactions. We should also note that the S22 test set contains quite a few π interactions (perhaps overemphasizing their importance compared to a typical study of noncovalent interactions). Thus, the good performance of SAPT0/aug-cc-pVDZ′ for the S22 molecules may not be indicative of its general applicability.

One of the most commonly applied many-body SAPT methods is SAPT2, which includes terms to second-order with respect to electron correlation. The performance of this method tends to be very similar to MP2. For electrostatics-dominated complexes, the performance of SAPT2/aug-cc-pVTZ is the best of any level of SAPT tested. SAPT2/aug-cc-pVDZ′

performs fairly well for dispersion dominated complexes, but very poorly for electrostatics-dominated complexes. Across the entire S22 test set, the performance of SAPT2 is fairly poor; it can only be recommended for application to electrostatics-dominated complexes with an aug-cc-pVTZ basis.

The higher-order methods SAPT2+ and SAPT2+(3) differ only by the $E^{(13)}_{elst,resp}$ and $E^{(30)}_{disp}$ terms that are included in the latter. Both of these methods are reliable for any interaction type considered when paired with an aug-cc-pVDZ basis; it should be noted that with this basis, SAPT2+(3) will be, on average, underbound relative to the CCSD(T) benchmark, whereas SAPT2+ is neither consistently underbound nor overbound. It follows that increasing the size of the basis to aug-cc-pVTZ leads to an improvement for SAPT2+(3) and a worsening for SAPT2+. Since the additional terms in SAPT2+(3) scale $\mathcal{O}(N^6)$ while, overall, the method scales $\mathcal{O}(N^7)$, and given the similar accuracy of SAPT2+ and SAPT2+(3) in an aug-cc-pVDZ basis, there does not seem to be a compelling reason to use SAPT2+ instead of the more complete SAPT2+(3). Perhaps the most accurate level of SAPT tested was SAPT2+(3)/aug-cc-pVTZ. This level of SAPT is, on average, slightly overbound, but within 1 kcal mol^{-1} of the CCSD(T) benchmark. The performance of this method is based around the error cancellation that occurs by including the $E^{(30)}_{disp}$ term without its exchange counterpart. The $E^{(30)}_{disp}$ term is usually small (less than 0.6 kcal mol^{-1} in all cases considered) and repulsive. The $E^{(13)}_{elst,resp}$ term is also small and usually repulsive. The addition of these terms to SAPT2+, which tends to overbind, results in a method that, fortuitously, predicts accurate interaction energies.

The SAPT2+3 method includes a proper and complete description of third-order interactions. Despite including several additional terms, the cost of SAPT2+3 is nearly identical to SAPT2+(3). One of the most robust levels of SAPT that we tested is SAPT2+3/aug-cc-pVDZ. This level is equally well suited for studying electrostatic- and dispersion-dominated complexes. With this small basis, the $\delta E^{(3)}_{HF}$ term should be included in all cases as it compensates for the basis set incompleteness in the dispersion term. With the larger, aug-cc-pVTZ basis, $\delta E^{(3)}_{HF}$ should only be included for electrostatics-dominated complexes, where the higher-order induction terms it captures become important. For dispersion-dominated

and mixed-influence complexes, it appears that the induction series is sufficiently converged at third-order and $\delta E_{HF}^{(3)}$ should be omitted. This finding is in agreement with the work of Patkowski *et al.*[162] It is possible that accounting for the effect of orbital response on the third-order induction, which is not considered here, would improve the performance of SAPT2+3 for the mixed-influence interactions.[163]

The levels of SAPT described above contain two different approximations to the exact second-order dispersion energy, $\epsilon_{disp}^{(2)}$. The $E_{disp}^{(20)}$ term is the simplest approximation to $\epsilon_{disp}^{(2)}$ and is found in the SAPT0 and SAPT2 methods. This description of dispersion is similar to that of MP2. The treatment of dispersion in SAPT2+ and higher SAPT methods is similar to that of MP4, which we will denote $\epsilon_{disp}^{(2)}[2]$, and sums the intramonomer corrections to $E_{disp}^{(20)}$ through second order:

$$\epsilon_{disp}^{(2)}[2] = E_{disp}^{(20)} + E_{disp}^{(21)} + E_{disp}^{(22)}. \tag{210}$$

At times, finite-order perturbation theory is insufficient to accurately compute dispersion energies. It is also possible to evaluate the second-order dispersion energy using a coupled-cluster based approach. The variant we will discuss was developed by Williams *et al.*[247] and is based on a CCD treatment of dispersion and the intramonomer correlation corrections to dispersion. This is augmented by a perturbative treatment of singles and triples. We refer to this treatment of dispersion as:

$$\epsilon_{disp}^{(2)}[\mathrm{CCD} + \mathrm{ST(CCD)}] = \mathrm{E}_{\mathrm{disp}}^{(2)}[\mathrm{CCD}] + \mathrm{E}_{\mathrm{disp}}^{(2)}[\mathrm{S(CCD)}] + \mathrm{E}_{\mathrm{disp}}^{(2)}[\mathrm{T(CCD)}]. \tag{211}$$

When we replace the $\epsilon_{disp}^{(2)}[2]$ treatment of dispersion with $\epsilon_{disp}^{(2)}[\mathrm{CCD} + \mathrm{ST(CCD)}]$ in the SAPT2+3 method, we will denote it as SAPT2+3(CCD).

In figures 14-16, SAPT is compared to CCSD(T) within the same basis, in contrast to the analysis of the S22A test set, where we compared to estimated complete basis CCSD(T) results. Here, we want to explore the accuracy of different treatments of dispersion without examining any error cancellation that occurs due to basis set incompleteness. The first case we will consider is the methane dimer (Figure 14). The dispersion interactions in the methane dimer are easily described, and all three SAPT methods are in good agreement with CCSD(T). Dispersion interactions involving delocalized π-orbitals are known to be

Figure 14: Methane dimer potential energy curves computed with various levels of SAPT/aug-cc-pVQZ (defined in Equations 14 and 19) and with CCSD(T)/aug-cc-pVQZ. The SAPT2+3 and SAPT2+3(CCD) curves are nearly coincident.

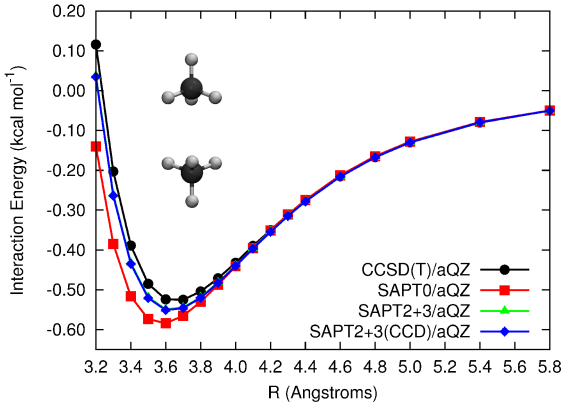
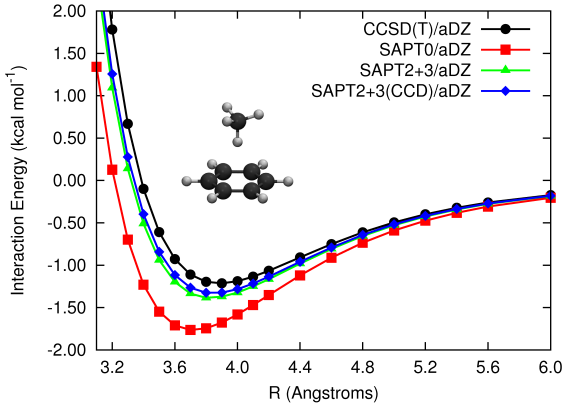
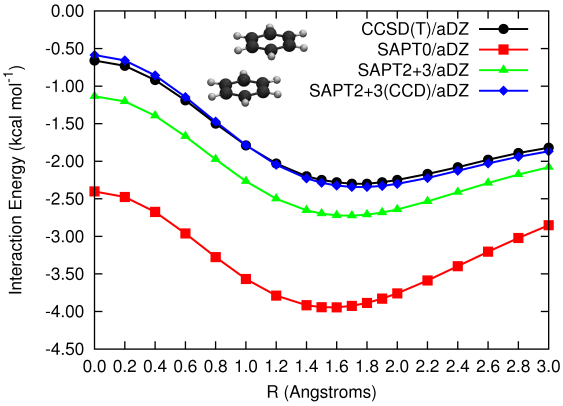


Figure 15: Methane-benzene potential curves computed with various levels of SAPT/aug-cc-pVDZ (defined in Equations 14 and 19) and with CCSD(T)/aug-cc-pVDZ.



more difficult to accurately compute. The methane-benzene complex is an example of such an interaction; the results for this system are shown in Figure 15. Here, the $E_{disp}^{(20)}$ term alone cannot accurately predict the dispersion in the methane-benzene complex. Some account of intramonomer correlation must be included to obtain a reliable dispersion energy for this system; either the $\epsilon_{disp}^{(2)}[2]$ or $\epsilon_{disp}^{(2)}[\text{CCD} + \text{ST(CCD)}]$ treatment of dispersion is sufficient. Dispersion interactions in the benzene dimer involve delocalized π-orbitals on both monomers and require a high degree of electron correlation to predict quantitatively. The SAPT results for the parallel-displaced benzene dimer are shown in Figure 16. For the

Figure 16: Parallel-displaced benzene dimer potential curves computed with various levels of SAPT/aug-cc-pVDZ (defined in Equations 14 and 19) and with CCSD(T)/aug-cc-pVDZ at a vertical separation of 3.6 Å. Although SAPT0 compares poorly to CCSD(T) in this case for a fixed basis set, SAPT0/aug-cc-pVDZ′ compares reasonably well with CCSD(T)/CBS estimates due to favorable error cancellation (see Ref. 94).



benzene dimer, the $E_{disp}^{(20)}$ term overestimates dispersion by roughly 2 kcal mol^{-1} for all the horizontal displacements considered. The $\epsilon_{disp}^{(2)}[2]$ treatment of dispersion is also insufficient in this case; although it is a significant improvement, there are still errors on the order of 0.5 kcal mol^{-1}. These problems can be addressed with the $\epsilon_{disp}^{(2)}[\mathrm{CCD} + \mathrm{ST(CCD)}]$ account of dispersion and leads to remarkable agreement with the CCSD(T) results.

These test cases demonstrate that dispersion interactions cannot all be treated with the same approximations. For simple alkane-alkane interactions, the $E_{disp}^{(20)}$ term, without any account of intramonomer electron correlation, is sufficient to accurately compute the dispersion energy. Other, more challenging interactions, such as methane–benzene, are reliably treated by intermediate descriptions of dispersion (*i.e.*, $\epsilon_{disp}^{(2)}[2]$). For the notoriously difficult π-π interactions, a rigorous $\epsilon_{disp}^{(2)}[\mathrm{CCD} + \mathrm{ST(CCD)}]$ treatment of dispersion is required if error cancellation is not to be relied upon. By applying DF and NO approximations, accurate dispersion energies can be computed with the many-body formulation of SAPT for much larger systems than were previously accessible.

CHAPTER III

APPLICATIONS OF SYMMETRY-ADAPTED PERTURBATION THEORY

3.1 Heteroatom effects on π-π interactions

The following was adapted from Ref. 92.

3.1.1 Introduction

Bimolecular complexes involving pyridine have been studied theoretically by several groups. Some of the first work studying the pyridine dimer was conducted by Megiel *et al.*;[140] the dependence of the chemical shift of the pyridine nitrogen was studied as a function of pyridine concentration in *n*-heptane, and hydrogen bonded configurations of the pyridine dimer were examined with Hartree-Fock and density functional theory (DFT). Optimized geometries and binding energies of the pyridine dimer were computed by Piacenza and Grimme using dispersion corrected density functional theory (DFT-D), second-order Møller-Plesset perturbation theory (MP2), and spin-component-scaled MP2 methods.[164] Binding energies for pyridine dimers and trimers were also computed by Mishra and Sathyamurthy at the MP2/6-311++G** level of theory.[146] Geerlings and co-workers studied complexes of pyridine, pyrimidine, and imidazole with substituted benzenes at the MP2/6-31G*(0.25) level of theory to examine the effect of substitution on binding energies and the H-bonding ability of the nitrogen lone pairs.[145] Tsuzuki *et al.*[226] have examined benzene-pyridine at a high level of theory as part of a study on interactions between benzene and pyridinium cations. However, relatively little work has sought to systematically explore the fundamental question of how heteroatoms affect π-π interactions. As this article was being prepared, two additional relevant studies were published. Tschumper and co-workers have examined complexes involving benzene, 1,3,5-triazine, cyanogen, and diacetylene.[13] Wang and Hobza[234] have presented high-quality interaction energies for selected configurations of benzene with isoelectronic nitrogen-containing heterocycles.

In this study, quantum chemical methods are used to compute benchmark quality binding energies and potential energy curves for benzene-pyridine and the pyridine dimer. Previous works have presented fully optimized geometries for these complexes,[164,226,234] which are important for understanding their spectroscopy. However, our present interest is not the spectroscopy of these clusters, but rather how heteroatoms tune π-π interactions across the energy landscape. Such information can be valuable because heteroatom-containing π-π interactions may occur in a wide variety of geometries in complex systems such as biopolymers. Because full six-dimensional intermolecular potential surfaces are difficult to visualize and compute, our strategy here is to compare several interesting configurations of the pyridine dimer to corresponding configurations in benzene-pyridine and the benzene dimer to ascertain the heteroatom effect. In addition, we plot selected potential energy curves.

To better understand the nature of heteroatom-influenced π-π interactions, it is also useful to analyze the interaction energy in terms of electrostatic, induction, dispersion, and exchange-repulsion components. Previous work on benzene-pyridine has considered such an analysis using approximate energy decomposition schemes.[226] Here we employ the rigorous symmetry-adapted perturbation theory (SAPT)[111,246] to analyze the sandwich, T-shaped, and parallel-displaced configurations of both benzene-pyridine and pyridine dimer, and we find that SAPT leads to somewhat different conclusions than previous analyses.

3.1.2 Theoretical Methods

Single-point energy computations were performed using second order perturbation theory (MP2) as well as coupled-cluster with singles and doubles including perturbative triples [CCSD(T)].[176] These methods were used with Dunning's aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.[49,118] In addition, the spin-component-scaled second-order perturbation theory (SCS-MP2) method of Grimme was used to analyze parallel-displaced configurations.[67] Within each complex, the monomer geometries were held rigid as the intermonomer distance was varied. Experimental geometries for each monomer were used. The benzene monomer geometry is that recommended by Gauss and Stanton: $r_{CC} = 1.3915$

Å and r_{CH} = 1.0800 Å.[59] The monomer geometry of pyridine used in this study is that reported by Innes *et al.*[103] The sandwich and T-shaped benzene dimer curves were obtained in a separate work.[198]

In order to correct for basis set superposition error, the Boys-Bernardi counterpoise correction scheme was employed for all energy computations.[21] Large-basis CCSD(T) results are estimated using an additive scheme which adds a "coupled-cluster correction," $\Delta\text{CCSD(T)} = E_{\text{CCSD(T)}}^{small-basis} - E_{\text{MP2}}^{small-basis}$, to a large-basis MP2 result: $E_{\text{CCSD(T)}}^{large-basis} \approx E_{\text{MP2}}^{large-basis} + \Delta\text{CCSD(T)}$. Previous work suggests that this correction is fairly well converged with the aug-cc-pVDZ basis set,[203] although more recent work suggests that the size of the correction might grow slightly if larger basis sets could be employed.[107]

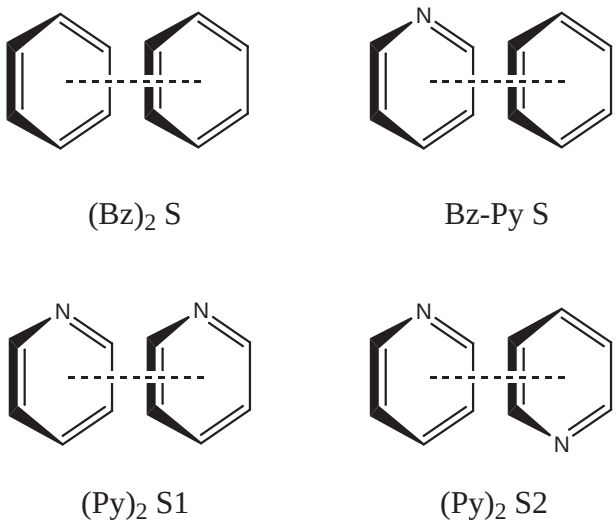
Dunning's correlation consistent basis sets have been shown to systematically approach the complete-basis-set (CBS) limit; this was exploited to obtain estimates of the MP2/CBS binding energies by using the two point extrapolation scheme of Halkier *et al.*[78] with aug-cc-pVTZ and aug-cc-pVQZ basis sets. All CCSD(T) and MP2 computations were performed with the core electrons frozen using the PSI 3.3 and MOLPRO programs.[41,151] A natural population analysis of benzene and pyridine for Hartree-Fock/6-311++G** wavefunction was performed using Jaguar.[106]

Energy component analysis is performed using symmetry adapted perturbation theory (SAPT).[111,246] All terms through second-order with respect to electron correlation are included in this work, thus designating this truncation of SAPT theory SAPT2. All SAPT2 computations were performed using the SAPT2006 program.[25] These computations were performed using the aug-cc-pVDZ′ basis set, which consists of the cc-pVDZ basis set with the diffuse s and p functions of aug-cc-pVDZ added to non-hydrogen atoms. In our experience, the SAPT2/aug-cc-pVDZ′ results are good approximations to large-basis CCSD(T) results through a favorable cancellation of errors.[203]

The configurations of benzene dimer, benzene-pyridine, and pyridine dimer studied here fall into three categories: sandwich (Figure 17), T-shaped (Figure 18) and parallel-displaced (Figure 19). For all configurations, the monomers were aligned based on the geometric centers of their rings. The vertical inter-monomer separation in all cases was measured from

these centers and is denoted as R. For the parallel-displaced configurations, the horizontal displacement is labeled H. The pyridine dimer has the most possible unique configurations, and because of this, the focus of this work was on the configurations that represented extremes for pyridine dimer: placing the nitrogen atoms as close and as far away from one another as possible, and aligning the dipole moments in parallel and anti-parallel arrangements. The analogous configurations for benzene-pyridine were also studied. Each of these configurations were compared to a similar configuration of benzene dimer. For convenience in the following discussion, we will frequently abbreviate pyridine as Py and benzene as Bz.

Figure 17: Sandwich and T-shaped configurations of benzene dimer, benzene-pyridine, and pyridine dimer.



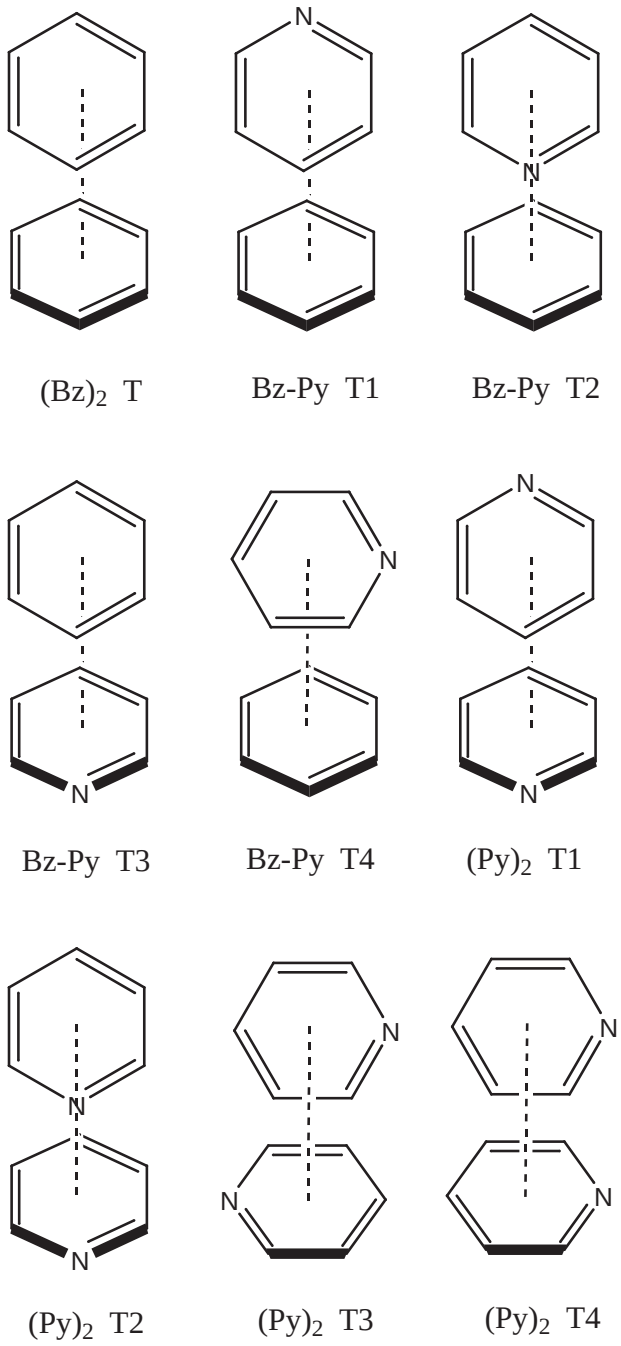
$(Bz)_2$ S Bz-Py S

$(Py)_2$ S1 $(Py)_2$ S2

Potential energy curves were computed for sandwich configurations of benzene-pyridine and pyridine dimer at the estimated CCSD(T) complete basis set limit. These computations show $(Py)_2$ S2 to be the most favorable sandwich configuration, followed by Bz-Py S (see Table 15). With a binding energy of 2.95 kcal mol^{-1}, $(Py)_2$ S2 binds nearly twice as strongly as $(Bz)_2$ S. The least favorable configuration, and the only configuration to be less favorable than the benzene dimer sandwich configuration, is $(Py)_2$ S1.

Five of the more favorable T-shaped configurations of $(Py)_2$ or Bz-Py were analyzed at the estimated CCSD(T) CBS limit, and potential curves for the remaining T-shaped configurations from Figure 18 were computed at the estimated CCSD(T)/aug-cc-pVTZ

Figure 18: Sandwich and T-shaped configurations of benzene dimer, benzene-pyridine, and pyridine dimer.



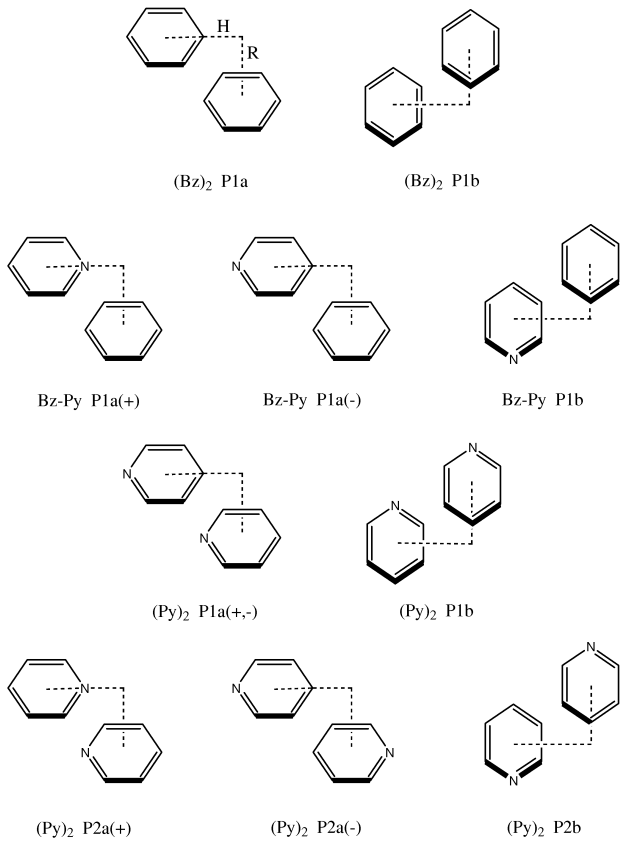
$(Bz)_2$ T Bz-Py T1 Bz-Py T2

Bz-Py T3 Bz-Py T4 $(Py)_2$ T1

$(Py)_2$ T2 $(Py)_2$ T3 $(Py)_2$ T4

level of theory. The most favorable T-shaped configuration was found to be Bz-Py T1, with $(Py)_2$ T3 being the second most favorable. These were the only two configurations found to be more favorable than $(Bz)_2$ T at the CCSD(T) CBS limit. Configurations of the T2

Figure 19: Parallel-displaced configurations of benzene dimer, benzene-pyridine, and pyridine dimer.



type (with a nitrogen of one ring pointed down at the center of another ring) were found to be more weakly bound with a shorter optimized inter-monomer separation.

Potential energy curves for the parallel-displaced configurations (Figure 19) were computed at vertical inter-monomer separations of 3.2, 3.4, 3.5, 3.6, and 3.8 Å. Horizontal displacements as large as 6 Å were considered. The parallel-displaced configurations of type *a* exhibit displacements over a vertex, while configurations of type *b* were displaced over a bond. Because of the large number of single point energies required for a thorough analysis of these configurations, CCSD(T) curves proved to be far too costly. Instead, the less computationally expensive SCS-MP2 method was employed.[67] Figure 20 clearly shows that the SCS-MP2/aug-cc-pVTZ curves are an excellent estimate of the estimated CCSD(T)/aug-cc-pVTZ curves. The SCSN scaling of MP2[87] was tested but did not work as well as SCS-MP2 for these complexes. These favorable results allowed SCS-MP2 to be

Table 15: Interaction energies of sandwich and T-shaped configurations of benzene dimer, benzene-pyridine, and pyridine dimer at various levels of theory.

	SCS-MP2/aTZ		Est. CCSD(T)/aTZ		Est. CCSD(T)/CBS	
	ΔE^a	R^b	ΔE^a	R^b	ΔE^a	R^b
$(Bz)_2$ S	-1.76	3.9	-1.64	3.9	-1.76	3.9
Bz-Py S	-2.19	3.8	-2.07	3.8	-2.22	3.8
$(Py)_2$ S1	-1.56	3.8	-1.48	3.9	-1.61	3.8
$(Py)_2$ S2	-2.88	3.7	-2.77	3.7	-2.95	3.7
$(Bz)_2$ T	-2.33	5.0	-2.59	4.9	-2.73	5.0
Bz-Py T1	-2.74	5.0	-3.02	5.0	-3.18	4.9
Bz-Py T2	-0.39	4.8	-0.64	4.7	—	—
Bz-Py T3	-1.80	5.1	-2.08	5.0	-2.20	5.0
Bz-Py T4	-2.36	5.0	-2.61	5.0	-2.74	5.0
$(Py)_2$ T1	-2.02	5.1	-2.32	5.0	-2.46	5.0
$(Py)_2$ T2	-0.95	4.7	-1.23	4.6	—	—
$(Py)_2$ T3	-2.55	5.0	-2.80	4.9	-2.95	4.9
$(Py)_2$ T4	-1.87	5.0	-2.15	5.0	—	—

^{a}CCSD(T) estimated CBS limit interaction energies in kcal mol^{-1}. ^{b}Inter-monomer separation in Å.

confidently applied to the remainder of the configurations. For completeness, SCS-MP2 can be compared to estimated CCSD(T) results with the aug-cc-pVTZ basis set for all complexes considered in this work (Table 15). SCS-MP2 performs well for sandwich configurations, further justifying its use for examining parallel-displaced complexes. The most favorable configuration of all those considered in this work is $(Py)_2$ P2b at $R = 3.4$ Å. This has a binding energy of 3.84 kcal mol^{-1} at the SCS-MP2/aug-cc-pVTZ level (Table 16). The most favorable benzene-pyridine configuration was Bz-Py P1a(-) at $R = 3.5$ Å with a binding energy of 3.23 kcal mol^{-1}.

The presence of nitrogen atoms in pyridine allows for the possibility of planar complexes with favorable CH$\cdots$N interactions. In a previous work by Piacenza and Grimme,[164] a configuration of pyridine dimer with C_{2h} symmetry containing two CH$\cdots$N interactions was examined. There are no analogous benzene dimer or benzene-pyridine configurations, and this configuration is the least significant for the π-π interactions which are the focus of this work. Nevertheless, due to the magnitude of the favorable interactions within this complex and the importance of hydrogen bonded interactions in biological complexes, the

Figure 20: Comparison of Bz-Py P1a(+,-) potential energy curves computed with various methods. Interaction energies in kcal mol^{-1}.

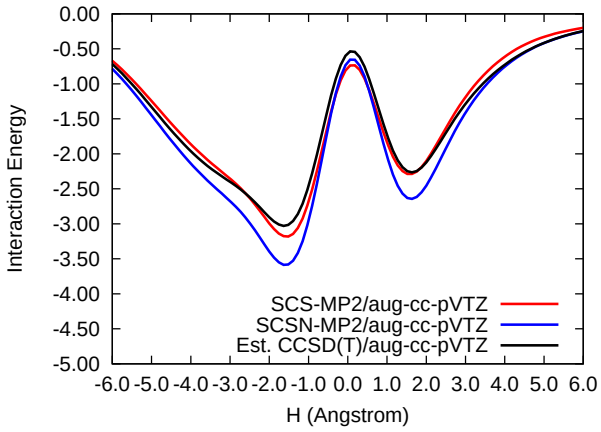


Table 16: Interaction energies of parallel-displaced configurations of benzene dimer, benzene-pyridine, and pyridine dimer computed at the SCS-MP2/aug-cc-pVTZ level of theory.

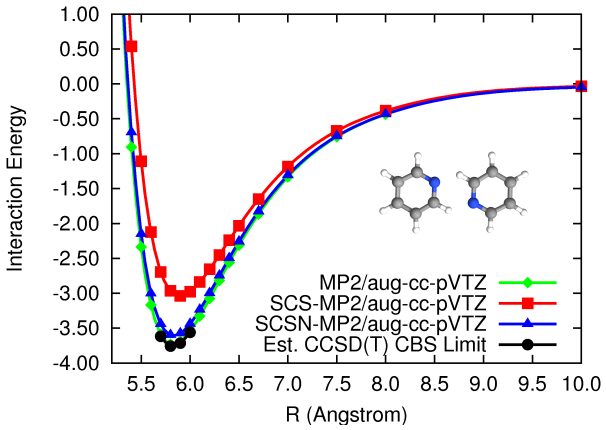
	SCS-MP2/aTZ		
	ΔE^a	R^b	H^b
$(Bz)_2$ P1a	-2.71	3.5	1.6
$(Bz)_2$ P1b	-2.70	3.5	1.6
Bz-Py P1a(+)	-2.36	3.5	1.4
Bz-Py P1a(-)	-3.23	3.5	1.6
Bz-Py P1b	-3.14	3.5	1.6
$(Py)_2$ P1a	-2.24	3.5	1.6
$(Py)_2$ P1b	-2.54	3.5	1.6
$(Py)_2$ P2a(+)	-2.78	3.5	1.4
$(Py)_2$ P2a(-)	-3.70	3.5	1.2
$(Py)_2$ P2b	-3.84	3.4	1.6

[a]SCS-MP2/aug-cc-pVTZ interaction energies in kcal mol^{-1}. [b]Distances given in Å.

hydrogen bonded pyridine dimer is interesting in its own right. A potential energy curve for this complex was computed at the estimated CCSD(T)/aug-cc-pVTZ level of theory (Figure 21). This complex is bound by 3.56 kcal mol^{-1} at an inter-monomer separation of 5.8 Å with CH$\cdots$N hydrogen bond distances of 2.5 Å. SCS-MP2 and SCSN-MP2 were tested for this hydrogen bonded pyridine dimer; MP2 and SCSN-MP2 perform well, while SCS-MP2 significantly underestimates the magnitude of the attractive interaction. Using

DFT-D, a much less computationally demanding technique than CCSD(T), Piacenza and Grimme report binding energies of 3.5-3.7 kcal mol^{-1} for the planar hydrogen bonded pyridine dimer.[164] These are in excellent agreement with our benchmark CCSD(T) results.

Figure 21: Potential energy curves for the hydrogen bonded pyridine dimer computed at the estimated CCSD(T)/aug-cc-pVTZ level of theory. Interaction energies in kcal mol^{-1}.



3.1.3 Sandwich Configurations

The most obvious difference in the intermolecular interactions of the $(Py)_2$ S1 and $(Py)_2$ S2 configurations is that they feature dipole-dipole interactions with opposite signs. The $(Py)_2$ S1 configuration has an unfavorable dipole-dipole interaction because the dipoles are parallel, while $(Py)_2$ S2 has a favorable dipole-dipole interaction because the dipoles are anti-parallel. In Bz-Py S, dipole-induced-dipole interactions are expected to contribute favorably to the binding energy. Note that all of these electrostatic interactions differ qualitatively from those in the benzene dimer, which lacks dipoles on the monomers; instead, the benzene dimer features quadrapole-quadrapole interactions. However, all of the sandwiches considered here, as well as the benzene dimer, do have in common favorable charge inter penetration terms due to an overlap of the π clouds.

Dispersion is also important in weakly bonded systems, and its magnitude can be related to the polarizability of the monomers. Pyridine is less polarizable than benzene.[45] This causes the dispersion interactions in complexes containing pyridine to be weaker than those

containing benzene. On the other hand, the contraction of the π electron cloud due to the heteroatom not only decreases the size of the favorable dispersion interactions, but it also decreases unfavorable exchange-repulsion interactions. Predictions about the relative size of these two changes are difficult to make; analysis with SAPT2 proves invaluable for quantifying these effects.

The SAPT2 results for the sandwich configurations at a separation of 3.8 Å (Table 17) generally confirm the above qualitative predictions about the various contributions to the interaction energy. Note that SAPT2 provides the same energetic ordering of sandwich complexes as does CCSD(T) estimated at the CBS limit. The $(Py)_2$ S2 configuration is predicted to have the most favorable electrostatic interactions and the least amount of dispersion, exchange-repulsion, and induction. These results stem from the reduced polarizability of pyridine and the anti-parallel alignment of the dipoles. As well as producing favorable electrostatic interactions, this alignment of the dipoles also maximizes separation of the electron density, thus lowering dispersion and exchange-repulsion. Each successive substitution of a pyridine for a benzene lowers the exchange-repulsion by roughly 0.5 kcal mol^{-1}. However, the magnitude of the favorable dispersion interaction is reduced by only about 0.3 kcal mol^{-1} per pyridine monomer. Because of this, the sum of dispersion and exchange-repulsion ("net dispersion") tends to become more favorable as benzenes are replaced by pyridines. As expected, the $(Py)_2$ S1 configuration is much less favorable than $(Py)_2$ S2 or $(Bz)_2$ S because of the parallel alignment of dipoles; this is reflected in a much less attractive electrostatic interaction in this configuration. The Bz-Py S configuration has a somewhat more favorable electrostatic contribution than $(Bz)_2$ S, but less favorable than that due to the anti-parallel dipoles of $(Py)_2$ S2. Perhaps surprisingly, the expected stabilization of Bz-Py S due to dipole-induced-dipole terms is not realized in the SAPT results; instead, the induction contribution for Bz-Py S is less attractive than in $(Bz)_2$ S. This unimportance of dipole-induced-dipole interactions has also been noted in substituted benzene dimers.[185,202]

Figure 22 shows the potential energy of the sandwich configurations as a function of the inter-monomer separation. For all distances considered, the Bz-Py heterodimer energy

Table 17: SAPT2 results for sandwich configurations of benzene dimer, benzene-pyridine, and pyridine dimer.[a]

	R^b	E_{elst}	E_{ind}	E_{exch}	E_{disp}	$E_{net\,disp}{}^c$	$E_{\mathrm{SAPT2}}{}^d$
$(Bz)_2$ S	3.8	-0.477	-0.275	4.516	-5.682	-1.166	-1.917
Bz-Py S	3.8	-0.800	-0.257	3.999	-5.335	-1.336	-2.393
$(Py)_2$ S1	3.8	-0.049	-0.208	3.565	-4.999	-1.435	-1.691
$(Py)_2$ S2	3.8	-1.294	-0.245	3.488	-4.996	-1.508	-3.047

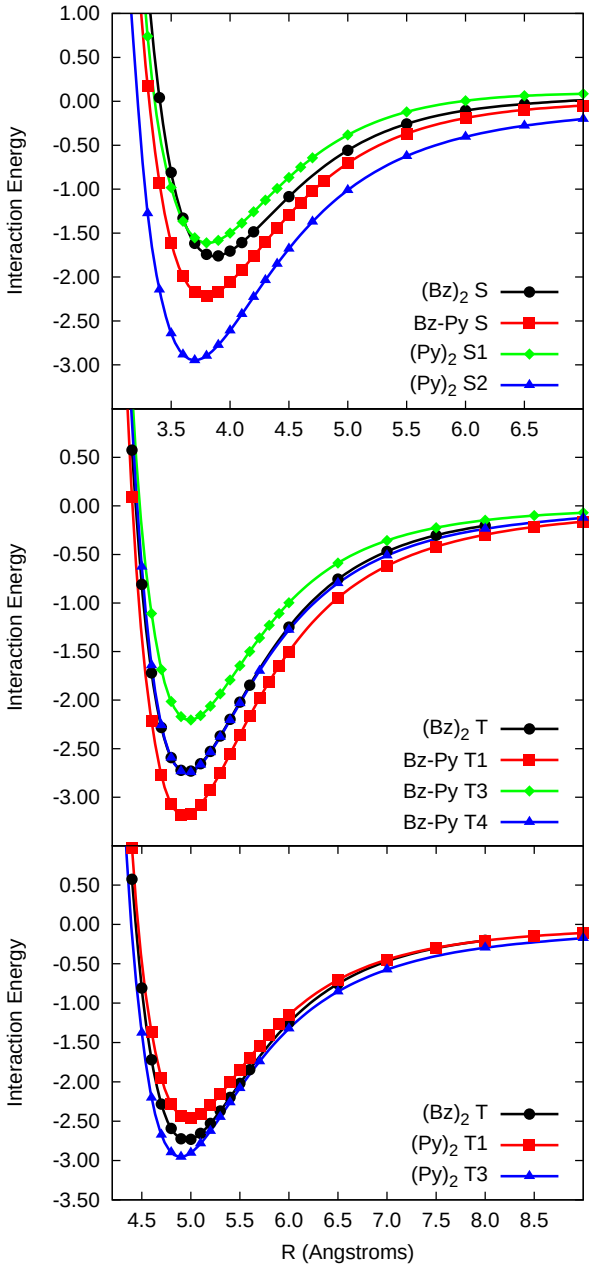
[a]Computations performed using the aug-cc-pVDZ′ basis set. Energies in kcal mol^{-1} and distances in Å. [b]Inter-monomer separation in Å. [c]Net dispersion is the sum of the exchange and dispersion components. [d]Total SAPT2 interaction energy.

lies in between those of the $(Bz)_2$ S and $(Py)_2$ S2 dimers, although the energy is somewhat closer to that of $(Bz)_2$ S. The $(Py)_2$ S1 configuration is the least favorable sandwich except at short distances (R < 3.5 Å) when it becomes slightly more favorable than $(Bz)_2$ S. This is due to the reduced spatial extent of the electron density for a pyridine monomer reducing the rate at which exchange-repulsion increases with decreasing inter-monomer separation relative to $(Bz)_2$ S.

3.1.4 T-shaped Configurations

The behavior of the T-shaped configurations containing pyridine monomers can be understood by examining how the interactions that stabilize the T-shaped benzene dimer are changed by the introduction of a heteroatom. As demonstrated by SAPT analysis (Table 18), the dominant stabilizing interaction in most of the T-shaped configurations is electrostatic; exchange-repulsion and dispersion terms are also large, but they tend to cancel, so their sum (net dispersion) is relatively small. The favorable electrostatic interaction can be rationalized by considering the attraction that will result between the negative π cloud on one ring and the positive charge on the hydrogen pointed toward it. In pyridine, the nitrogen atom pulls electron density away from the hydrogen para to it, increasing that hydrogen's positive charge relative to its value in benzene, making pyridine more effective as a "π hydrogen bond" donor. However, because the presence of this nitrogen also distorts the π electron cloud by pulling electron density away from the center of the ring and toward

Figure 22: Potential energy curves for sandwich and T-shaped configurations computed at the estimated CCSD(T) CBS limit. Interaction energies in kcal mol^{-1}.



the nitrogen (see Figure 23), one might expect that this makes pyridine somewhat less effective than benzene as a "π hydrogen bond" acceptor; indeed, Bz-Py T3 is less favorable than $(Bz)_2$ T, and $(Py)_2$ T1 is less favorable than Bz-Py T1. The Bz-Py T1 configuration

should be the most favorable of those considered, and this conclusion is supported by the estimated CCSD(T) CBS limit potential energy curves (see Table 15 and Figure 22). Because $(Py)_2$ T1 is not as strongly bound as $(Bz)_2$ T, it appears that nitrogen heteroatoms have a larger unfavorable effect on "π hydrogen bond" accepting ability than they have a favorable effect on "π hydrogen bond" donating ability. The weak binding in the Bz-Py T2 and $(Py)_2$ T2 complexes occurs because the electrostatic contribution is much less favorable or even unfavorable as the negative charge on nitrogen points down at the negative π cloud below (see Table 15 and Figure 24). These complexes remain weakly bound because of favorable induction and dispersion contributions. The minimum energy geometries of these two complexes have shorter inter-monomer separations than the other T-shaped complexes.

Table 18: SAPT2 results for T-shaped configurations of benzene dimer, benzene-pyridine, and pyridine dimer.[a]

	R^b	E_{elst}	E_{ind}	E_{exch}	E_{disp}	$E_{net\,disp}{}^c$	$E_{\mathrm{SAPT2}}{}^d$
$(Bz)_2$ T	5.0	-1.753	-0.518	3.517	-3.730	-0.213	-2.484
Bz-Py T1	5.0	-2.118	-0.635	3.540	-3.696	-0.156	-2.909
Bz-Py T2	4.7	0.329	-0.616	3.381	-3.754	-0.372	-0.659
Bz-Py T3	5.0	-1.209	-0.399	3.271	-3.512	-0.241	-1.850
Bz-Py T4	5.0	-1.804	-0.485	3.209	-3.359	-0.150	-2.439
$(Py)_2$ T1	5.0	-1.391	-0.498	3.288	-3.477	-0.189	-2.078
$(Py)_2$ T2	4.7	-0.392	-0.536	3.171	-3.534	-0.363	-1.292
$(Py)_2$ T3	5.0	-1.780	-0.378	2.674	-3.106	-0.431	-2.589
$(Py)_2$ T4	5.0	-1.138	-0.383	2.776	-3.124	-0.348	-1.869

[a]Computations performed using the aug-cc-pVDZ′ basis set. Energies in kcal mol^{-1} and distances in Å. [b]Inter-monomer separation in Å. [c]Net dispersion is the sum of the exchange and dispersion components. [d]Total SAPT2 interaction energy.

In complexes $(Py)_2$ T3 and $(Py)_2$ T4, the top pyridine is rotated 90°, leading to significant contributions from dipole-dipole interactions. We also examined a similar Bz-Py configuration, T4. The direct interaction between a single hydrogen with the π electron cloud below it is replaced by a less direct interaction of two hydrogens with the π cloud. The estimated CCSD(T) CBS limit potential energy curves show $(Py)_2$ T3 to be the most favorable of these three configurations, which would be expected because this configuration features anti-parallel dipoles. This is the most favorable T-shaped pyridine dimer considered, and the only one that binds more strongly than $(Bz)_2$ T. Bz-Py T4 binds more weakly

Figure 23: Electrostatic potential computed at the Hartree-Fock/6-31G* level of theory. The scale is -25 (red) to 25 (blue) kcal mol^{-1}. A benzene molecule is shown on the left, pyridine is on the right.

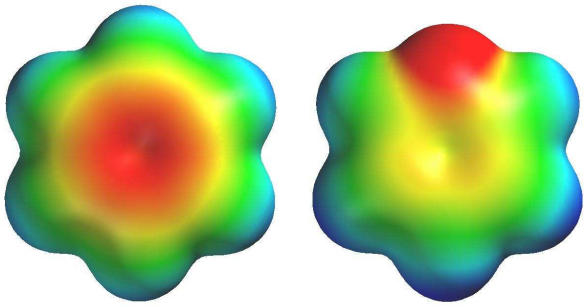
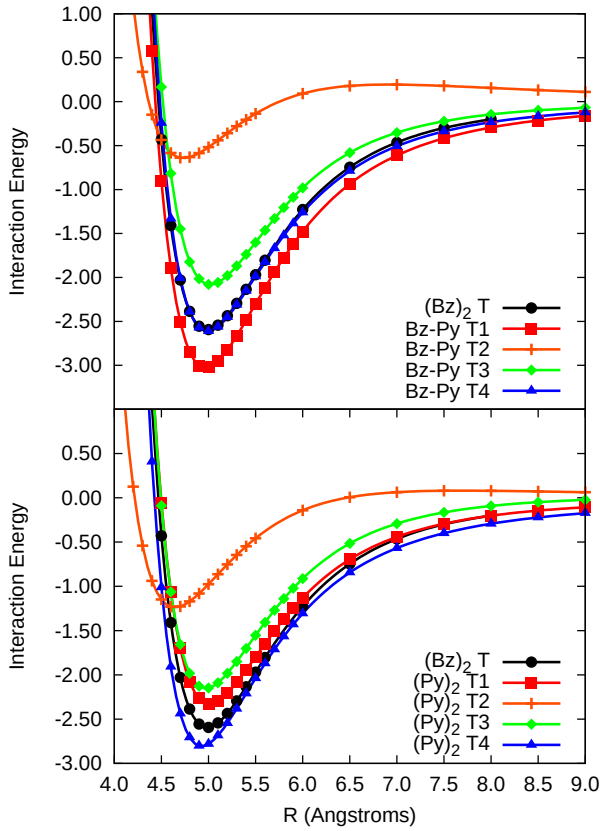


Figure 24: Potential energy curves for T-shaped configurations computed at the estimated CCSD(T)/aug-cc-pVTZ level of theory. Interaction energies in kcal mol^{-1}.



than the benzene dimer by a small amount. The least favorable complex is of course $(Py)_2$ T4, in which the dipoles are parallel.

To quantify the change in the positive charge on the interacting hydrogen, natural population analysis charges were computed for benzene and pyridine with a Hartree-Fock/6-311++G** wavefunction. The para carbon and hydrogen in pyridine become more positive relative to benzene. This causes the electrostatic interaction between the para position and the π cloud of the benzene monomer below it to be larger in Bz-Py T1 than in $(Bz)_2$ T. As shown in Table 18, the sum of dispersion and exchange-repulsion is comparable to the size of the electrostatic contribution only in the case of the weakly bound Bz-Py T2 and $(Py)_2$ T2 complexes; for all other configurations, the electrostatic contribution dominates and is stabilizing by one kcal mol^{-1} or more. The SAPT2 computations show that inductive effects are also an important factor stabilizing the T-shaped complexes. This contribution is stabilizing by 0.52 kcal mol^{-1} in the benzene dimer due to the quadrapole induced-multipole interactions, and perhaps surprisingly it remains close to this size in all of the Bz-Py and $(Py)_2$ T-shaped complexes considered, even though pyridine features a dipole rather than a quadrapole moment. Because pyridine is less polarizable than benzene, configurations in which the lower monomer is pyridine tend to have less favorable induction contributions than the benzene dimer. Induction is enhanced in the two Bz-Py T-shaped configurations (T1 and T2) in which the dipole of pyridine is parallel to the C_6 axis of the benzene below it.

3.1.5 Parallel-Displaced Configurations

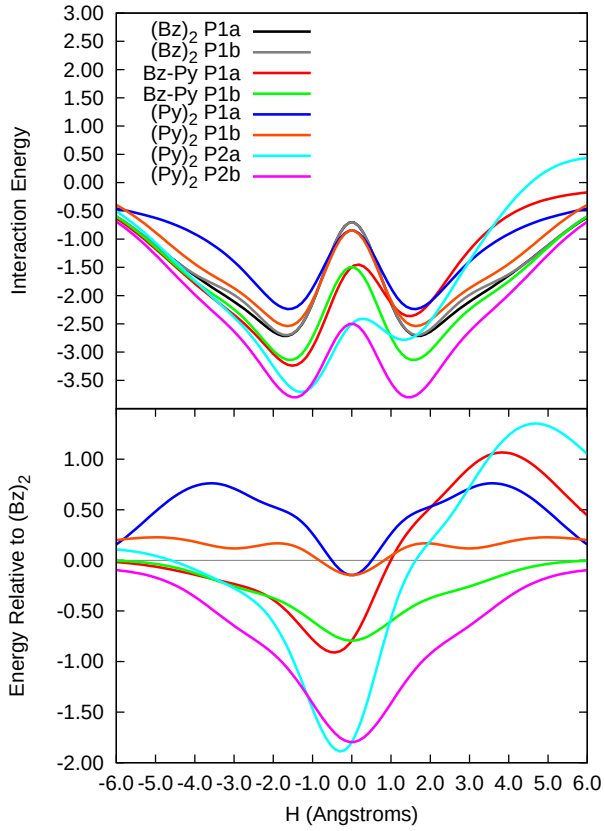
The parallel-displaced configurations were analyzed at the SCS-MP2/aug-cc-pVTZ level of theory. The sign of the horizontal displacement is shown by (+) and (-) for the cases that are not symmetric with respect to horizontal displacements away from the sandwich configuration; in all of the "edgewise" displaced configurations, labeled *b*, the geometries are symmetric with respect to the horizontal displacement. The dipole-dipole interactions and the interplay between dispersion and exchange-repulsion seen in sandwich configurations are also observed in the parallel-displaced complexes. The interactions between hydrogen

atoms and π electron clouds important in the T-shaped configurations are also seen here. Due to the complicated interplay between these various contributions, SAPT2 analysis and potential energy curves relative to benzene dimer are essential to understanding the behavior of these interactions.

The potential energy curves for the parallel-displaced configurations at a vertical separation of 3.5 Å are shown in Figure 25. The most favorable geometries found in this work and their corresponding interaction energies are contained in Table 16. The most favorable of the complexes examined is $(Py)_2$ P2b; this is the $(Py)_2$ S2 geometry displaced "edgewise." $(Py)_2$ P2a(-) lies only 0.14 kcal mol^{-1} above $(Py)_2$ P2b; both of these complexes contain anti-parallel dipoles. The next most favorable complex is Bz-Py P1a(-), and also in this case the Bz-Py P1b complex is nearly isoenergetic, differing by only 0.09 kcal mol^{-1}. In this case the "edgewise" displacement is not as favorable as the "over vertex" displacement. As a result of the higher symmetry in benzene compared to pyridine, the difference between displacements over a vertex or over an edge is even smaller in the benzene dimer, merely 0.01 kcal mol^{-1} at the SCS-MP2/aug-cc-pVTZ level of theory. The most favorable pyridine-containing parallel-displaced complexes are bound more strongly than the benzene dimer. Not surprisingly, the $(Py)_2$ P1a and $(Py)_2$ P1b dimers are the least favorable as a result of their parallel dipoles. With regard to the conclusion that the Bz-Py P1a(-) configuration is the most favorable benzene-pyridine complex, the data reported in this work agrees with that of Tsuzuki *et al.*[226] Tsuzuki *et al.* report estimated CCSD(T) CBS limit binding energies of 3.04 and 2.22 kcal mol^{-1} for complexes very similar to our $(Py)_2$ P2a(-) and $(Py)_2$ P2a(+) configurations, respectively. This is in good agreement with the less computationally demanding SCS-MP2/aug-cc-pVTZ binding energies reported in this work of 3.23 and 2.36 kcal mol^{-1}, respectively.

The parallel-displaced complexes pass through sandwich configurations at $H = 0$. As can be seen in Figure 25, all of the pyridine containing complexes are more favorable than the benzene dimer at $H = 0$ Å for a vertical distance of $R = 3.5$ Å (although the $(Py)_2$ P1 configurations are less favorable than benzene dimer at larger vertical separations due to the parallel pyridine dipoles). This is likely due to the contracted π electron clouds seen in

Figure 25: Potential energy curves for parallel-displaced configurations computed at a vertical displacement of 3.5 Å. Potential curves relative to $(Bz)_2$ P1a for *a* configurations or $(Bz)_2$ P1b for *b* configurations are also shown. Interaction energies in kcal mol^{-1}.



pyridine monomers. As discussed above, the contraction of these clouds leads to a reduction in exchange-repulsion relative to benzene dimer. In all of the complexes except $(Py)_2$ P1, there is also an increased electrostatic attraction. The "edgewise" displaced Bz-Py P1b and $(Py)_2$ P2b complexes remain more favorable than benzene dimer for the entire range of horizontal displacements examined. The Bz-Py P1a(+) and $(Py)_2$ P2a(+) complexes become less favorable than the benzene dimer as the nitrogen in the pyridine monomers begins interacting with the π electron cloud of the other monomer. The $(Py)_2$ P1a and $(Py)_2$ P1b dimers become less favorable than the benzene dimer once the horizontal displacement increases sufficiently to lessen the importance of the reduced exchange-repulsion. Although these complexes remain less favorable than the benzene dimer for all horizontal

displacements studied past the sandwich-like configurations, one can infer from the slope of the $(Py)_2$ P1a potential energy curve that it will become more favorable than the benzene dimer at larger horizontal displacements. This is likely due to slightly more favorable electrostatic interactions at large separations.

The SAPT2 decomposition of the interaction energies within these parallel-displaced complexes at a vertical separation of 3.4 Å and horizontal separation of 1.6 Å is presented in Table 19. The distance between ring centers at this displacement is comparable to that examined for the sandwich configurations (roughly 3.76 Å compared to 3.8 Å). The electrostatic and induction energies play a major role in the binding of these complexes. The net dispersion terms computed here are all repulsive for these complexes except for $(Py)_2$ P2a(+) and $(Py)_2$ P2b, although this may be different at other geometries. Both electrostatics and net dispersion can change significantly among the different complexes and orientations considered. The orientation of the dipoles is obviously important for the electrostatic interactions. However, the orientation of the pyridine monomers can also strongly influence the exchange-repulsion and dispersion terms. The induction terms, which depend on polarizability, are weakly affected by heteroatoms in parallel-displaced complexes, decreasing by about 0.1-0.2 kcal mol^{-1} compared to their value in the benzene dimer. At the geometry considered, the electrostatic term is larger than induction or net dispersion for the parallel-displaced configurations.

In a recent work by Tsuzuki *et al.*,[226] the interaction energy of parallel-displaced benzene-pyridine is decomposed into its physically relevant components, but with significantly different results than we present in this work using SAPT2. Those authors conclude that parallel-displaced pyridine-benzene complexes are bound primarily by net dispersion interactions. Although we concur that dispersion is the largest single stabilizing factor, our SAPT2 results suggest that net dispersion (the sum of dispersion and exchange-repulsion) is generally less important than electrostatics in these configurations. To examine this discrepancy, we performed a decomposition of the interaction energies for the geometries reported in Tsuzuki *et al.*[226] using SAPT2. The results are presented in Table 20. For the Bz-Py P1a

Table 19: SAPT2 results for parallel-displaced configurations of benzene dimer, benzene-pyridine, and pyridine dimer.[a]

	R^b	H^c	E_{elst}	E_{ind}	E_{exch}	E_{disp}	$E_{net\,disp}{}^d$	$E_{\mathrm{SAPT2}}{}^e$
$(Bz)_2$ P1a	3.4	1.6	-2.774	-0.882	8.584	-7.879	0.705	-2.952
$(Bz)_2$ P1b	3.4	1.6	-2.805	-0.912	8.677	-7.883	0.795	-2.922
Bz-Py P1a(+)	3.4	1.6	-1.905	-0.733	7.244	-7.156	0.088	-2.550
Bz-Py P1a(-)	3.4	1.6	-3.235	-0.853	8.248	-7.587	0.661	-3.427
Bz-Py P1b	3.4	1.6	-2.948	-0.807	7.768	-7.383	0.386	-3.369
$(Py)_2$ P1a	3.4	1.6	-1.742	-0.682	6.969	-6.897	0.072	-2.353
$(Py)_2$ P1b	3.4	1.6	-2.039	-0.663	7.008	-6.927	0.081	-2.621
$(Py)_2$ P2a(+)	3.4	1.6	-1.781	-0.655	5.904	-6.456	-0.553	-2.988
$(Py)_2$ P2a(-)	3.4	1.6	-3.538	-0.786	7.926	-7.302	0.624	-3.701
$(Py)_2$ P2b	3.4	1.6	-3.258	-0.703	6.851	-6.894	-0.043	-4.004

[a]Computations performed using the aug-cc-pVDZ′ basis set. Energies in kcal mol^{-1} and distances in Å. [b]Vertical separation in Å. [c]Horizontal separation in Å. [d]Net dispersion is the sum of the exchange and dispersion components. [e]Total SAPT2 interaction energy.

geometries, the most striking difference is seen in the electrostatic contributions to the interaction. Our quantum mechanically based SAPT2 results predict electrostatic interactions to be a major factor stabilizing these complexes, while the decomposition from Tsuzuki *et al.* predicts that electrostatic interactions destabilize these complexes. This discrepancy results from Tsuzuki *et al.* using distributed multipoles to compute the electrostatic interaction. This procedure does not account for the favorable electrostatic interactions originating from the interpenetration of π electrons, an effect which has been shown to be important for the stabilization of the benzene dimer.[202] The lower exchange energies obtained by Tsuzuki *et al.* are a direct result of the method used for the computation of the electrostatic interaction because the exchange energy was reported as the remainder of the interaction energy after dispersion, electrostatic, and induction energy had been computed explicitly. For completeness, we also used SAPT2 to examine the T-shaped complexes reported by Tsuzuki *et al.* As would be expected, the electrostatic interactions in the T-shaped complexes are described fairly well by the multipole analysis.

Table 20: Energy component analysis for optimized benzene-pyridine complexes; present results using SAPT2, literature results (from Ref. 226) in parentheses.[a]

	E_{elst}	E_{ind}	E_{exch}	E_{disp}	$E_{net\ disp}{}^{c}$	ΔE^{d}
Bz-Py P1a(+)	-1.012	-0.503	4.915	-6.000	-1.085	-2.599
	(0.99)	(-0.21)	(2.85)	(-5.84)	(-2.99)	(-2.22)
Bz-Py P1a(-)	-2.497	-0.698	6.472	-6.807	-0.334	-3.529
	(0.39)	(-0.22)	(3.27)	(-6.48)	(-3.21)	(-3.04)
Bz-Py T1	-1.882	-0.555	2.858	-3.343	-0.485	-2.922
	(-1.01)	(-0.20)	(1.26)	(-2.86)	(-1.60)	(-2.81)
Bz-Py T3	-1.382	-0.328	2.227	-3.074	-0.847	-2.557
	(-0.57)	(-0.08)	(0.88)	(-2.80)	(-1.92)	(-2.57)
Bz-Py T4	-2.003	-0.523	3.177	-3.626	-0.449	-2.976
	(-0.96)	(-0.19)	(1.60)	(-3.31)	(-1.71)	(-2.87)

[a]Energies in kcal mol^{-1}. SAPT2 Computations preformed using the aug-cc-pVDZ$'$ basis set using geometries reported by Tsuzuki *et al.*[226] [b]Decomposition performed by Tsuzuki *et al.*[226] [c]Net dispersion is the sum of the exchange and dispersion components. [d]Total interaction energy.

3.1.6 Conclusions

The parallel-displaced configurations of benzene-pyridine and pyridine dimer were the most favorable complexes studied in this work. In the case of the benzene dimer, the T-shaped and parallel-displaced configurations are nearly isoenergetic; by substituting nitrogen atoms (and consequently introducing dipole moments), the parallel-displaced configurations become favored. $(Py)_2$ P2b and Bz-Py P1a(-) were the most favorable configurations found for the pyridine dimer and benzene-pyridine complex, respectively. The most favorable benzene-pyridine complex was found to bind more strongly than benzene dimer by roughly 0.5 kcal mol^{-1}. The most favorable pyridine dimer was found to bind about 1 kcal mol^{-1} more strongly than benzene dimer. The inter-monomer separation for the minimum energy structures of each configuration of benzene-pyridine and pyridine dimer did not change substantially relative to benzene dimer.

The substitution of a nitrogen atom into a benzene molecule creates a dipole in the molecule, reduces its polarizability, and reduces the spatial extent of the electron density. The presence of a heteroatom in pyridine makes the electrostatic interactions within

pyridine-containing dimers much more sensitive to the orientation of the monomers. In general, the substitution of benzene monomers in benzene dimer with pyridine molecules will reduce the magnitude of the dispersion and induction interactions as a result of the reduced polarizability of pyridine. Similarly, the reduced spatial extent of the π-electron cloud in a pyridine molecule leads to reduced exchange-repulsion. These general trends observed here can be expected to persist in larger and more complex heteroatom-containing π systems.

For sandwich and parallel-displaced configurations, pyridine monomers cause dipole-induced-dipole interactions in benzene-pyridine and dipole-dipole interactions in pyridine dimer. The former is found to be relatively unimportant, while the latter is very important and can lead to more favorable or less favorable electrostatic interactions, depending on the configuration. The other important considerations, stemming from the decreased polarizability and reduced spatial extent of the electron density, are a reduction in the magnitude of the dispersion and exchange-repulsion energies relative to benzene dimer. Because the exchange-repulsion and dispersion terms are of opposite sign but with roughly equal magnitude, it is convenient to consider their sum, "net dispersion." The electrostatic and net dispersion interactions both play an important role in the interaction energy of the sandwich configurations. Limited SAPT2 analysis at selected geometries suggests that the electrostatic term tends to dominate the interaction energy near the equilibrium geometries of parallel-displaced configurations; in addition, the most strongly bound parallel-displaced pyridine-containing complexes studied in this work had electrostatic interactions that were much more favorable than the complexes which were bound more weakly. Generally speaking, electrostatics are also the dominant stabilizing factor in the T-shaped complexes, although there are also favorable induction and net dispersion contributions; in the $(Bz)_2$ T, Bz-Py T1, Bz-Py T3 and $(Py)_2$ T1 complexes the electrostatic attraction is related to what might be called a "π hydrogen bond." For Bz-Py T4 and $(Py)_2$ T3 the favorable electrostatics originate from dipole effects. The less favorable Bz-Py T2, $(Py)_2$ T2 and $(Py)_2$ T4 complexes do not contain either the "π hydrogen bonds" or stabilizing dipole effects. In all the configurations considered, there were benzene-pyridine and pyridine dimer complexes found that were more favorable than the analogous benzene dimer complex.

Previous work had indicated that parallel-displaced benzene-pyridine complexes are bound primarily due to dispersion effects. Our SAPT-based analysis indicates that dispersion is a major stabilizing force, but it is mostly canceled be exchange-repulsion. The sum of these two terms is usually repulsive at the near-equilibrium geometries considered. Electrostatic interactions are very important and are significantly stabilizing according to the quantum-mechanical SAPT method, which can capture charge interpenetration effects neglected in a multipole analysis.

3.1.7 Epilogue

The work described above can be found in Ref. 92. The question of heteroatom effects was subsquently studied by Kim and co-workers,[61] who examined dimers with multiple heteroatom substitutions. The conclusions from Ref. 92 were shown to generalize; additional heteroatom substitutions continue to reduce polarizability and the spatial extent of the π-orbitals. The benzene-pyridine and pyridine dimer complexes were also explored with the effective fragment potential method (EFP) by Slipchenko and co-workers.[205] The potential energy curves and SAPT2 decompositions reported in Ref. 92 were used extensively as a benckmark for the performance of the EFPs.

3.2 Characterizing the indole-benzene complex

The following was adapted from Ref. 60.

The interaction energy of selected T-shaped configurations of the indole-benzene complex was decomposed using symmetry-adapted perturbation theory (SAPT).[111,246] For ease of comparison of energy components between similar dimer configurations, for all axial T-shaped configurations SAPT results were evaluated for the same distance between the center of benzene and the plane of indole (4.9 Å); likewise, for all equatorial T-shaped configurations, a fixed distance of 2.48 Å was used for the distance betwen the center of benzene to the closest hydrogen in indole. The SAPT0 computations were performed with a truncated aug-cc-pVDZ basis set denoted aug-cc-pVDZ′. This basis set removes all diffuse functions from hydrogen atoms and diffuse d functions from non-hydrogen atoms. In our experience, fortuitous error cancellation occurs for π-π interactions when this basis set is paired with

MP2-like methods.[203] The two-electron integrals necessary for SAPT0 were computed with the density fitting (DF) approximation. A truncated version of the aug-cc-pVDZ-RI fitting basis set[81] was designed to match the aug-cc-pVDZ′ orbital basis. The diffuse functions were removed from the fitting basis for hydrogen atoms and diffuse f functions were removed for non-hydrogen atoms.

The benzene monomer geometry was set to those suggested by Gauss and Stanton, R(C-C)=1.3915 Å and R(C-H)=1.0800 Å.[59] The indole monomer geometry was optimized at the MP2/aug-cc-pVDZ level. Monomer geometries were kept frozen as the intermolecular distances were varied. Three "axial" T-shaped configurations are shown in Figure 26 and seven "equatorial" T-shaped configurations are shown in Figure 27 (where the plane of the indole ring is defined as the 'equator').

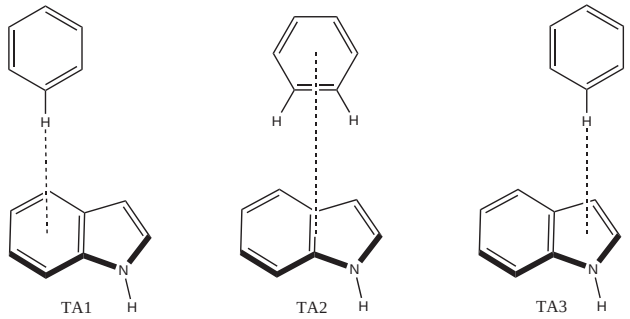


Figure 26: Three axial T-shaped configurations of the indole-benzene complex.

The trend observed in Figure 28 can be explained by examining the electrostatic potentials (ESPs) of benzene and indole as shown in Figure 29 and the decomposition of the interaction energies with SAPT as shown in Table 21. Favorable interactions occur for all configurations due to the interaction between the positive hydrogens of benzene and the negative π-cloud of indole. Furthermore, these interactions are enhanced via favorable dispersion interactions between the π-clouds of indole and benzene. Considering the ESP of indole, one might expect configuration TA1 to be more favorable than TA3 due to stronger electrostatic interactions; the ESP above the 6-membered ring appears more negative than above the 5-membered ring. However, the SAPT analysis of Table 21 indicates that TA3 has an electrostatic attraction about 0.4 kcal mol^{-1} stronger than that in TA1. A comparision of electrostatics and exchange for the TA1 and TA3 configurations shows, surprisingly,

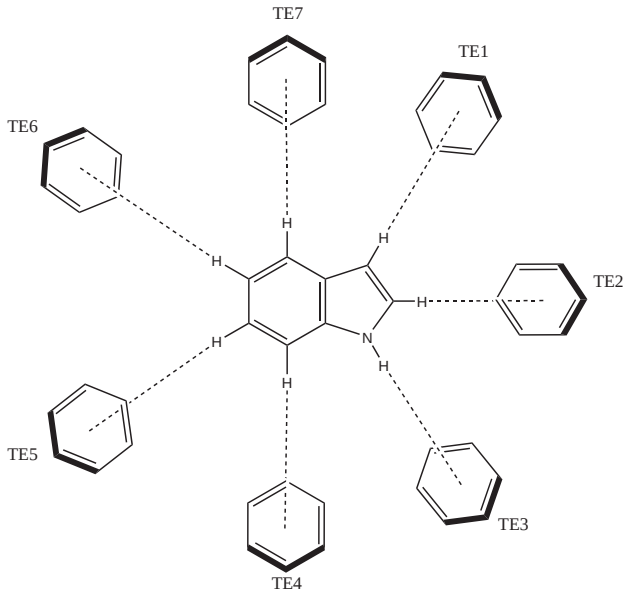


Figure 27: Seven equatorial T-shaped configurations of the indole-benzene complex.

larger values for the TA3 configuration. This is possibly indicative of increased charge penetration for the TA3 configuration. Net dispersion (the sum of dispersion and exchange)[92] is more favorable (by 0.6 kcal mol^{-1}) for TA1, in which the benzene is above the larger 6-membered ring. Overall, the interaction energies of TA1 and TA3 are very similar, with TA1 being slightly favored in the MP2, SCS-MP2, and SAPT computations.

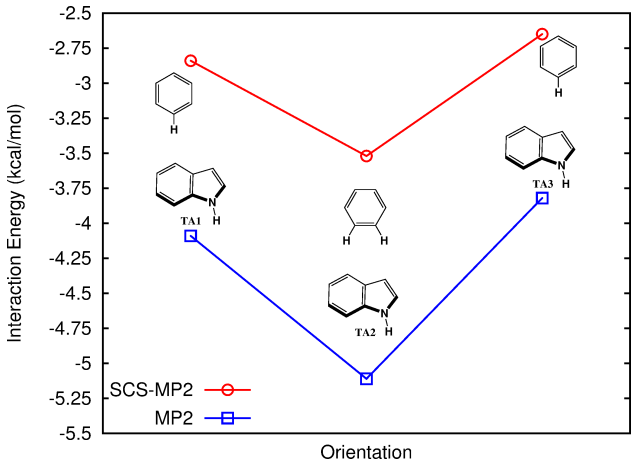


Figure 28: MP2 and SCS-MP2/aug-cc-pVDZ interaction energies (kcal mol^{-1}) for the three axial T-shaped configurations.

The center configuration, TA2, is more favorable than TA1 or TA3 because it has two C-H/π contacts and allows the benzene ring to interact with the π-clouds of both the 5- and

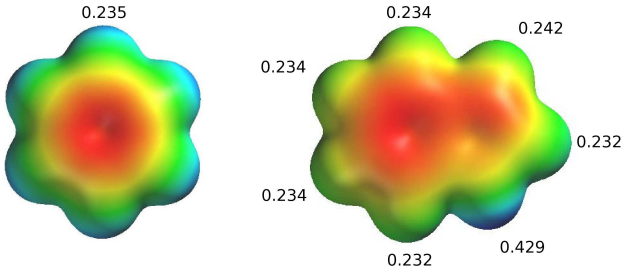


Figure 29: B3LYP/6-31G* electrostatic potential mapping for benzene (left) and indole (right), where blue is positive and red is negative. Numbers are the hydrogen atomic charges.

Table 21: Physical components (kcal mol^{-1}) of the total interaction energy determined using SAPT0 for the axial T-shaped configurations.[a,b]

	E_{elst}	E_{exch}	E_{ind}	E_{disp}	$E_{net\ disp}$[c]	E_{SAPT0}
TA1	-1.94	3.13	-0.57	-4.32	-1.19	-3.70
TA2	-2.19	3.25	-0.60	-4.95	-1.71	-4.49
TA3	-2.32	4.27	-0.68	-4.83	-0.56	-3.55
TE1	-1.71	3.77	-0.55	-4.28	-0.50	-2.77
TE2	-2.31	3.48	-0.72	-3.74	-0.26	-3.29
TE3	-3.31	3.09	-1.01	-4.16	-1.07	-5.40
TE4	-2.18	4.05	-0.76	-4.63	-0.57	-3.46
TE5	-1.66	3.83	-0.53	-4.19	-0.36	-2.55
TE6	-1.71	3.74	-0.54	-4.15	-0.41	-2.66
TE7	-1.86	4.02	-0.58	-4.70	-0.68	-3.12
PD(MP2)	-3.24	10.49	-1.22	-11.00	-0.51	-4.96
PD(SCS)	-2.99	6.37	-0.91	-7.58	-1.21	-5.11

[a]Computed using the aug-cc-pVDZ′ basis set. [b]For TA configurations, distances from the center of benzene to the plane of indole are 4.9 Å. For TE configurations, distances from the center of benzene to each respective hydrogen of indole are 2.48 Å. For PD configurations, minima are (R1=3.4, R2=0.1, R3=-1.0 Å, MP2/aug-cc-pVDZ) and (R1=3.4, R2=1.3, R3=-1.8 Å, SCS-MP2/aug-cc-pVDZ). [c]Net dispersion is the sum of the exchange and dispersion components.

6-member rings of indole. A similar trend was observed for the methane-indole complex in a study by Ringer *et al.*[181] For that system, a decomposition of the interaction energy through SAPT suggested that the configuration in which two hydrogens of methane point toward the rings of indole is stabilized primarily by increased electrostatic attraction relative to the configurations with one interacting hydrogen. However, in the case of the indole-benzene complex, the electrostatic interaction of configuration TA2 strengthens compared to that in

configuration TA1, but it does not become as strong as that in TA3. Instead, the increased net dispersion from the interaction of the benzene ring with both π-clouds is the primary component that stabilizes configuration TA2 over configurations TA1 and TA3; the net dispersion of configuration TA2 is more favorable than configurations TA1 and TA3 by 0.52 kcal mol^{-1} and 1.15 kcal mol^{-1}, respectively. Unlike the indole-methane complex, the benzene molecule has the potential for long-range CH-π interactions in the TA1 and TA3 configurations (due to the hydrogen pointing downwards 60° off-axis); in indole-methane, the analogous TA1 and TA3 configurations have three hydrogens pointing upwards. These long-range contacts may be responsible for the increased electrostatic interactions in the TA1 and TA3 configurations.

The equatorial T-shaped indole-benzene interaction energy trend (Figure 30) can be partially rationalized by examining the ESPs and hydrogen atomic charges in Figure 29, and more completely by the components of SAPT in Table 21. The hydrogen bonded to the nitrogen of the 5-member indole ring exhibits the largest positive charge. Therefore, the most favorable interaction is TE3, where the most positive hydrogen interacts with the negative benzene π-cloud. The SAPT analysis confirms that this configuration has the most stabilizing electrostatic interaction by 1.0 kcal mol^{-1}. The partial charges of the other hydrogens in indole are all fairly similar, and so we might expect that configurations besides TE3 would have similar electrostatic contributions. Instead, we see a significant stabilization (0.3 - 0.7 kcal mol^{-1}) of configurations TE2 and TE4 relative to the remaining configurations. The hydrogens pointed at benzene in these configurations are those hydrogens which are on either side of the most positive hydrogen, and we believe that the extra electrostatic stabilization of these configurations is due to longer-range attractions between the π-cloud of benzene and that most positive hydrogen. We note that the configurations with the largest electrostatic stabilization also have the largest induction contributions; the polarization of the benzene π-orbitals seems to correlate with the partial charge on the indole hydrogens. We also note that configurations TE1, TE3, TE4, and TE7 are those with the largest net dispersion interaction; in the other configurations, the benzene is much closer to one ring than the other, and so interactions with the second ring are greatly diminished.

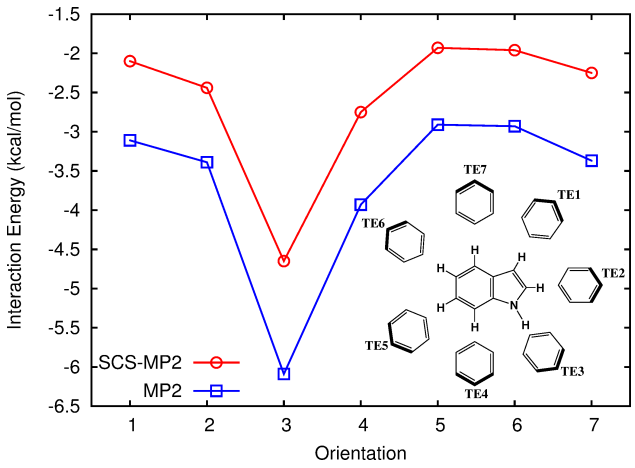


Figure 30: MP2 and SCS-MP2/aug-cc-pVDZ interaction energies (kcal mol^{-1}) for the seven equatorial T-shaped configurations.

The parallel configurations are stabilized by π-π dispersion interactions between the benzene ring and the two rings of indole, as well as favorable electrostatic effects due to charge interpenetration. As shown by the top views of Figure 31, the MP2 method locates a minimum corresponding to the center of the benzene ring laying over the shared bond of indole and shifted towards the nitrogen. On the other hand, at this same vertical separation, the SCS-MP2 minimum places the center of the benzene ring almost directly over the nitrogen of indole. The SAPT investigation on these two structures, shown in Table 21, suggests that the SCS-MP2 minimum is preferred over the MP2 minimum (according to SAPT) due to a dramatic decrease in the exchange contribution, which leads to a more favorable net dispersion interaction. The total SAPT interaction energies, utilizing the truncated basis set, indicates that the SCS-MP2 minimum is slightly more favored than the MP2 minimum. Further analysis by coupled-cluster methods demonstrate that the minimum predicted by SCS-MP2 is indeed lower in energy than the minimum predicted by MP2 (-4.64 and -4.46 kcal mol^{-1}, respectively). More advanced SAPT computations, SAPT2+(3)/aug-cc-pVDZ, are in excellent agreement with CCSD(T) (-4.72 and -4.38 kcal mol^{-1}, respectively). The SAPT computations suggest that the failure of MP2 to predict the correct PD minimum is a result of its overestimation of the dispersion energy.

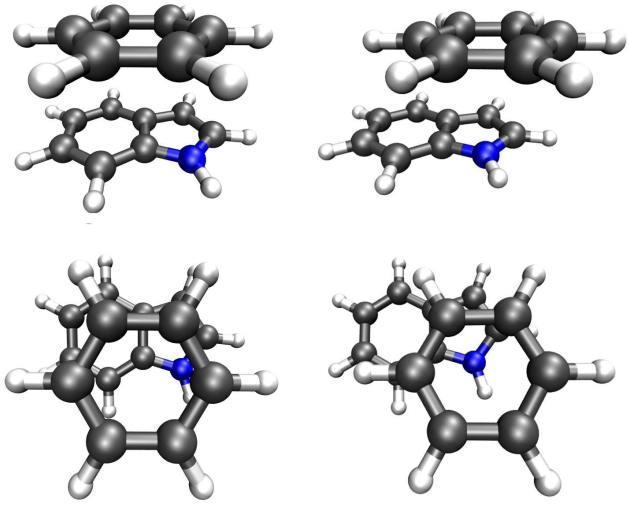


Figure 31: Representation of the minimum parallel displaced geometries predicted by the MP2 (left column) and SCS-MP2 (right column) methods.

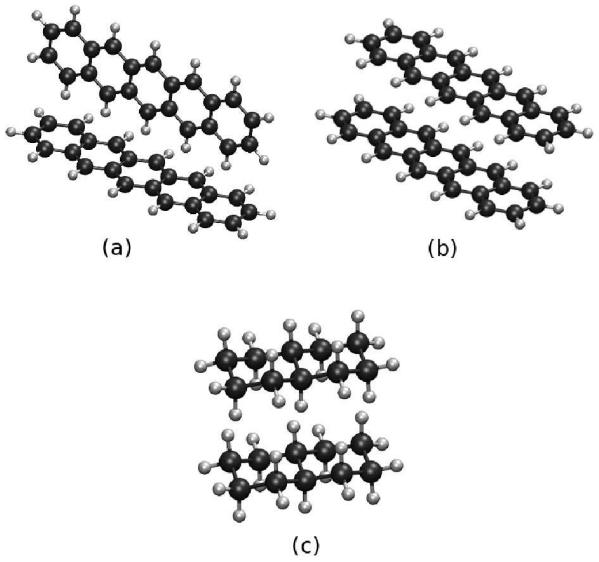
3.3 π-π interactions in linear acenes

The following was adapted from Ref. 93.

Our DF-SAPT0 program has been applied to study the differences between parallel-displaced and T-shaped configurations of linear acenes as shown in Figure 32. We will consider the $n = 1$ (benzene dimer) through $n = 5$ (pentacene dimer) cases. In order to isolate the changes in the interaction due to the additional rings, the CC and CH bond distances are held at their lengths in benzene, 1.3915 and 1.080 Å respectively.[59] Additionally, the intermolecular displacements are fixed at the values that are optimal for the benzene dimer.[198] For comparison purposes, we also considered saturated, stacked dimers beginning with the cyclohexane dimer, using the geometry of Grimme.[71] Larger saturated dimers were constructed from the cyclohexane dimer without reoptimization of geometrical coordinates (consistent with our method of constructing the aromatic dimer geometries). These constraints greatly simplify the energy component analysis because some of the terms are very sensitive to geometry and therefore difficult to compare in dimers with different intermolecular distances. The SAPT0 computations on the acene dimers were performed with a truncated aug-cc-pVDZ basis set denoted aug-cc-pVDZ′. This basis set removes all diffuse functions from hydrogen atoms and diffuse d functions from carbon atoms. In our experience, fortuitous error cancellation occurs for aromatic dimers when this basis set is

paired with MP2-like methods.[203] This basis is somewhat smaller than needed for accurate interaction energies of the saturated dimers (see below), but it should suffice for a semi-quantitative analysis of the energy components. A truncated aug-cc-pVDZ-RI fitting basis set was paired with the aug-cc-pVDZ′ orbital basis: the diffuse functions were removed from the fitting basis for hydrogen atoms and diffuse f functions were removed for carbon atoms.

Figure 32: Geometries of (a) T-shaped, (b) parallel-displaced pentacene dimer and (c) the saturated analogue of naphthalene dimer. The centers of the rings in the T-shaped dimers are separated by 5.0 Å. The parallel-displaced geometries are separated by 3.5 Å vertically and 1.7 Å horizontally. These intermolecular distances are those which are optimal for the benzene dimer.[198] The carbon atoms in the saturated dimers are all separated by approximately 4.3 Å.[71]



The results for the SAPT0 decomposition of the acene dimers are given in Table 22. The total SAPT0 interaction energies for the acene dimers show good agreement with the SCS-MP2 and B97-D values reported by Grimme.[71] For the saturated stacked dimers, the dimers are somewhat underbound (by 27-45%) compared to Grimme's MP2 results. We ascribe this difference to poorer error cancellation for the SAPT0/aug-cc-pVDZ′ level of theory for the saturated systems. Nevertheless, we believe the qualitative trends in the SAPT energy components will be reliable enough for our analysis; we will focus particularly on the dispersion energies, and our SAPT dispersion energies are quite similar to Grimme's

B2PLYP-D/TZV(2d,p) dispersion energies.[71]

Table 22: SAPT0 decomposition of the interactions within aromatic T-shaped and parallel-displaced acene dimers as well as the stacked, saturated dimers.[a]

N	E_{elst}	E_{exch}	E_{abab}[b]	E_{ind}	E_{disp}	$-\frac{C_6}{R^6}$[c]	E_{SAPT0}
Parallel-Displaced, Aromatic							
1	-1.62	6.08	-4.42	-0.67	-6.66	-4.73	-2.87
2	-2.95	9.65	-8.86	-0.90	-13.07	-8.97	-7.27
3	-4.31	13.15	-13.64	-1.13	-19.95	-13.32	-12.24
4	-5.69	16.60	-18.62	-1.38	-27.03	-17.69	-17.49
5	-7.07	20.01	-23.74	-1.66	-34.20	-22.06	-22.91
T-Shaped, Aromatic							
1	-1.74	3.19	-2.35	-0.50	-3.73	-3.24	-2.78
2	-3.03	6.09	-5.47	-0.93	-7.95	-6.57	-5.81
3	-4.25	8.94	-9.06	-1.35	-12.48	-9.98	-9.13
4	-5.46	11.77	-12.93	-1.78	-17.13	-13.40	-12.60
5	-6.66	14.58	-16.98	-2.22	-21.84	-16.83	-16.15
Stacked, Saturated							
1	-2.72	8.75	-6.36	-0.84	-6.57	-7.56	-1.37
2	-5.24	16.46	-14.46	-1.56	-13.09	-14.32	-3.43
3	-7.75	24.17	-23.77	-2.28	-19.70	-21.20	-5.55
4	-10.26	31.89	-33.83	-2.99	-26.33	-28.10	-7.69
5	-12.77	39.60	-44.38	-3.71	-32.97	-35.01	-9.85

[a]Computations performed with the aug-cc-pVDZ′ basis. Energies given in kcal mol^{-1}. [b]Contribution to the exchange energy from *abab* type elements of the intermolecular potential. [c]Atomic C_6 values taken from Reference 69.

For the acene dimers, the SAPT decomposition shows a linear increase in the E_{exch}, E_{elst}, and E_{ind} terms with the number of rings. The E_{disp} term shows a nonlinear increase that is consistent with the decomposition reported by Grimme.[71] Since the dispersion energy can be approximated as a pairwise $-C_6R^{-6}$ interaction, a nonlinear increase is expected. However, as shown in Table 23, because of the finite range of the empirical dispersion correction, it increases only linearly past anthracene dimer, whereas the quantum mechanical SAPT0 dispersion energy shows nonlinearity in all the cases considered. It is important to note that this same behavior is seen in both the parallel-displaced and T-shaped dimers. The magnitude of the dispersion energy in the parallel-displaced dimers is obviously greater due to the closer interaction between π-clouds. These observations lead to the conclusion that the intrinsic dispersion interaction between π-electrons in both configurations is the same

qualitatively. The difference in magnitude is simply a result of the separation between π-clouds. The second order dispersion energy in SAPT is computed as $4\,(ar|bs)^2\,/(\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s)$ where ϵ represents a HF orbital energy. As the separation between monomers increases, obviously, the $(ar|bs)$ integrals will decrease in magnitude.

Table 23: Changes in energy components of aromatic and saturated interactions (in kcal mol^{-1}) as the number of rings are increased.[a]

$N_1 \rightarrow N_2$	ΔE_{elst}	ΔE_{exch}	ΔE_{ind}	ΔE_{disp}	$\Delta(-C_6R^{-6})$	ΔE_{SAPT0}
Parallel-Displaced, Aromatic						
$1 \rightarrow 2$	-1.33	3.57	-0.23	-6.41	-4.24	-4.40
$2 \rightarrow 3$	-1.37	3.50	-0.23	-6.88	-4.35	-4.97
$3 \rightarrow 4$	-1.38	3.45	-0.25	-7.08	-4.37	-5.25
$4 \rightarrow 5$	-1.38	3.41	-0.29	-7.17	-4.37	-5.42
T-Shaped, Aromatic						
$1 \rightarrow 2$	-1.28	2.89	-0.43	-4.22	-3.33	-3.03
$2 \rightarrow 3$	-1.23	2.85	-0.42	-4.52	-3.41	-3.32
$3 \rightarrow 4$	-1.21	2.83	-0.43	-4.65	-3.43	-3.46
$4 \rightarrow 5$	-1.20	2.81	-0.44	-4.72	-3.43	-3.55
Stacked, Saturated						
$1 \rightarrow 2$	-2.52	7.71	-0.72	-6.53	-6.76	-2.06
$2 \rightarrow 3$	-2.51	7.71	-0.72	-6.60	-6.88	-2.11
$3 \rightarrow 4$	-2.51	7.72	-0.72	-6.63	-6.90	-2.15
$4 \rightarrow 5$	-2.51	7.72	-0.72	-6.64	-6.91	-2.15

[a] Computations performed with the aug-cc-pVDZ′ basis. Energy differences are computed as E_{N_2} - E_{N_1}.

In Tables 22 and 23, the only geometric changes that occur are the addition of rings to the monomers. This isolates the electronic effects from geometric effects. Our analysis suggests that the nonlinear increase of the dispersion interaction seen in the aromatic complexes is purely an electronic effect that originates from the interaction of large, delocalized π-orbitals in relatively close proximity. In contrast to the aromatic dimers, the saturated dimers do not show a nonlinearity in the dispersion term beyond what is expected. Moreover, the changes in all of the energy components for the saturated complexes are more linear than in the aromatic complexes.

The $n = 1$ to $n = 4$ cases were previously studied by Grimme[71] in an attempt to determine if there is anything unique about π-π interactions. That work relies upon the results from a Morokuma style energy decomposition analysis (EDA).[122,152] The specifics of the EDA implementation used by Grimme are explained in Reference 74. The EDA is benchmarked against SAPT and a discrepancy for the electrostatic and exchange terms is noted:[74] "Larger systematic differences between SAPT and EDA are observed for E_{exr} and

E_{es}, *i.e.*, the former is higher and latter is always lower in EDA. The reasons for this are presently not clear and deserve more research." When this approach is applied to acene dimers, the electrostatic term appears to be more attractive than the dispersion term by about a factor of two. This is curious since the interaction between non-polar molecules is expected to be dispersion-dominated. Due to this counter-intuitive result, Grimme uses the sum of the exchange and electrostatic terms in his analysis.

In the implementation of EDA used by Grimme, the stabilizing *abab* type two-electron integrals that enter the HF energy expression are incorrectly included in the electrostatic term instead of the exchange term. In Table 22, we show the contribution from the *abab* interaction to the $E_{exch}^{(10)}$ term separately. In the case of pentacene dimer, this stabilizing interaction is actually three times as large as the entire electrostatic term. By misplacing these contributions, the exchange and electrostatic results from Grimme's EDA appear much too large in magnitude and are difficult to reconcile physically. It should be noted that E_{abab}, reported here, is computed from complete elements of the intermolecular potential, whereas the terms misplaced in Grimme's EDA are only the two-electron contribution to E_{abab}. The EDA implementation reported in Ref. 74 could be fixed by separating the two-electron energy into coulomb and exchange contributions and adding each to the appropriate grouping. This issue is symptomatic of the specific implementation, not of the EDA outlined by Morokuma.

Because Grimme sums the exchange and electrostatic terms into a total first order interaction in his work on acene dimers, the problem with the EDA implementation did not affect the final conclusions of the paper, namely, that the increase in interaction energy with respect to system size is similar for T-shaped acenes and saturated hydrocarbons, whereas it is significantly larger in magnitude for parallel-displaced acenes. Moreover, there is a non-linear increase in the interaction energy for the parallel-displaced acene dimers. Based on these considerations, Grimme concludes that stacked aromatics feature a "special" π-π interaction, not present in saturated hydrocarbons, which results from stabilizing electron correlation terms that only become significant when the two monomers are in close contact (leading to dispersion contributions which are more favorable than would be predicted by

pairwise $-C_6R^{-6}$ terms). While our SAPT results and analysis support these conclusions, we note that the T-shaped configurations also feature a non-linear increase in interaction energy, which is not present in the saturated hydrocarbons, all the way up to the largest dimer considered (pentacene dimer). Moreover, the dispersion energies for the T-shaped dimers are also larger than predicted by pairwise $-C_6R^{-6}$ terms. Hence, we see evidence in the T-shaped configurations, as well as the parallel-displaced configurations, of "special" π-π interactions. The dispersion terms in the T-shaped configurations are certainly smaller than in the parallel-displaced configurations, but they remain about 60% as large. The difference between the SAPT dispersion energy and the empirical $-C_6R^{-6}$ estimate remains about 40% as large. On this basis, even though the interaction energy of T-shaped acenes remains similar to that of stacked, saturated hydrocarbons of the same size, in our view special π-π interactions are also present in the T-shaped configurations, albeit to a lesser (but non-negligible) degree. This leads us to conclude that the close agreement between T-shaped acenes and stacked saturated dimers in plots of the interaction energy vs. dimer size (see Figure 2 of Grimme's work[71]) is not an indication that the nature of the interactions is similar, but is an accident resulting from the very different geometries of those two sets of molecules. One could attempt to probe this hypothesis directly, by computing T-shaped saturated dimer energies, but unfortunately we could not come up with a reasonable chemical model that wold fit this description without adding too many additional short-range contacts.

We have applied our DF-SAPT code to the interactions within acene dimers. The new code is efficient enough that we were easily able to include the pentacene dimer in our tests. We determined the source of a discrepancy between SAPT and the EDA implementation previously used to study acene dimers by Grimme.[71] This difference was due to a problem with the EDA implementation. Both the T-shaped and the parallel-displaced configurations feature a non-linear increase in interaction energy with respect to system size, all the way through pentacene dimer. Moreover, both types of configurations feature dispersion energies which are significantly larger than one would predict using pairwise $-C_6R^{-6}$ terms. Although these effects are smaller in magnitude for the T-shaped configurations due to the

larger separation between the π-clouds, they do not become negligible. This suggests that special π-π interactions are present not only in parallel-displaced configurations, but also in T-shaped configurations of aromatic hydrocarbons.

3.4 Dispersion in problematic complexes: homogeneous dimers of NCCN, P_2 and PCCP

The following was adapted from Ref. 98.

3.4.1 Introduction

With the number of recently developed methods aimed at describing noncovalent interactions, it is important to have reliable and challenging benchmarks available. One of the most popular is the S22 test set of Hobza and co-workers.[117] This set of benchmark interaction energies has been used extensively to test and train new methods.[180] For most wavefunction based methods, two of the most difficult systems in this test set are the stacked benzene dimer and indole-benzene complexes.[172,213] Even the original benchmark energy for the stacked indole-benzene differs by approximately 15% from the best estimate currently available.[172,213] It is useful to study systems that contain dispersion interactions similar to the stacked aromatic π-π complexes in the S22 test set, but for which more accurate benchmarks can be established.

Here, we examine NCCN, P_2, and PCCP dimer as such model systems. All three of these systems are much smaller than the benzene dimer or indole-benzene allowing for the computation of the non-relativistic, electronic interaction energy in the complete basis set limit. We apply our new CCD+ST(CCD) SAPT program to examine the nature of the dispersion interactions present in NCCN, P_2, and PCCP dimer compared to those in stacked aromatic π-π complexes. Through this analysis, the problems encountered by finite-order perturbation theory are explored. Furthermore, we compare the molecular interaction between an extensive set of the aforementioned methods and our new benchmark data over multiple slices of the interaction potential for these three dimers.

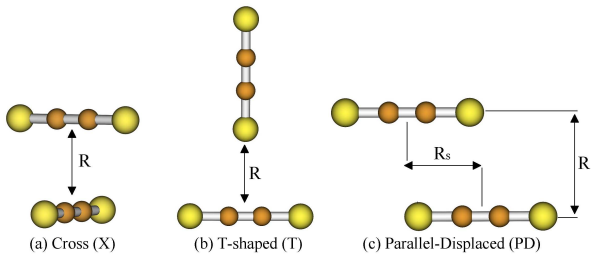
3.4.2 Theoretical Methods

Rigid, linear monomer geometries were adopted for all computations. Experimental bond lengths were taken from Herzberg[82,83] for P_2 and NCCN [R(PP) = 1.8943Å, R(CC) = 1.3839 Å and R(CN) = 1.1578 Å]. Although PCCP has been observed experimentally, structural characterization was not feasible, and no experimentally inferred geometrical parameters were reported.[24] As such, the bond lengths for PCCP used in this study [R(CC)=1.35560Å and R(CP)=1.58597 Å] were obtained from a low-level geometry optimization. These values, however, are entirely consistent with CCSD(T) optimizations with correlation consistent triple-ζ basis sets.[75]

Potential energy curves (PECs) of the dimer structures were computed in three different configurations: cross (X), parallel-displaced (PD) and T-shaped (T) that belong to the D_{2d}, C_{2h} and C_{2v} point groups, respectively. These configurations are depicted in Figure 33 for $(PCCP)_2$, but the general definitions of the intermolecular geometrical parameters also apply to $(P_2)_2$ and $(NCCN)_2$. The D_{2d} cross configuration is depicted in Figure 33(a), where the arrow indicates the intermolecular distance (R) between the mid-points of the central bond of each monomer. For the C_{2v} T-shaped structures, the arrow shown in Figure 33(b) denotes the intermolecular distance (R) from the mid-point of the central bond that is perpendicular to the C_2 rotational axis of symmetry to the nearest atom in the other monomer that lies on the C_2 axis of symmetry. The C_{2h} parallel displaced structures are defined by 2 intermolecular parameters. R is again used to indicate the separation between the monomers, specifically the distance between the two parallel lines defined by the linear monomers (denoted by the vertical arrow in Figure 33(c)). The other intermolecular geometrical parameter for the PD configurations is the displacement of the monomers along the aforementioned parallel lines relative to a rectangular (or sandwich) D_{2h} structure. In Figure 33(c), this "horizontal slip" distance is labeled R_S and denoted by the horizontal arrow. The R_S coordinate was fixed at a value of 2.80 Å for $(NCCN)_2$ and 2.31 Å for $(P_2)_2$, 2.66 Å for $(PCCP)_2$. These values roughly correspond to the average of MP2 and CCSD(T) optimized R_S parameters.

The PECs in this work were generated by scanning over R for each configuration of

Figure 33: Three prototypical dimer configurations of NCCN, PCCP and P_2 dimers included in this study.



the three homogeneous dimers. HF, MP2, CCSD and CCSD(T) electronic energies were computed at each point along the curve with the aug-cc-pVDZ, aug-cc-pVTZ and aug-cc-pVQZ basis sets. Larger, aug-cc-pV(X+d)Z type basis sets were determined to provide nearly identical results. The 1s-like core orbitals of C and N were constrained to be doubly occupied during the electron correlation computations, whereas this constraint was applied to the 1s-, 2s- and 2p-like core orbitals of P (i.e., the frozen core approximation). The electronic energies were converged to at least 1×10^{-10} E_h for the SCF and 1×10^{-8} E_h for the coupled-cluster procedures. The single point energy computations were performed using both the 2006.1 and 2010.1 versions of the `Molpro` software package.[151]

Electronic interaction energies were computed at the complete basis set (CBS) limit along the PECs for the X, PD, and T configurations of $(NCCN)_2$, $(P_2)_2$, and $(PCCP)_2$ by extrapolating the energy with respect to the cardinal number of the basis set. Within the supermolecular approach, CBS-limit interaction energies are computed by subtracting the extrapolated monomer electronic energies from the extrapolated electronic energies of the complex. Extrapolations were performed for the monomer energies in the monomer basis and for the dimer energies in the dimer basis. The electronic energy was separated into Hartree-Fock and correlation energies. HF energies were extrapolated with aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ according to a three-parameter formula.[54,77] The correlation energies were extrapolated to the CBS limit using the two-point formula of Halkier *et al.* with aug-cc-pVTZ and aug-cc-pVQZ basis sets.[76]

The simplest wavefunction based methods tested in this work are the spin-component scaled MP2 methods. These methods have been shown to be capable of accurately computing noncovalent interactions with a triple-ζ quality basis.[44,67,215] In this work, the original parameterization, SCS-MP2, and a molecular interaction specific parameterization, SCS(MI)-MP2, are tested.[44,67] The SCS- and SCS(MI)-MP2 computations in this work use the cc-pVTZ basis;[49,118] the HF and MP2 computations are performed under the DF approximation using the cc-pVTZ-JK and cc-pVTZ-RI auxiliary basis sets, respectively.[236,238] The spin-component scaled CCSD method of Takatani *et al.*, SCS-CCSD,[214] and its recent reparameterization for molecular interactions, SCS(MI)-CCSD,[169] are also tested. The SCS-CCSD method has been found to yield excellent results with large basis sets;[65] in the present work, SCS-CCSD/aug-cc-pVQZ and SCS(MI)-CCSD/aug-cc-pVQZ interaction energies are computed. The SCS parameters for these methods can be found in Table 3.4.2. The midground in terms of computation expense between SCS-MP2 and SCS-CCSD is the scaled MP3 method (MP2.5) of Pitoňak *et al.*[167] This method includes half of the third-order correction to MP2 (or, equivalently, averages MP2 and MP3 energies). Similarly to SCS-CCSD, this method performs well with large basis sets, and MP2.5/aug-cc-pVQZ interaction energies are reported. The counterpoise correction is applied to these wavefunction based methods.[21] These computations are performed with `Molpro`.[151]

Table 24: Scaling parameters for opposite-spin (S_{os}) and same-spin (S_{ss}) components of the correlation energy.

	S_{os}	S_{ss}
SCS-MP2	1.20	0.33
SCS(MI)-MP2	0.17	1.75
SCS-CCSD	1.27	1.13
SCS(MI)-CCSD	1.11	1.28

We also test the promising MP2C method.[84,165] This method is a composite of a counterpoise corrected MP2 interaction energy and dispersion energies from intermolecular perturbation theory. The uncoupled Hartree-Fock (UCHF) dispersion energy contained in MP2 is replaced with a dispersion energy computed with time-dependent density functional

theory (TDDFT) response functions.

$$E_{\mathrm{MP2C}} = E_{\mathrm{MP2}} - E_{disp}(\mathrm{UCHF}) + E_{disp}(\mathrm{TDDFT}) \tag{212}$$

The MP2C interaction energies are computed with an aug-cc-pVQZ basis and use aug-cc-pVQZ-JK and aug-cc-pVQZ-RI auxiliary basis sets. The HF and MP2 energies required for the MP2 interaction energy are computed under the DF approximation. The UCHF dispersion energy is computed with DF-HF eigenvectors and eigenvalues and is evaluated under the DF approximation. The TDDFT dispersion is computed with local HF eigenvectors and eigenvalues and the TDDFT response functions are evaluated with the adiabatic local density approximation exchange-correlation kernel.[84] For the PCCP dimer aug-cc-pVTZ, aug-cc-pVTZ-JK and aug-cc-pVTZ-RI basis sets were used. A related method (in terms of its description of dispersion) is density functional based symmetry-adapted perturbation theory, SAPT(DFT).[85,147–149,245] Here, we use PBE0 with local HF exchange to describe the monomers. One caveat of SAPT(DFT) is that the monomer DFT computations need to be asymptotically corrected in order to produce accurate interaction energies; this requires the ionization potential of the monomers, which we compute at the PBE0/TZVPP level. The SAPT(DFT) computations use the DF approximation and the same basis sets as the MP2C computations. The MP2C and SAPT(DFT) computations are performed with `Molpro`.[151]

Many DFT methods have been developed in recent years that attempt to accurately describe noncovalent interactions.[72] Here we test two hybrid meta-GGA functionals, M05-2X and M06-2X.[250,251,253] These functionals have been found to perform well when paired with the aug-cc-pVDZ and aug-cc-pVTZ basis sets, respectively.[29] Meta-GGA's are known to be susceptible to numerical errors related to the integration grid.[113] For this reason, we use a large, 100,302 (radial points, angular points) grid for the M05-2X and M06-2X computations. The ωB97X-D of method of Chai and Head-Gordon[35] and Grimme's B97-D3 method[69] are both evaluated with aug-cc-pVTZ basis sets.[29] A dense numerical integration grid was employed for the ωB97X-D computations, a pruned grid composed of 99 radial shells and 590 angular points per shell. We also test two double-hybrid DFT methods that

include a perturbative MP2-like correlation correction, specifically, B2PLYP-D3/aug-cc-pVTZ and XYG3/6-311+G(3df,2p).[29,70,249] The XYG3 functional is evaluated with B3LYP orbitals and densities.[64] Only the ωB97X-D interaction energies are counterpoise corrected. The M05-2X and M06-2X computations were performed with QChem 3.2.[195] The ωB97X-D computations were performed with the `Gaussian 09` software package.[58] The B97-D3, B2PLYP-D3, and XYG3 computations were performed with NWChem 6.0.[229]

Wavefunction-based SAPT computations were performed with a development version of the `PSI4` program.[41,93] All SAPT computations use the density fitting approximation. SAPT computations were performed with the aug-cc-pVQZ basis and use the aug-cc-pVQZ-RI auxiliary basis (with the exception of the PCCP dimer, where aug-cc-pVTZ and aug-cc-pVTZ-RI sets were used). To reduce the expense of including triple excitations, a truncated virtual space constructed from MP2 natural orbitals is used. This approximation has been shown to greatly improve efficiency without introducing significant errors.[95] A similar approximation can be applied to the evaluation of the CCD dispersion energy and will be discussed in a forthcoming publication.[100] The highest level of SAPT applied in this work is denoted SAPT2+3(CCD), which is defined as follows:

$$\begin{aligned} E_{\text{SAPT2+3(CCD)}} = {} & E^{(10)}_{elst,resp} + E^{(12)}_{elst,resp} + E^{(13)}_{elst,resp} + E^{(10)}_{exch} + E^{(11)}_{exch}(S^2) + E^{(12)}_{exch}(S^2) \\ & + E^{(20)}_{ind,resp} + E^{(20)}_{exch\text{-}ind,resp}(S^2) + E^{(30)}_{ind} + E^{(30)}_{exch\text{-}ind}(S^2) + {}^tE^{(22)}_{ind} + {}^tE^{(22)}_{exch\text{-}ind} \\ & + \epsilon^{(2)}_{disp}[\text{CCD} + \text{ST(CCD)}] + E^{(30)}_{disp} + E^{(20)}_{exch\text{-}disp} + E^{(30)}_{exch\text{-}disp} \\ & + E^{(30)}_{ind\text{-}disp} + E^{(30)}_{exch\text{-}ind\text{-}disp}. \end{aligned} \tag{213}$$

Following from Reference 162, the supermolecular HF interaction energy is not included in the SAPT energy, since the third-order treatment of induction is expected to be sufficient when nonpolar monomers are considered. Approximate exchange terms are scaled according to $E^{(10)}_{exch}/E^{(10)}_{exch}(S^2)$ in order to account for higher-order exchange effects that are neglected in the S^2 approximation.

3.4.3 SAPT Analysis of the Dispersion Energy

The SAPT computations on the NCCN, PCCP, and P_2 dimers allow the dispersion energy to be analyzed separately from the total interaction energy. The dependence of interaction energies on the theoretical treatment of dispersion can be estimated from the relative importance of the dispersion component. To identify which systems would most likely be sensitive to the treatment of dispersion, Table 25 shows the magnitude of the dispersion energy relative to the total SAPT2+3(CCD) interaction energy computed at the estimated CCSD(T)/CBS limit equilibrium geometry. In the cross configurations of the NCCN, PCCP, and P_2 dimers, for example, the magnitude of the dispersion energy is 2–3 times larger than the total interaction energy. The relative contribution from dispersion is appreciably smaller for the PD and T-shaped configurations of the NCCN dimer, but remains large for P_2 and PCCP dimers. From this simple analysis, one would expect the PCCP and P_2 dimers to be more sensitive to the treatment of dispersion than the NCCN dimer.

Table 25: The magnitude of the dispersion energy relative to the total SAPT2+3(CCD) interaction energy at estimated CCSD(T)/CBS limit equilibrium geometries.

	NCCN	PCCP	P_2
Cross	333%	205%	240%
PD	116%	216%	250%
T-shaped	96%	194%	234%

A more detailed analysis of the dispersion energy in these complexes can be found in Table 26. Here, the dispersion energy is reported as computed at various truncations of the MBPT expansion. For the moment, we will consider only the $E_{disp}(2)$, $E_{disp}(4)$, and E_{disp}(CCD) treatments of the dispersion energy. The $E_{disp}(2)$ term is an MP2-like (UCHF) dispersion energy. The $E_{disp}(4)$ term contains perturbative intramonomer correlation corrections to dispersion through second-order. The E_{disp}(CCD) dispersion is the most reliable, and uses CCD wavefunctions to correct the dispersion energy for intramonomer correlation. The dispersion energies are also presented as a percentage of the $E_{disp}(2)$ dispersion energy. These percentages can be used as a means of gauging how difficult the dispersion energies

in a given complex are to compute. For simple systems, there will be little change between $E_{disp}(2)$ and $E_{disp}(\mathrm{CCD})$. For more difficult systems, there will be a significant difference between $E_{disp}(2)$ and $E_{disp}(4)$, but not $E_{disp}(4)$ and $E_{disp}(\mathrm{CCD})$. For the most difficult systems, there will be large differences between all three of these treatments of dispersion.

Table 26: Dispersion energies computed with various levels of SAPT.[a]

	$E_{disp}(2)^b$	$E_{disp}(3)^c$	$E_{disp}(2.5)^d$	$E_{disp}(4)^e$	$E_{disp}(\mathrm{CCD})^f$
$(NCCN)_2$ Cross	-2.3	-1.5 (65%)	-1.9 (83%)	-2.0 (85%)	-1.9 (81%)
$(NCCN)_2$ PD	-2.9	-1.8 (65%)	-2.4 (82%)	-2.5 (87%)	-2.3 (82%)
$(NCCN)_2$ T-shaped	-2.5	-1.6 (66%)	-2.1 (83%)	-2.3 (92%)	-2.2 (87%)
$(PCCP)_2$ Cross	-8.0	-3.9 (49%)	-6.0 (74%)	-6.3 (79%)	-5.8 (72%)
$(PCCP)_2$ PD	-7.9	-3.7 (47%)	-5.8 (73%)	-6.5 (82%)	-5.8 (73%)
$(PCCP)_2$ T-shaped	-4.7	-2.6 (55%)	-3.7 (77%)	-4.0 (85%)	-3.7 (78%)
$(P_2)_2$ Cross	-3.2	-1.9 (58%)	-2.5 (79%)	-2.4 (74%)	-2.4 (75%)
$(P_2)_2$ PD	-3.6	-2.1 (58%)	-2.8 (79%)	-2.8 (78%)	-2.8 (78%)
$(P_2)_2$ T-shaped	-2.0	-1.2 (58%)	-1.6 (79%)	-1.7 (82%)	-1.6 (79%)
$(CH_4)_2$	-1.3	-1.2 (97%)	-1.2 (99%)	-1.3 (103%)	-1.3 (106%)
CH_4-Bz	-2.9	-2.4 (82%)	-2.7 (91%)	-2.7 (93%)	-2.7 (93%)
$(Bz)_2$ PD	-7.9	-5.2 (65%)	-6.6 (83%)	-6.7 (84%)	-6.5 (82%)

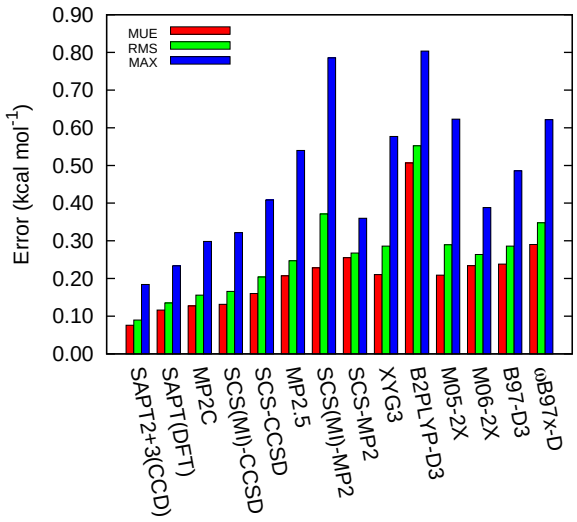
[a] Dispersion energies are given in kcal mol^{-1} and as percentages of $E_{disp}(2)$. [b] $E_{disp}(2) = E_{disp}^{(20)}$. [c] $E_{disp}(3) = E_{disp}^{(20)} + E_{disp}^{(30)} + E_{disp}^{(21)}$. [d] $E_{disp}(2.5) = E_{disp}^{(20)} + \frac{1}{2}(E_{disp}^{(30)} + E_{disp}^{(21)})$. [e] $E_{disp}(4) = E_{disp}^{(20)} + E_{disp}^{(30)} + E_{disp}^{(21)} + E_{disp}^{(22)}$. [f] $E_{disp}(\mathrm{CCD}) = E_{disp}^{(2)}(\mathrm{CCD}) + E_{disp}^{(22)}(\mathrm{ST})(\mathrm{CCD})$.

For the sake of comparison with more commonly studied systems, methane dimer, methane-benzene, and benzene dimer dispersion energies are also presented in Table 26. For typical interactions, such as these, the $E_{disp}(4)$ treatment of dispersion has been found previously to be quite reliable.[94] The dispersion interactions between alkanes can be treated accurately with $E_{disp}(2)$. The interactions involving π orbitals are more difficult and require, at least, the $E_{disp}(4)$ treatment of dispersion. Dispersion within the NCCN dimer is comparable to the dispersion in the benzene dimer. There is a large difference between $E_{disp}(2)$ and $E_{disp}(4)$ and a small difference between $E_{disp}(4)$ and $E_{disp}(\mathrm{CCD})$. The P_2 dimer differs in that there is an even larger difference between $E_{disp}(2)$ and $E_{disp}(4)$, but the $E_{disp}(\mathrm{CCD})$ correction is unimportant. The PCCP dimer has the most difficult dispersion interactions to compute. In this case, not only is the difference between $E_{disp}(2)$ and $E_{disp}(4)$ large, but the difference between $E_{disp}(4)$ and $E_{disp}(\mathrm{CCD})$ can be as large as 0.7 kcal mol^{-1}. This analysis shows that only the most robust methods will be capable of accurately computing dispersion interactions within the PCCP dimer.

3.4.4 Performance of Supermolecular Methods

The methods tested in this work can be grouped in terms of their treatments of the dispersion interaction. We test four methods that contain highly parameterized dispersion corrections (M05-2X, M06-2X, B97-D3, and ωB97X-D). We test four methods that contain empirically corrected MP2-like dispersion terms (SCS-MP2, SCS(MI)-MP2, XYG3, and B2PLYP-D3). The other methods compute dispersion with more robust techniques than MP2; MP2.5 contains contributions from MP3, SCS- and SCS(MI)-CCSD contain a CCSD treatment of dispersion, and MP2C and SAPT(DFT) use TDDFT-based dispersion corrections.

Figure 34: Errors for NCCN, PCCP and P_2 dimers computed at equilibrium with various methods.



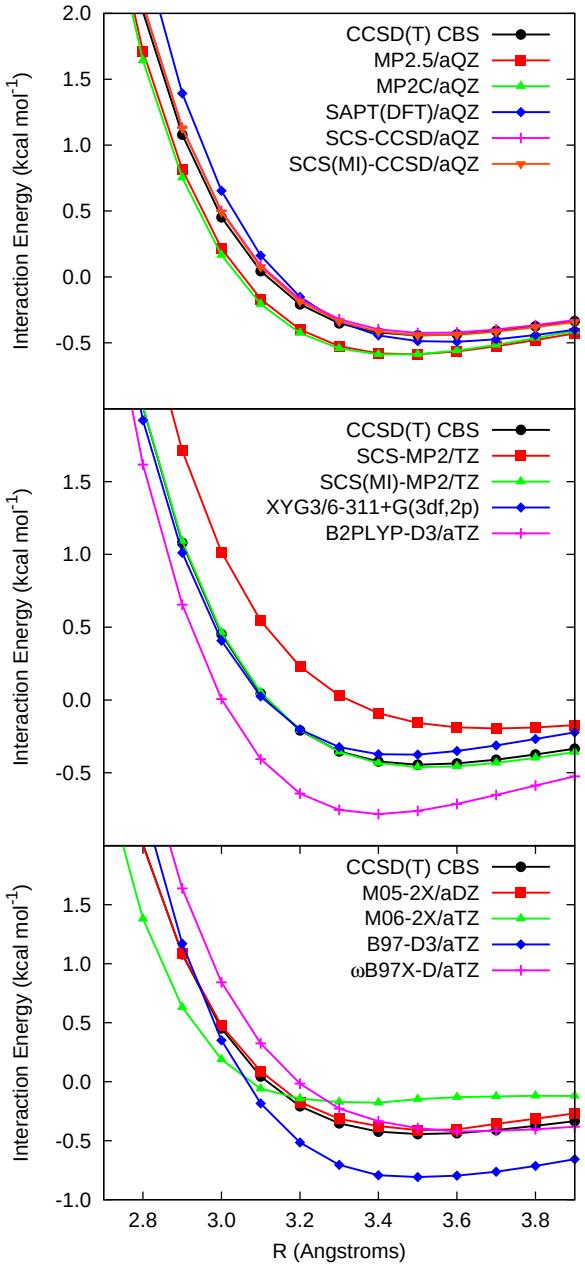
First, we will consider the M05-2X and M06-2X density functionals. Although these functionals do not contain the correct physics to describe long-range dispersion interactions, through extensive parameterization they appear to capture "medium-range" dispersion (up to perhaps 4-5 Å).[29,91,198,221,251,254] This deficiency is evident in Figures 35-43; M05-2X tends to predict reasonable interaction energies near equilibrium and underbinds at long range. M06-2X often predicts reasonable interaction energies but equilibrium distances that are too short; at long-range, the performance of M06-2X degrades quickly. The PD PCCP

dimer (Figure 43) is an example of typical behavior for M05-2X and M06-2X. Although lacking long-range dispersion interactions, these functionals can provide reasonable interaction energies for small, closely interacting complexes with relatively low computational expense. In Figure 34, the average errors of each method are presented for the equilibrium configurations of the nine dimers considered in this work. For these equilibrium geometries, the performance of M05-2X and M06-2X is slightly better than the DFT-D methods. Their performance for the difficult dispersion interactions included in this work is similar to their performance for less difficult dispersion bound systems included in other test sets (e.g. the S22 and NBC10 test sets).[29] The fact that the accuracy of these functionals does not degrade for more difficult systems is a desirable characteristic. These functionals can be a practical solution for studying near-equilibrium configurations of dispersion bound complexes.

The B97-D3 and ωB97X-D methods rely entirely on pairwise C_6R^{-6} terms to account for dispersion. Therefore, the accuracy of these functionals for dispersion bound complexes is tied to the empirical parameterization of these -D terms. B97-D3 uses C_6 coefficients that are provided some information about the chemical environment of each atom.[73] This is done through a rather ingenious atom typing procedure that is completely black-box and varies continuously with the molecular geometry. The performance of these DFT-D methods at equilibrium is not as good as the M0N-2X methods; however, at long-range, the behavior of these functionals improves rather than degrading (see Figures 35-43). ωB97X-D tended to be underbound relative to the CCSD(T) benchmark, which could be indicative of C_6 coefficients that were not optimal for these highly polarizable molecules. Overall, the B97-D3 functional outperforms ωB97X-D for the dimers considered in this work. This is a useful result, since the B97-D3 functional exhibits $\mathcal{O}(N^3)$ scaling as opposed to the $\mathcal{O}(N^4)$ scaling of the hybrid ωB97X-D method.

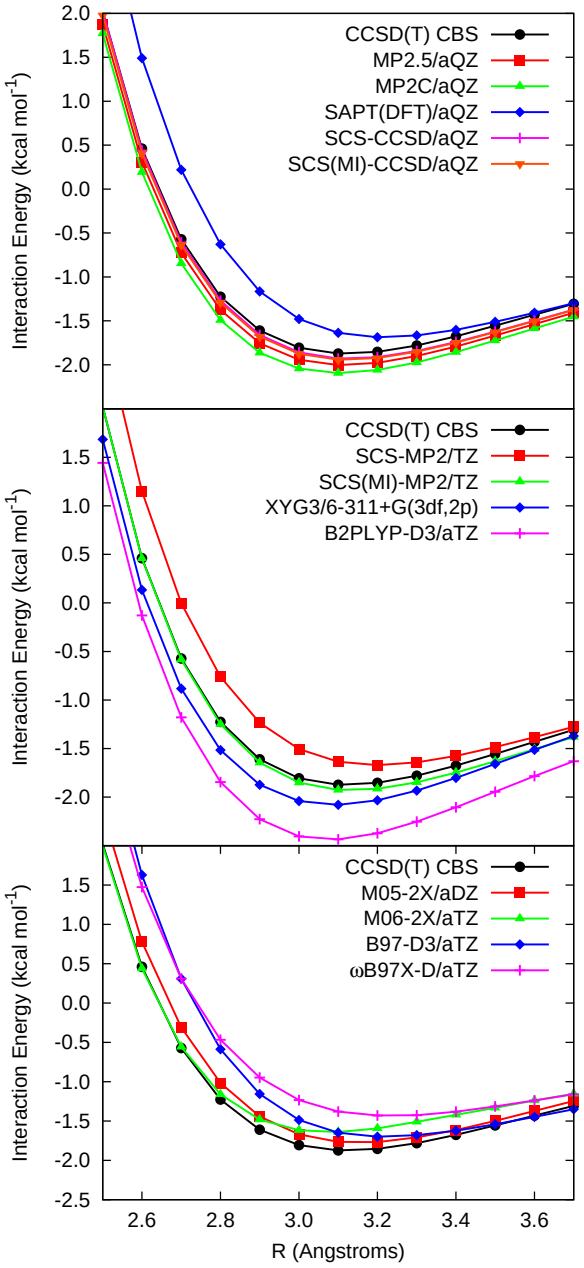
Two promising double hybrid density functionals, XYG3 and B2PLYP-D3, were applied to the NCCN, P_2, and PCCP dimers. XYG3 accounts for dispersion with an MP2-like term that is evaluated using B3LYP orbitals and scaled by 0.3211. This small scaling parameter is required because short-range correlation is included elsewhere in the functional; additionally, the DFT orbitals have, relative to Hartree-Fock orbitals, a smaller HOMO-LUMO

Figure 35: Cross NCCN dimer potential energy curves computed with various methods.



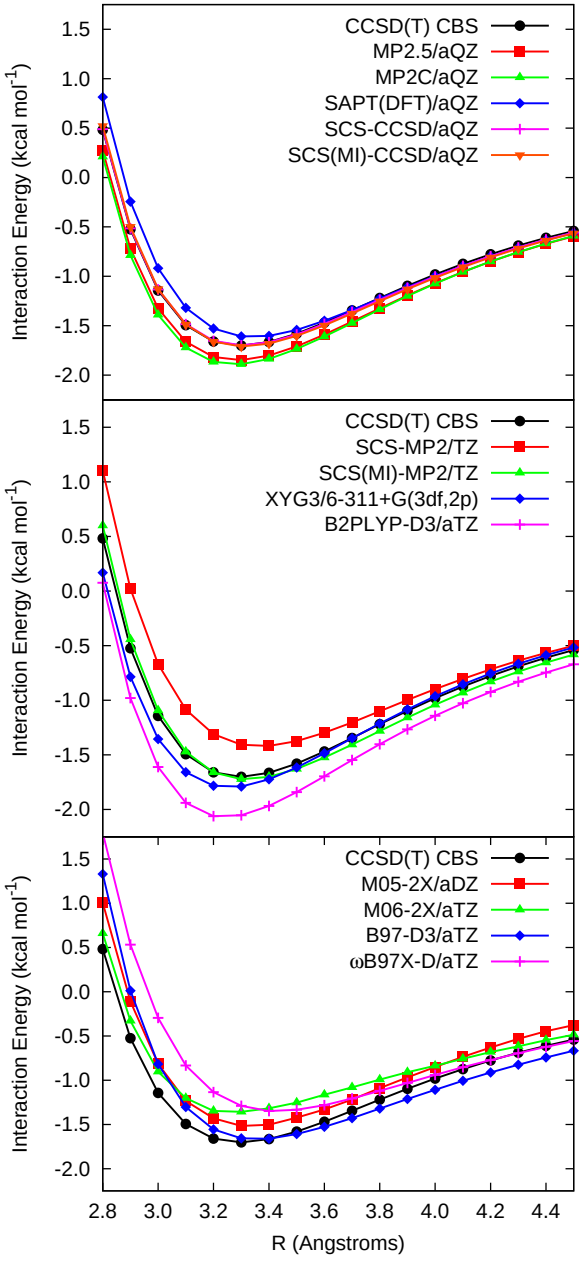
gap, resulting in a larger perturbative correction. The B2PLYP functional contains a similar scaled perturbative correction, but still requires a small empirical dispersion correction in order to provide accurate results for dispersion bound complexes. The results for B2PLYP-D3 for these dimers is very poor. Perhaps the limited data set used to parameterize this

Figure 36: T-shaped NCCN dimer potential energy curves computed with various methods.



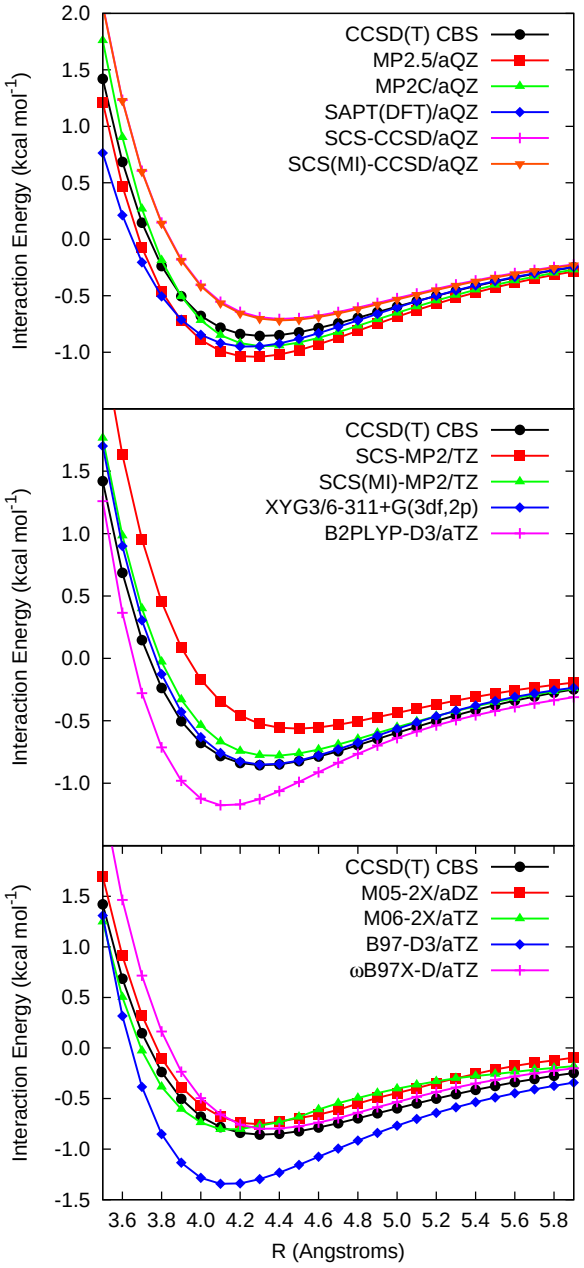
functional, which contains no third row elements, contributes to this failing. This poor performance is unusual, as the B2PLYP-D3 method is typically reliable for treating dispersion dominated interactions.[29] XYG3 performs as well as any DFT based method that was tested; this is consistent with previous findings for other test sets.[29,231] Unfortunately, it

Figure 37: Parallel-displaced NCCN dimer potential energy curves computed with various methods.



merely equals the performance of M05-2X and M06-2X, which both scale $\mathcal{O}(N^4)$ whereas XYG3 scales $\mathcal{O}(N^5)$ due to the perturbative correction. An important observation is that XYG3 does not inherit the poor performance of MP2 in its own perturbative correction,

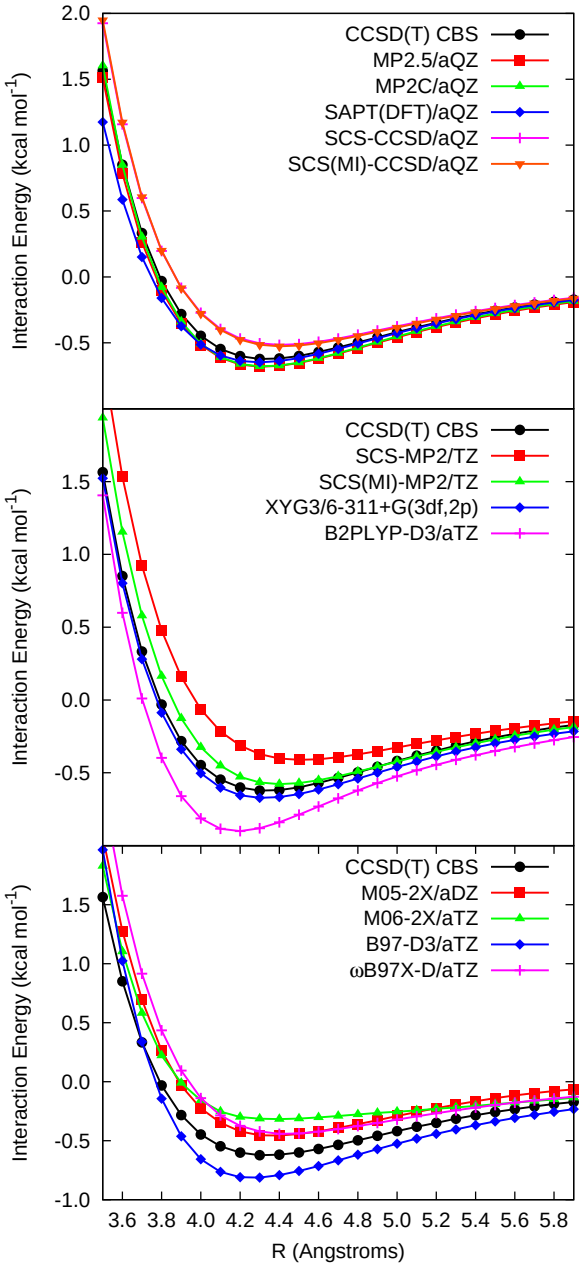
Figure 38: Cross P_2 dimer potential energy curves computed with various methods.



even for these particularly difficult cases. This can be attributed to the small fraction of the perturbative correction that needs to be included.

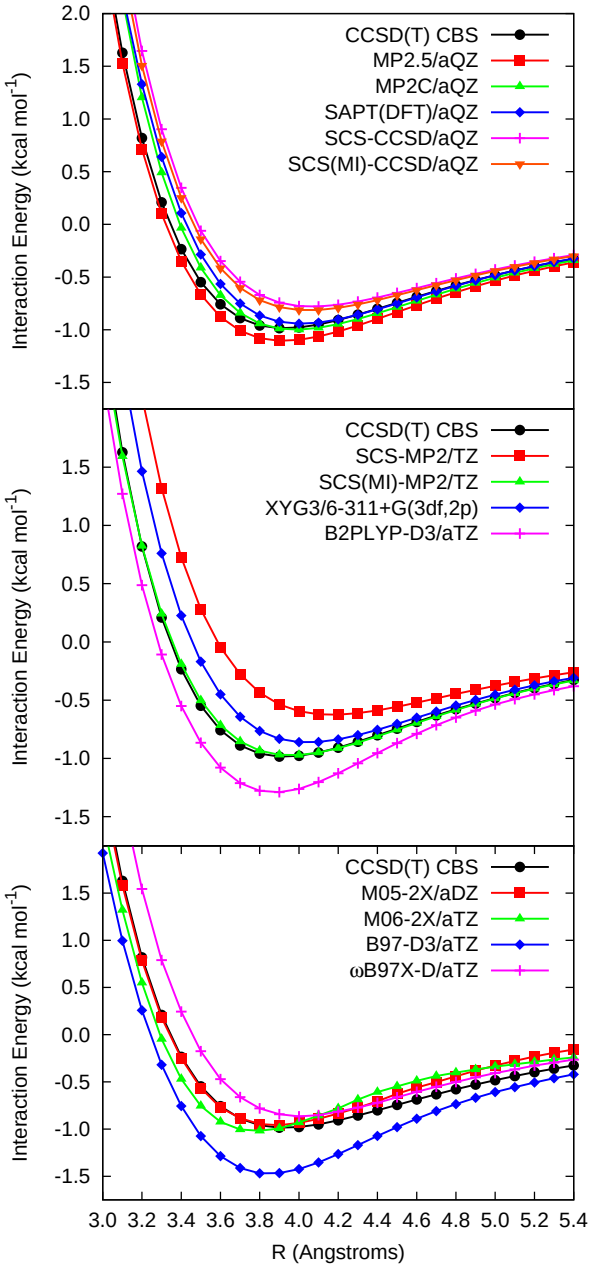
The SCS-MP2 methods we tested offer tremendous improvement over unscaled MP2

Figure 39: T-shaped P_2 dimer potential energy curves computed with various methods.



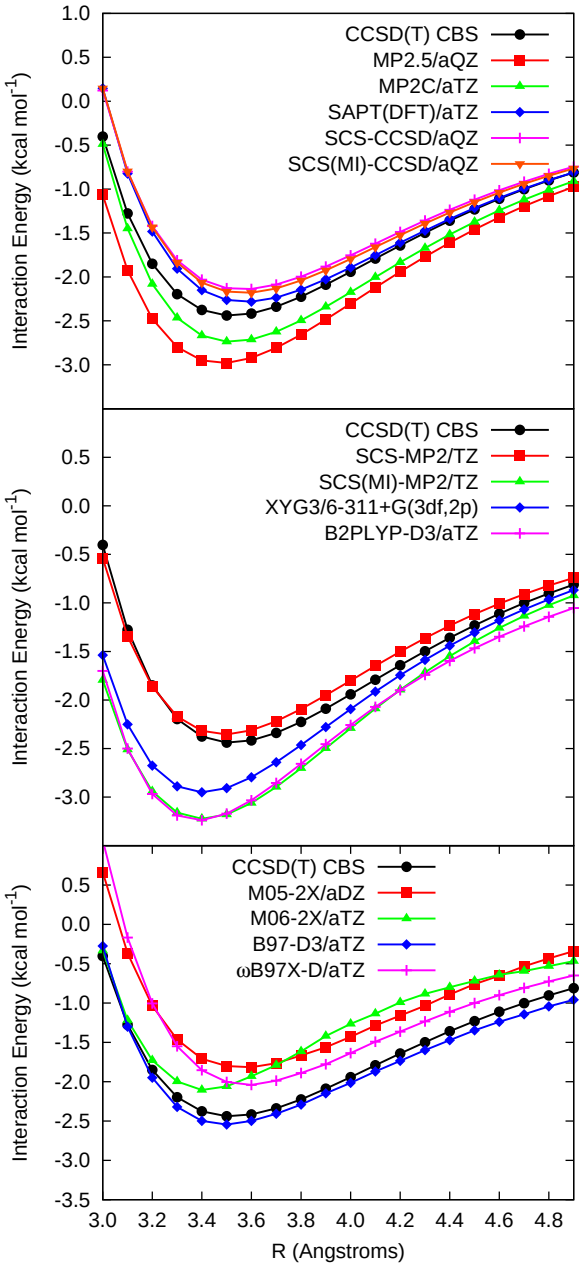
interaction energies. However, their performance is erratic; SCS(MI)-MP2, which was parameterized against the S22 test set, performs brilliantly for NCCN and P_2 dimers, but severely overbinds the PCCP dimer. The SCS-MP2 method is more consistent, but underbinds every dimer considered. It is probable that SCS-MP2/aug-cc-pVTZ would provide

Figure 40: Parallel-displaced P_2 dimer potential energy curves computed with various methods.



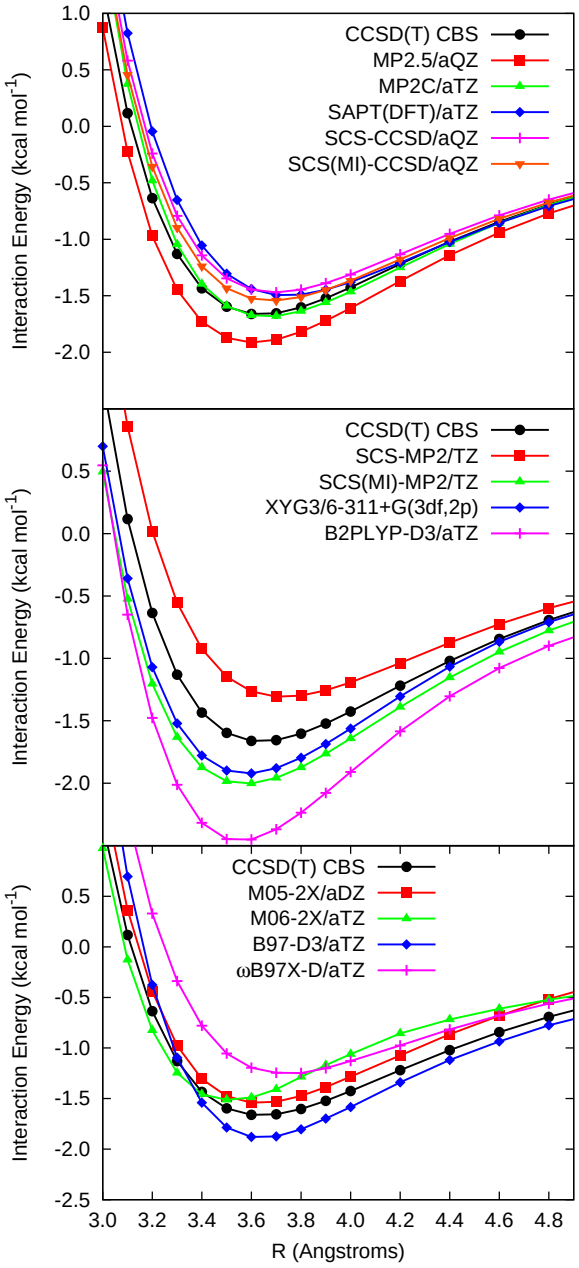
better results than SCS-MP2/cc-pVTZ, which is tested here. However, SCS-MP2 methods have an underlying problem that prevents any single parameterization from providing accurate results for a wide variety of dispersion bound complexes. The leading dispersion term,

Figure 41: Cross PCCP dimer potential energy curves computed with various methods.



$E_{disp}^{(20)}$, which is included in MP2, is composed of $\frac{1}{2}$ same-spin and $\frac{1}{2}$ opposite-spin correlation (assuming a closed shell reference). For an SCS-MP2 method, this means that the scaling of the dispersion term is effectively an average of the same-spin and opposite-spin parameters,

Figure 42: T-shaped PCCP dimer potential energy curves computed with various methods.

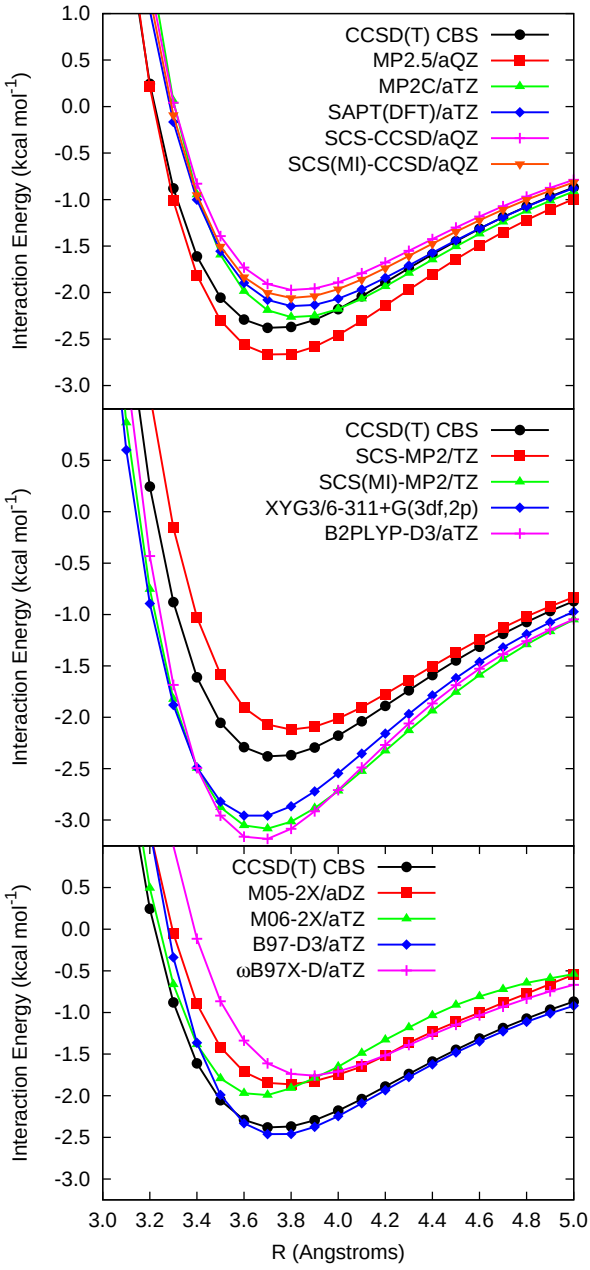


i.e.,

$$E_{disp}^{(20)}(\text{SCS-MP2}) = \frac{S_{os} + S_{ss}}{2} E_{disp}^{(20)}(\text{MP2}). \tag{214}$$

With this in mind, the percentages of $E_{disp}^{(20)}$ in Table 26 given for $E_{disp}(\text{CCD})$ represent

Figure 43: Parallel-displaced PCCP dimer potential energy curves computed with various methods.



nearly ideal values of $(S_{os} + S_{ss})/2$ for each dimer considered. For the dimers in Table 26, the ideal value for $(S_{os} + S_{ss})/2$ ranges from 1.06 to 0.72. In SCS-MP2 and SCS(MI)-MP2 these averaged values are 0.77 and 0.96, respectively. The use of a non-augmented basis set,

cc-pVTZ, essentially has the same effect as scaling the dispersion energy. The realization that the dispersion energy is scaled by a single parameter shows that it is not possible to find one set of SCS parameters that will provide consistent results for a diverse set of nonbonded interactions.

The problems encountered by SCS-MP2 methods are a result of the inconsistent behavior of MP2 for treating dispersion interactions. The SCS-CCSD methods circumvent this problem by attempting to correct a method that behaves in a much more consistent manner. Previous benchmarking of SCS-CCSD has shown that the only drawback is the $\mathcal{O}(N^6)$ scaling of CCSD; admittedly, this is a considerable limitation. The tests of SCS- and SCS(MI)-CCSD for NCCN dimer were consistent with previous results, indeed, the SCS-CCSD methods provide interaction energies within the uncertainties of the CCSD(T) benchmarks. For the P_2 and PCCP dimers, however, their performance was not as good. Both parameterizations of SCS-CCSD consistently underbind both of these dimers. Despite this slight problem, the overall performance of both SCS-CCSD methods was still excellent. The new SCS(MI) parameterization outperforms the original SCS-CCSD parameters for nearly all of the test cases, but, on average, only by a few hundredths of one kcal mol^{-1}. The parameters themselves are very similar; the original opposite- and same-spin scaling parameters of 1.27 and 1.13, respectively, and the SCS(MI) parameters of 1.11 and 1.28. The apparent insensitivity of SCS-CCSD methods to the choice of parameters is an obvious advantage of SCS-CCSD over SCS-MP2 and a desirable trait in general for an SCS method to possess.

In a similar spirit to SCS-MP2 and SCS-CCSD, MP2.5 attempts to correct the behavior of MP2 and MP3 by combining their energies in an empirical manner. MP2.5 is an average of MP2 and MP3 interaction energies; equivalently, it is also MP2 plus half of the third-order correction. The motivation for this method is obvious from the results in Table 26. The third-order dispersion energy [$E_{disp}(3)$] is always higher than the second-order dispersion energy [$E_{disp}(2)$]. In the cases where the second-order dispersion energy is a good estimate of the CCD dispersion energy [E_{disp}(CCD)], the third-order correction to the dispersion energy is small. Where there is a large difference between the second-order dispersion energy and

the CCD dispersion energy, there is an even larger difference between the second-order and third-order dispersion energies. We also report the second-order dispersion energy with half of the third-order correction included [$E_{disp}(2.5)$]. There is excellent agreement between $E_{disp}(2.5)$ and our best estimate of the dispersion energy, E_{disp}(CCD); this is the origin of the excellent performance of MP2.5 for dispersion bound complexes. For the complexes considered in this work, the performance of MP2.5 is slightly better than SCS-MP2 methods and slightly worse than SCS-CCSD methods. MP2.5 tends to overbind somewhat relative to the CCSD(T) benchmark. For more typical interactions, the simple averaging of MP2 and MP3 in the MP2.5 method provides rather accurate results.[65] For these more difficult dispersion bound complexes, a larger fraction of MP3 would need to be included for similar accuracy.

The least empirical method tested for these complexes is the MP2C method of Hesselmann.[84,165] This method uses TDDFT response functions to correct the account of dispersion in MP2. The initial tests of this method have been very promising.[65,165] Here, we apply MP2C to more difficult systems than were included in the previous tests. The performance of this method, however, does not degrade for these systems. The performance of MP2C is slightly better than the SCS-CCSD methods at greatly reduced cost. A related method, in terms of the treatment of dispersion, SAPT(DFT) also performs extremely well for these complexes. These methods perform better than any of the other approximate methods tested in this work. Additionally, both MP2C and SAPT(DFT) scale $\mathcal{O}(N^5)$, which is better than or equivalent to all but the DFT methods (excluding the double hybrids). SAPT(DFT) is slightly better than MP2C, and this is likely due to a more accurate treatment of dispersion and a more rigorous treatment of exchange-dispersion. Both of these methods contain a treatment of the dispersion energy that does not degrade for more difficult systems. They can both be recommended as generally applicable for the study of nonbonded interactions.

3.4.5 Conclusions

We have introduced the NCCN, P_2, and PCCP dimers as model systems for dispersion dominated π-π interactions. These dimers involve dispersion interactions that are problematic to study computationally. Dispersion interactions of this type are typically found in much larger complexes (*e.g.* benzene dimer or indole-benzene). SAPT computations show that the dispersion in NCCN dimer is similar in nature to the dispersion in the benzene dimer, while dispersion in P_2 and PCCP dimers is potentially more difficult to accurately characterize. The advantage of these model systems lies in their relatively small size; near complete basis set limit CCSD(T) interaction energies were obtained for NCCN, P_2, and PCCP dimers.

The performance of many recently developed quantum mechanical methods was tested for the NCCN, P_2, and PCCP dimers. DFT based methods provide reasonable results at relatively low computational expense. An advantage of these methods is that their performance does not degrade for these more difficult systems. Spin-component scaled methods perform much better than their unscaled, parent methods; although generally good, the performance of SCS methods can be somewhat erratic. The best results for the NCCN, P_2, and PCCP dimers come from the methods that use TDDFT response functions to describe dispersion interactions, MP2C and SAPT(DFT). These methods provide accuracy that could otherwise only be achieved through the inclusion of the effect of triple excitations on the dispersion energy. However, there are some limitations for general applicability of MP2C and SAPT(DFT) due to their roots in intermolecular perturbation theory (*i.e.* the need to fragment the system and the current lack of analytic gradients). Despite this, both methods are very promising for accurately characterizing π-π dispersion interactions in extended systems.

3.5 The Influence of Electrostatics on Substituent Effects in π-π Interactions

The following was adapted from Ref. 97.

Understanding how π-π interactions can be modified by substituents is of fundamental

importance for advances in drug design, exploration of non-natural nucleic acid analogs, and crystal engineering of organic materials. The conventional wisdom, inferred from numerous experiments, is encoded in the Hunter-Sanders rules,[102] which state that substituent effects can be understood in terms of how substituents change electrostatic energies by donating or withdrawing electron density from the π cloud. Theoretical studies of gas-phase mono-substituted benzene dimers, however, indicated that *both* electron-donating and electron-withdrawing substituents increase the attraction between two benzenes in an idealized face-to-face (sandwich) orientation,[201] contradicting this picture. Moreover, symmetry-adapted perturbation theory (SAPT)[111] analysis indicates that for substituents like methyl, the majority of the substituent effect is due to dispersion, not electrostatics.[201] T-shaped and parallel-displaced configurations are more prevalent in actual chemical systems, and some studies have explored their substituent effects as well.[10,134,194,202,241] Nevertheless, near-sandwich configurations are observed in some model systems,[37,39,141,243] and the sandwich structures have remained popular in theoretical studies because they are simpler geometrically.

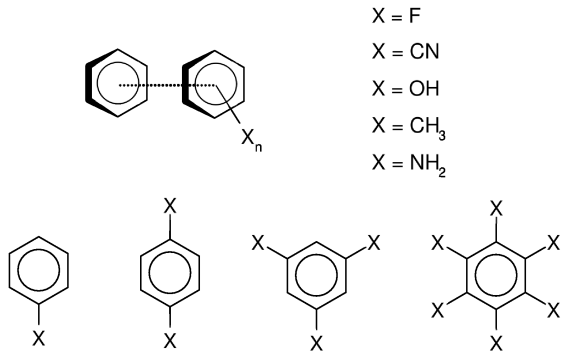
Wheeler and Houk made the amazing discovery that computed substituent effects in sandwich $C_6H_5X \cdots C_6H_6$ complexes are essentially the *same* as they are in $HX \cdots C_6H_6$ complexes in corresponding geometries.[240] That is, the substituent effects are due to direct substituent-π interactions, and not to tuning of the π-π interaction itself. For an expanded collection of substituents, a good correlation was observed between stabilization due to substituent and the Hammett parameter σ_m of the substituent, suggesting that electrostatic effects do determine the trend with respect to substituents, while dispersion merely serves to shift the trend line down from the origin (stabilizing even complexes with electron-donating substituents, in agreement with earlier theoretical studies).

However, if dispersion effects are noticeable in mono-substituted benzene dimers, then they should become large in multiply-substituted benzene dimers. Indeed, Ringer and Sherrill[184] showed that the correlation between relative binding energy and $\sum \sigma_m$ is destroyed for multiply-substituted sandwich dimers. Moreover, several monomers with radically different electrostatic potentials all exhibited similar binding to benzene. This data appears

to support the hypothesis that differential dispersion effects can be large in multiply substituted benzene dimers, and that electrostatics effects alone are not sufficient to understand substituent effects in sandwich benzene dimers. However, energy decompositions were not reported.

While an energy component analysis was being performed in our laboratory, a similar study was published by Lewis and co-workers.[235] Their study found a good correlation between computed binding energies and a model containing σ_m and M_r values, where M_r is the molar refractivity, taken to describe the polarizability of a substituent (which should be proportional to its dispersion contribution). However, SAPT analysis indicated that the sum of dispersion, exchange-repulsion, and induction was relatively constant for the substituted dimers (variations of a few tenths of one kcal mol^{-1}), whereas changes in the electrostatic term were much larger. Counterintuitively, the electrostatic term itself was found to be more favorable than in benzene dimer for *all* substituents. While increased binding for electron-donating substituents is easy to rationalize as arising from dispersion terms, it is not obvious how electron-donating substituents could lead to increased binding in the electrostatic term itself. Here we explain this surprising result.

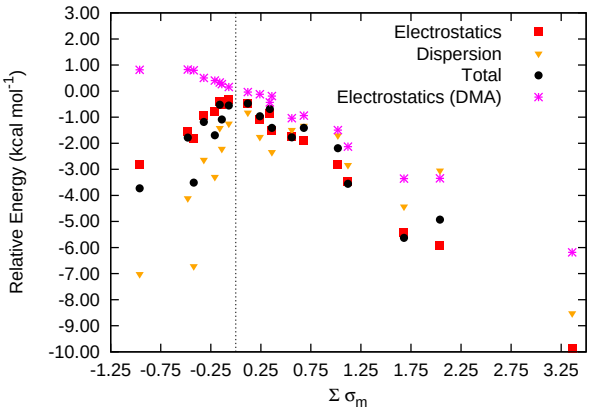
Figure 44: Depiction of the substituted sandwich benzene dimers considered; only one of the benzene rings is substituted, according to the substitution pattern displayed. Only hexahydroxybenzene is omitted.



SAPT0/aug-cc-pVDZ′ interaction energies were computed for substituted sandwich benzene dimers using a development version of the PSI4 program.[41,93] This level of theory can

accurately predict benzene dimer interaction energies.[94] The benzene and substituted benzene monomer geometries are optimized at the MP2/aug-cc-pVDZ level of theory using Q-Chem 3.2.[195] The sandwich configurations of the substituted benzene dimers from Reference 184 are considered and are aligned on the geometric center of the benzene rings. The specific substituents and geometries considered are shown in Figure 44.

Figure 45: SAPT electrostatic, dispersion, and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at their respective equilibrium geometries. For comparison, relative electrostatic energies computed from a distributed multipole analysis (DMA) are also included.



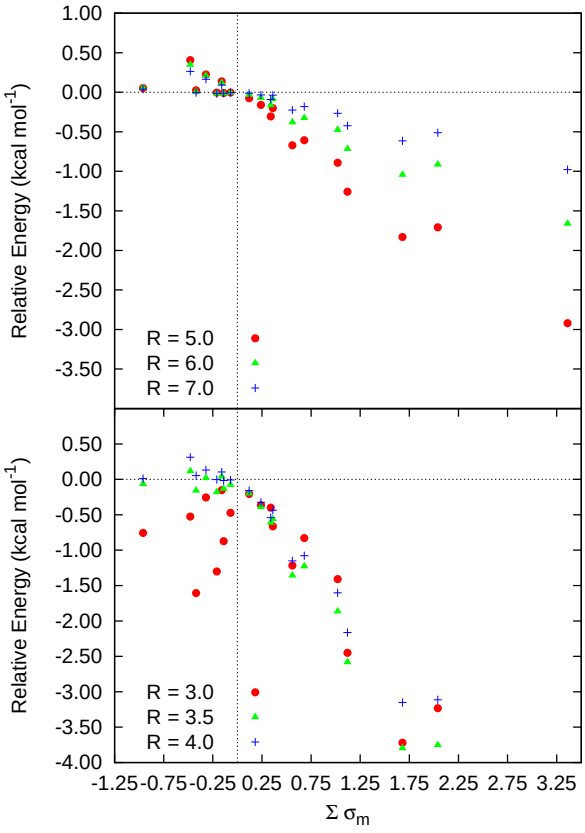
In Figure 45, we present the SAPT electrostatic, dispersion, and total interaction energies of each substituted benzene dimer relative to the unsubstituted benzene dimer, each at their equilibrium geometries. We also include relative electrostatic energies estimated by a distributed multipole analysis (DMA), to be discussed below. The reader may notice that the relative energies and dispersion energies for dimers with $\sum \sigma_m < 0$ appear to form two separate trend lines; one line results from dimers various numbers of -NH_2 substituents, and the other line results from dimers various numbers of -CH_3 substituents (substituent effects in sandwich dimers tend to be additive[185]). Dimers with more total electron-withdrawing character (positive $\sum \sigma_m$), have stronger interactions, and the electrostatic energies behave similarly to the total interaction energies for these substituents. These results are in accord with the Hunter-Sanders rules.[102] Dimers with more electron-donating character are also

more strongly bound, in agreement with Reference 184 but in contradiction to the Hunter-Sanders rules. This increase in binding is consistent with the stronger dispersion terms exhibited by all substituents (which grow with the number of substituents), as anticipated by Reference 184. The surprising result is that the SAPT electrostatic energy itself also becomes more attractive with more strongly electron-donating substituents, as reported by Lewis and co-workers.[235]

In order to understand the origin of this curious effect, the multipole picture of electrostatics must be abandoned. As monomer electron densities begin to overlap, charge penetration effects become important.[210] These are attractive electrostatic interactions due to the interaction of the electrons of one monomer with the nuclei of the other, and they increase with orbital overlap. At long range, the sandwich benzene dimer has repulsive electrostatics due to unfavorable quadrapole-quadrapole interactions. At short range, however, it has attractive electrostatic interactions due to charge penetration. Any typical multipole-based description of the electrostatics in the benzene dimer would incorrectly predict repulsive electrostatics at short range. To demonstrate the limitations of the multipole model explicitly, a distributed multipole analysis (DMA)[208] was performed on Hartree-Fock/6-311G** densities for all the monomer geometries considered here, using the Molpro program.[151] Using an in-house program developed for the purpose, electrostatic energies based on the multipole analysis were computed for the unsubstituted and substituted benzene dimers, including terms through quadrapole-quadrapole. Figure 45 compares the DMA-predicted electrostatic energies vs. the more rigorous SAPT electrostatic energies. We clearly see that the DMA electrostatic energies fail to capture the charge penetration terms that become important at the equilibrium intermolecular separations.

The fact that all substituents lead to an increased electrostatic interaction at equilibrium separations is a direct result of increased charge penetration. Any substituent increases the spatial extent of the substituted benzene's electron density and thus increases the overlap with the unsubstituted benzene (with the possible exception of fluorine). Therefore, both electron donating and electron withdrawing substituents increase the electrostatic interaction as a result of increased charge penetration.

Figure 46: SAPT electrostatic energies of substituted benzene dimers relative to the unsubstituted benzene dimer at fixed intermolecular displacements (given in Ångströms).



The effect of charge penetration on the results in Figure 45 is, perhaps, exaggerated because all substituted benzene dimers have shorter intermolecular distances than the unsubstituted benzene dimer. To simplify the analysis, in Figure 46 we present relative electrostatic energies at various fixed intermolecular distances. At long range, 5 to 7 Å, the electrostatic interactions in the substituted benzene dimers correlate with $\sum \sigma_m$ (as might be predicted by the Hunter-Sanders rules). The electron withdrawing substituents have attractive electrostatic interactions that become stronger as the intermonomer separation is reduced; the electron donating substituents have repulsive electrostatic interactions that become more repulsive at shorter intermonomer separations. At short range, 3 to 4 Å, the correlation falls apart. The electron withdrawing substituents continue to behave as

expected, but the electron donating substituents have increasingly more attractive electrostatic interactions as the intermonomer separation is reduced. At long range, there is very little charge penetration present; multipole-multipole interactions dominate the electrostatic interaction. As the separation is reduced and orbital overlap increases, charge penetration begins to dominate the electrostatic interaction. Notably, the equilibrium geometries here are all in the region (3.45–3.95 Å, see Table 27) where charge penetration effects are very important.

Table 27: The electrostatic, exchange, induction and dispersion components of substituted benzene dimer interactions at equilibrium. Results from SAPT0/aug-cc-pVDZ′ computations in kcal mol^{-1}.

	$\sum \sigma_m$	R (Å)	Elst	Exch	Ind	Disp	SAPT0
$C_6H_6 \cdots C_6H_6$	0.00	3.95	0.52	3.15	-0.24	-4.77	-1.34
$C_6H_5F \cdots C_6H_6$	0.34	3.85	-0.35	3.96	-0.24	-5.39	-2.03
$C_6H_4F_2 \cdots C_6H_6$	0.68	3.75	-1.39	5.00	-0.26	-6.10	-2.75
$C_6H_3F_3 \cdots C_6H_6$	1.02	3.70	-2.29	5.53	-0.31	-6.47	-3.53
$C_6F_6 \cdots C_6H_6$	2.04	3.55	-5.40	7.67	-0.73	-7.81	-6.27
$C_6H_5CN \cdots C_6H_6$	0.56	3.80	-1.23	4.63	-0.26	-6.26	-3.11
$C_6H_4(CN)_2 \cdots C_6H_6$	1.12	3.70	-2.94	6.02	-0.38	-7.60	-4.90
$C_6H_3(CN)_3 \cdots C_6H_6$	1.68	3.60	-4.91	7.85	-0.72	-9.19	-6.97
$C_6(CN)_6 \cdots C_6H_6$	3.36	3.45	-9.36	12.11	-2.02	-13.29	-12.56
$C_6H_5CH_3 \cdots C_6H_6$	-0.07	3.85	0.20	4.26	-0.34	-6.01	-1.89
$C_6H_4(CH_3)_2 \cdots C_6H_6$	-0.14	3.80	-0.01	4.97	-0.40	-6.98	-2.43
$C_6H_3(CH_3)_3 \cdots C_6H_6$	-0.21	3.75	-0.26	5.80	-0.52	-8.06	-3.03
$C_6(CH_3)_6 \cdots C_6H_6$	-0.42	3.60	-1.29	8.61	-0.69	-11.48	-4.85
$C_6H_5NH_2 \cdots C_6H_6$	-0.16	3.80	0.11	4.60	-0.39	-6.18	-1.86
$C_6H_4(NH_2)_2 \cdots C_6H_6$	-0.32	3.70	-0.42	5.81	-0.51	-7.40	-2.52
$C_6H_3(NH_2)_3 \cdots C_6H_6$	-0.48	3.60	-1.05	7.53	-0.74	-8.88	-3.13
$C_6(NH_2)_6 \cdots C_6H_6$	-0.96	3.50	-2.29	9.81	-0.80	-11.78	-5.07
$C_6H_5OH \cdots C_6H_6$	0.12	3.85	0.05	4.02	-0.29	-5.59	-1.81
$C_6H_4(OH)_2 \cdots C_6H_6$	0.24	3.75	-0.58	5.13	-0.33	-6.53	-2.30
$C_6H_3(OH)_3 \cdots C_6H_6$	0.36	3.70	-1.01	5.74	-0.38	-7.10	-2.75

Although Figure 45 emphasizes the two most important attractive forces (electrostatics and dispersion), it is worth commenting on the other SAPT components, induction and exchange-repulsion. In previous work,[92,185,202] we have found that although substituents or heteroatoms create a dipole not present in unsubstituted benzene, the resulting dipole-induced dipole forces (included in the induction term) are relatively weak. Figures 47-53 show that variations in the induction term due to substituents are quite modest (typically a few tenths of one kcal mol^{-1}) and are generally much smaller than variations in the other energy components.

For π-π interactions, the dispersion and exchange-repulsion terms are often of roughly

Figure 47: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at their respective equilibrium geometries. The signs of the relative exchange energies are reversed for ease of comparison.

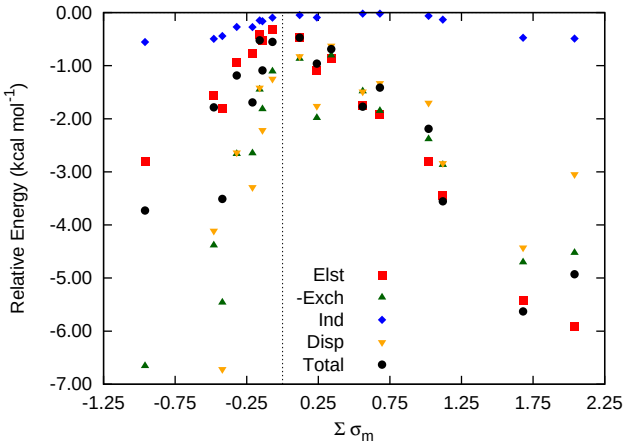
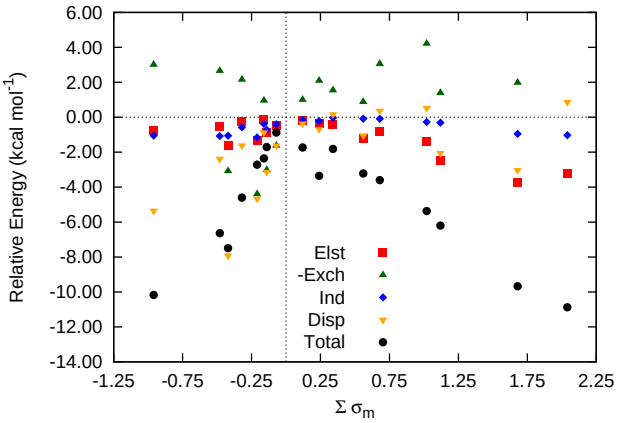


Figure 48: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 3.0 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.



equal magnitude (but opposite sign), leading them to approximately cancel.[92,202] This is not a general phenomenon, as exchange-repulsion will tend to cancel whatever the dominant attraction is in the complex; for hydrogen-bonded systems, exchange is closer in magnitude to the dominant electrostatic term than it is to the dispersion term.[221] Figure 47 shows

Figure 49: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 3.5 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.

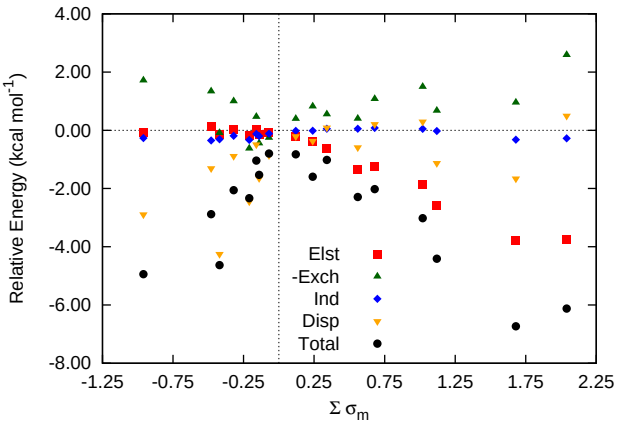
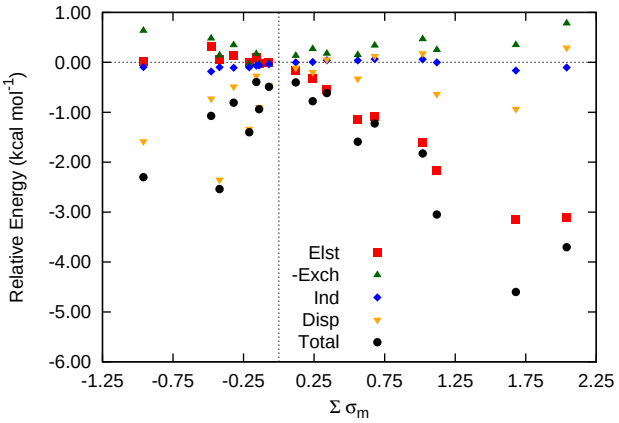


Figure 50: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 4.0 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.



that exchange roughly cancels dispersion at equilibrium, although there are significant differences (up to 2.9 kcal mol^{-1}) in some cases. However, much of this appears to be a geometry effect arising because the substituted benzene dimers are bound more strongly and achieve shorter intermolecular separations. At fixed separations of from 3.0–4.0 Å, the

Figure 51: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 5.0 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.

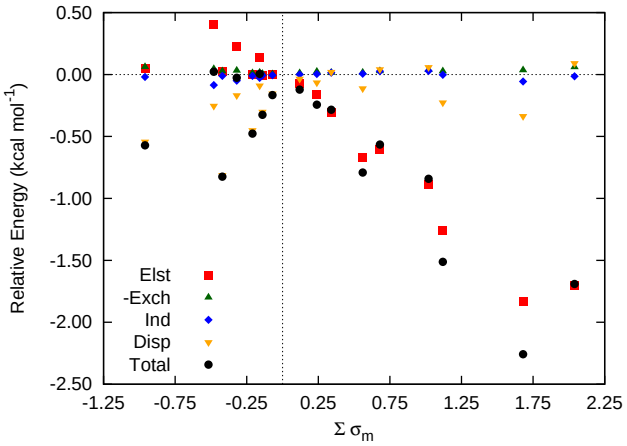
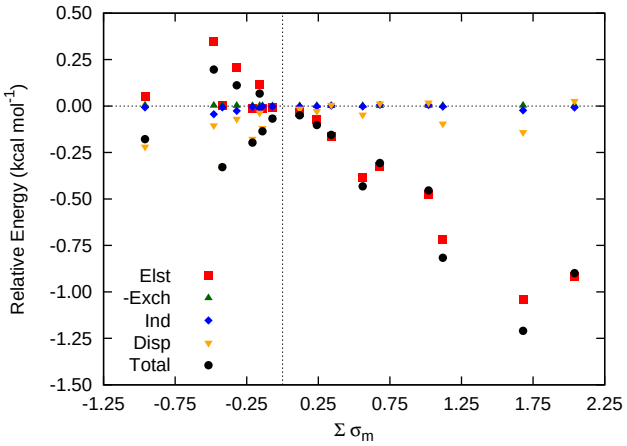
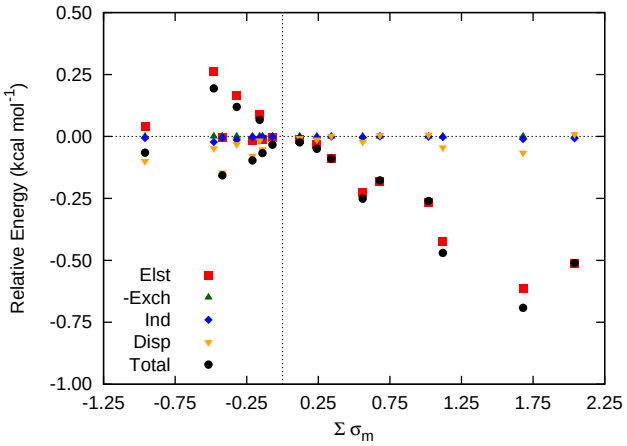


Figure 52: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 6.0 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.



substituted dimers usually feature *less* unfavorable exchange-repulsion than in the benzene dimer. This is surprising given that, on the Wheeler-Houk view,[240–242] the main effect of the substituent should be direct exchange-repulsion between the substituent and the unsubstituted benzene, which one would imagine as always being more repulsive than in the benzene

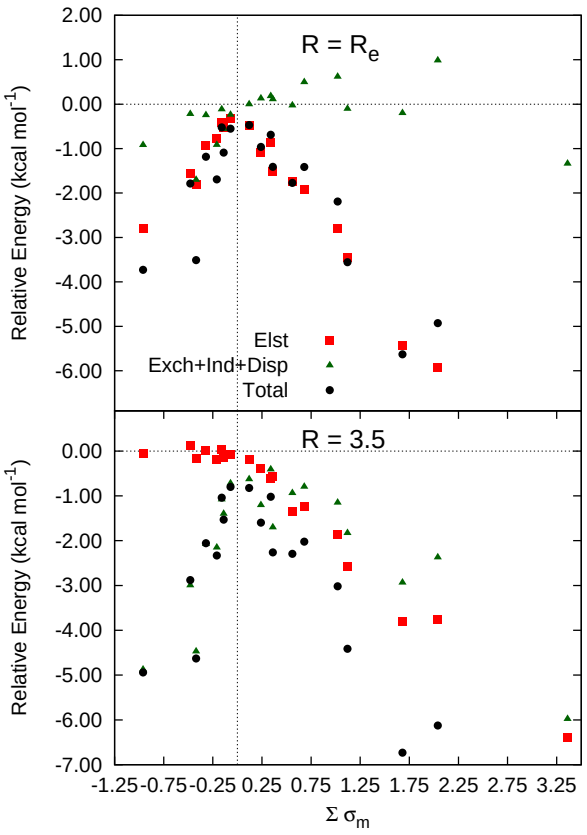
Figure 53: Electrostatic, exchange, induction, dispersion, and and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at constant 7.0 Å displacements. The signs of the relative exchange energies are reversed for ease of comparison.



dimer. We hypothesize that because exchange-repulsion is such a short-range phenomenon and so sensitive to the orbital overlap, it remains dominated by small changes in the π-electron density induced by substituents. All electron-withdrawing substituents considered here lead to reduced exchange repulsion, whereas the electron-donating methyl groups lead to enhanced exchange repulsion. (Curiously, the electron-donating amine group breaks this pattern and leads to reduced exchange repulsion). Although the behavior of the relative exchange-repulsion term is hard to reconcile with the Wheeler-Houk view, the magnitude of this term is typically comparable to or smaller than the relative dispersion and electrostatic terms (see Figures 48-53), so that substituent effects on the total interaction energy may remain well-described by the Wheeler-Houk picture.

Finally, we wish to further explore the other intriguing finding of Lewis and co-workers,[235] that the sum of all non-electrostatic terms is roughly constant (thus commending the electrostatic term as the primary descriptor for substituent effects). Similar results have been noted[134,194] for parallel-displaced configurations of substituted benzene dimers at their equilibrium positions. For the present sandwich systems at their equilibrium separations, the upper panel of Figure 54 plots the electrostatic, non-electrostatic (exchange + induction

Figure 54: Electrostatic, non-electrostatic, and total interaction energies of substituted benzene dimers relative to the unsubstituted benzene dimer at (upper panel) equilibrium geometries and (lower panel) at an intermolecular separation of 3.5 Å.



+ dispersion), and total SAPT energies vs. $\sum \sigma_m$. Consistent with the findings of Lewis and co-workers,[235] the electrostatic energies generally track the total interaction energies (although there are differences as large as 3.6 kcal mol^{-1}), and the non-electrostatic terms are roughly constant and near zero (although they can be as large as 1.7 kcal mol^{-1}). We find a good correlation ($R^2 = 0.94$, see Table 28) between the SAPT electrostatic energies and the total SAPT energies.

Unfortunately, however, the quality of this correlation degrades significantly for non-equilibrium geometries. π-π and other non-covalent interactions often occur in the context of larger systems, where backbone constraints or competing interactions prevent individual contacts from reaching what would otherwise be their optimal geometries. Hence, for a

Table 28: Correlation (R^2) between interaction energies (relative to benzene dimer) and various energy components, $\sum \sigma_m$, and $\sum |\sigma_m|$, at various intermolecular distances.

	Elst (SAPT)	Elst (DMA)[a]	Exch+Ind+Disp	Disp	$\sum \sigma_m$	$\sum \|\sigma_m\|$
3.0 Å	0.73	0.59	0.91	0.23	0.42	0.85
3.5 Å	0.76	0.68	0.69	0.28	0.56	0.90
4.0 Å	0.81	0.79	0.48	0.25	0.67	0.92
5.0 Å	0.91	0.90	0.28	0.18	0.79	0.91
6.0 Å	0.96	0.95	0.21	0.15	0.84	0.88
7.0 Å	0.98	0.97	0.19	0.14	0.85	0.85
R_{eq}[b]	0.94	0.71	0.14	0.64	0.60	0.91

[a]Distributed multipole analysis is performed with HF/6-31G* densities. Electrostatic interactions through quadrupole-quadrupole are included. [b]Dimers are at their respective equilibrium geometries.

correlation to remain useful, it must hold at a range of geometries. As shown in the lower panel of Figure 54, for fixed separations, the relative non-electrostatic energies are no longer nearly constant, varying as much as 6.0 kcal mol^{-1} at 3.5 Å or as much as 3.2 kcal mol^{-1} at 4.0 Å for $C_6(CN)_6$. As shown in Table 28, the R^2 metric of the correlation between the relative electrostatic and total energies reduces to 0.81 at 4.0 Å, and to only 0.76 at 3.5 Å (where the correlation with the non-electrostatic components, 0.69, starts to become as good). The correlation between relative interaction energies and electrostatic energies improves at larger distances where other contributions start to die off. Interestingly, the good correlation noted by Lewis and co-workers[235] between relative interaction energies and the sum of the *absolute values* of the Hammett parameters, $\sum |\sigma_m|$, persists here (although somewhat diminished in quality), with R^2 ranging from 0.85 to 0.92, depending on the intermolecular distance. However, as there is no precedent or theoretical justification for using $\sum |\sigma_m|$ as a descriptor of interaction energies, without additional testing we are unsure whether it will prove useful in future studies.

The present study has shown that charge penetration effects are crucial for understanding the electrostatic component of π-π interactions and their substituent effects. There is no reason to expect that Hammett σ_m parameters should be capable of describing complicated effects such as charge penetration, and indeed at shorter ranges, correlation between the electrostatic energy and $\sum \sigma_m$ breaks down. As shown previously,[184] for multiply-substitued sandwich benzene dimers, there is not a good correlation between relative binding energies and $\sum \sigma_m$ (see also Table 28). In agreement with Lewis and co-workers,[235] we do find a

fairly good correlation between relative binding energies and $\sum|\sigma_m|$, but we are reluctant to advocate use of these parameters without further study.

Although additional questions remain, we are optimistic that all of the main fundamental concepts necessary for an understanding of substituent effects in the simplest π-π model system (gas-phase sandwich benzene dimer) are now recognized. These appear to be: (a) Except at large intermolecular separations ($\sim$ 6 Å or more) all substituents lead to increased binding, regardless of electron-donating or electron-withdrawing character.[201] (b) Substituent effects are largely due to direct substituent-π interactions, not an indirect modulation of π density[240] (although as indicated here, this may not be true for the particular contribution from exchange). (c) Substituents influence not only electrostatic contributions, but also exchange, induction, and London dispersion contributions.[202] Although substituent effects in sandwich and parallel-displaced configurations correlate well with just the (SAPT) electrostatic contribution at equilibrium,[134,194,235] at other geometries this correlation is not nearly as good. (d) The electrostatic contributions are modified by a substituent not only through changes in dipoles, quadrupoles, etc., but also through charge penetration effects that are not easily modeled by multipoles (and certainly not by atom-centered charges[197]). To demonstrate this point, a simplified electrostatic model complete through quadrupole-quadrupole interactions was shown to fail at reproducing the more rigorous SAPT electrostatic energy. We stress that parallel π-π interactions are often observed for interplanar distances of 3.6 Å or less (3.3 Å in B-DNA), and that charge penetration effects are large (i.e., multipole models fail) at these distances. It now remains to explore how these fundamental concepts in the physics of π-π interactions play out in other geometrical arrangements, for larger systems, and in the presence of solvent.

CHAPTER IV

CONCLUSION

The application of density fitting (DF) and Cholesky decomposition (CD) approximations to wavefunction-based SAPT has proven to greatly improve its scalability. Both DF and CD approaches have negligibly small errors compared to SAPT results using standard electron repulsion integrals. For our test cases, the auxiliary basis required for the one-center CD approach is smaller than the corresponding DF basis. This could prove very useful for computations on large systems, where the expense associated with the formation of the three-index integrals is negligible compared to the formation of four-index MO integrals from three-index integrals. Recent algorithmic advances have allowed us to perform SAPT0 computations on systems with as many as 220 atoms and 2850 basis functions.

Likewise, the application of DF approximations to the treatment of intramonomer electron correlation in higher-order SAPT is also fruitful. Through the formation of a group of reusable intermediates based on the DF representation of the two-electron integrals, we are able to greatly improve the efficiency of our DF-SAPT program. In some cases, the DF intermediates are used to reduce the overall scaling of certain SAPT corrections. These intermediates can also be used to avoid the need to handle certain classes of integrals explicitly. As a result the dependence on disk I/O, which has plagued other SAPT implementations, is greatly reduced in our program. In order to get highly accurate interaction energies from SAPT, the triples contribution to dispersion must be included; this terms scales as $\mathcal{O}(o^3v^4)$. By using DF intermediates, the evaluation of this term can be streamlined to avoid a bottleneck due to excessive disk I/O. Our implementation of DF-SAPT has allowed us to perform the largest SAPT computations to date that include all intramonomer correlation corrections through second-order.

Despite the improvements from introduction of density fitting approximations, the

triples correction to dispersion remained a bottleneck in SAPT computations. The development of an approximation to the triples correction to dispersion in SAPT that uses MP2 natural orbitals (NOs) to reduce the number of virtual orbitals and a scaling relation to reduce the size of the error incurred has significantly reduced this bottleneck. By truncating the virtual space and scaling the resulting energy according to Equation 204, roughly half of the virtual orbitals can be removed with negligible errors. When this approximation is used in conjunction with the frozen core approximation, $E_{disp}^{(22)}$(T) can be evaluated 50-60 times faster for the aug-cc-pVDZ basis set, with even greater speedups for larger basis sets. For the S22 test set, these approximations result in errors of, at most, only a few hundredths of one kcal mol^{-1}. With the approximations developed in this work, it should now be possible to include the important $E_{disp}^{(22)}$(T) term for most systems where SAPT2 is applicable. The introduction of density fitting and MP2 NO approximations have allowed us to perform correlated SAPT computations on systems with as many as 44 atoms and 840 basis functions (a double-ζ quality basis) or 30 atoms and 1130 basis functions (a triple-ζ quality basis).

The advances we have made in the development of an efficient wavefunction-based SAPT implementation have allowed us to perform the largest such computations to date. The energy decompositions provided by our SAPT program have been used to elucidate the nature of interactions in many complex systems. The first chemical application of our SAPT program was to the indole-benzene complex, where qualitative explanations for the energetic ordering of numerous geometries were desired.[60] The fundamental nature of π-π interactions in extended systems was probed with SAPT0 to determine how different energy components change with the length of linear acenes.[93] After the higher-order SAPT terms were efficiently implented, correlated SAPT computations were performed on small hydrogen bonded complexes in order to study the distance dependance of the interactions.[221] We have performed high-level SAPT computations on the protonated benzene dimer (the benzene-benzenium interaction) in order to illustrate the difference between T-shaped and parallel displaced configurations of the complex; the SAPT analysis allowed the protonated complex to be

compared to the neutral benzene dimer.[104] Our CCD dispersion program, the first production level implementation of the $\epsilon_{disp}^{(2)}(\mathrm{CCD} + \mathrm{ST(CCD)})$ correction, was applied to the pathological dispersion interactions in the NCCN, PCCP, and P_2 dimers.[98] Our new implementation of SAPT0 has allowed large data sets and large molecules to be treated efficiently. This has allowed us to explore the rise and twist of stacked DNA base pair duplexes.[160] We have also applied our SAPT0 program to study the interactions in dimers of arsenolite (As_4O_6) and realgar (As_4S_4).[63] We have probed the nature of electrostatic interactions in substituted benzene dimers to ascertain the importance of charge penetration.[97] Using the scalability of the SAPT0 implementation, we have studied the intercalation of DNA by proflavine, while including backbone and next-nearest neighbor interactions.[99] SAPT is an extremely useful tool for providing chemical insight into the nature of noncovalent interactions. Our new DF-SAPT implementation makes a wider range of systems accessible with wavefunction-based SAPT.

REFERENCES

[1] Abrams, M. L. and Sherrill, C. D., "A comparison of polarized double-zeta basis sets and natural orbitals for full configuration interaction benchmarks," *J. Chem. Phys.*, vol. 118, pp. 1604–1609, 2003.

[2] Abrams, M. L. and Sherrill, C. D., "Natural orbitals as substitutes for optimized orbitals in complete active space wavefunctions," *Chem. Phys. Lett.*, vol. 395, no. 4-6, pp. 227–232, 2004.

[3] Adamowicz, L. and Bartlett, R. J., "Optimized virtual orbital subspace for high-level correlated calculations," *J. Chem. Phys.*, vol. 86, pp. 6314–6324, 1987.

[4] Allen, M. J. and Tozer, D. J., "Helium dimer dispersion forces and correlation potentials in density functional theory," *J. Chem. Phys.*, vol. 117, no. 24, pp. 11113–11120, 2002.

[5] Almlöf, J., "Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform appraoch," *Chem. Phys. Lett.*, vol. 96, p. 319, 1991.

[6] Antony, J. and Grimme, S., "Is spin-component scaled second-order Møller-Plesset perturbation theory an appropriate method for the study of noncovalent interactions in molecules?," *J. Phys. Chem. A*, vol. 111, pp. 4862–4868, 2007.

[7] Aquilante, F., Lindh, R., and Pedersen, T. B., "Unbiased auxiliary basis sets for accurate two-electron integral approximations," *J. Chem. Phys.*, vol. 127, p. 114107, 2007.

[8] Aquilante, F., Lindh, R., and Pedersen, T. B., "Analytic derivatives for the Cholesky representation of the two-electron integrals," *J. Chem. Phys.*, vol. 129, p. 034106, 2008.

[9] Aquilante, F., Pedersen, T. B., and Lindh, R., "Low-cost evaluation of the exchange fock matrix from Cholesky and density fitting representations of the electron repulsion integrals," *J. Chem. Phys.*, vol. 126, p. 194106, 2007.

[10] Arnstein, S. A. and Sherrill, C. D., "Substituent effects in parallel-displaced π-π interactions," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 2646–2655, 2008.

[11] Axilrod, B. M. and Teller, E., "Interaction of the van der Waals type between three atoms," *J. Chem. Phys.*, vol. 11, pp. 299–300, 1943.

[12] Bagus, P. S., Hermann, K., and Bauschlicher, C. W., "A new analysis of charge transfer and polarization for ligand-metal bonding: Model studies of Al_4CO and Al_4NH_3," *J. Chem. Phys.*, vol. 80, pp. 4378–4386, 1984.

[13] Bates, D. M., Anderson, J. A., Oloyede, P., and Tschumper, G. S., "Probing the effects of heterogeneity on delocalized π-π interaction energies," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 2775–2779, 2008.

[14] Becke, A. D. and Johnson, E. R., "A density-functional model of the dispersion interaction," *J. Chem. Phys.*, vol. 123, p. 154101, 2005.

[15] Becke, A. D. and Johnson, E. R., "Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients," *J. Chem. Phys.*, vol. 124, p. 014104, 2006.

[16] Beebe, N. H. F. and Linderberg, J., "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," *Int. J. Quantum Chem.*, vol. 12, pp. 683–705, 1977.

[17] Bofill, J. M. and Pulay, P., "The unrestricted natural orbital-complete active space (UNO-CAS) method: An inexpensive alternative to the complete active space-self-consistent-field (CAS-SCF) method," *J. Chem. Phys.*, vol. 90, pp. 3637–3646, 1989.

[18] Bondarev, D. A., Skawinski, W. J., and Venanzi, C. A., "Nature of intercalator amiloride-nucleobase stacking. An empirical potential and ab initio electron correlation study," *J. Phys. Chem. B*, vol. 104, pp. 815–822, 2000.

[19] Boström, J., Aquilante, F., Pedersen, T. B., and Lindh, R., "Ab initio density fitting: Accuracy assessment of auxiliary basis sets from Cholesky decompositions," *J. Chem. Theory Comput.*, vol. 5, pp. 1545–1553, 2009.

[20] Boström, J., Delcey, M. G., Aquilante, F., Serrano-Andrés, L., Pedersen, T. B., and Lindh, R., "Calibration of Cholesky auxiliary basis sets for multiconfigurational perturbation theory calculations of excitation energies," *J. Chem. Theory Comput.*, vol. 6, pp. 747–754, 2010.

[21] Boys, S. F. and Bernardi, F., "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.*, vol. 19, no. 4, pp. 553–566, 1970.

[22] Braess, D. and Hackbusch, W., "Approximation of 1/x by exponential sums in [1,∞)," *IMA J. Numer. Anal.*, vol. 25, p. 685, 2005.

[23] Braun, J., Neusser, H. J., and Hobza, P., "N-H$\cdots\pi$ interactions in Indole$\cdots$Benzene-H_6,D_6 and Indole$\cdots$Benzene-H_6,D_6 radical cation complexes. Mass analyzed threshold ionization experiments and correlated ab initio quantum chemical calculations," *J. Phys. Chem. A*, vol. 107, pp. 3918–3924, 2003.

[24] Bronstrup, M., Gottfriedsen, J., Kretzschmar, I., Blanksby, S. J., Schwarz, H., and Schumann, H., "PCCP does exist," *Phys. Chem. Chem. Phys.*, vol. 2, no. 10, pp. 2245–2250, 2000.

[25] Bukowski, R., Cencek, W., Jankowski, P., Jeziorska, M., Jeziorski, B., Lotrich, V. F., Kucharski, S. A., Misquitta, A. J., Moszyński, R., Patkowski, K., Podeszwa, R., Rybak, S., Szalewicz, K., Williams, H. L., Wheatley, R. J., Wormer, P. E. S., and Żuchowski, P. S., "*SAPT2008*: An *ab initio* program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies." University of Delaware and University of Warsaw (http://www.physics.udel.edu/∼szalewic/SAPT/SAPT.html).

[26] Bukowski, R., Cencek, W., Patkowski, K., Jankowski, P., Jeziorska, M., Kolaski, M., and Szalewicz, K., "Portable parallel implementation of symmetry-adapted perturbation theory code," *Mol. Phys.*, vol. 104, pp. 2241–2262, 2006.

[27] Bukowski, R., Podeszwa, R., and Szalewicz, K., "Efficient calculation of coupled Kohn-Sham dynamic susceptibility functions and dispersion energies with density fitting," *Chem. Phys. Lett.*, vol. 414, pp. 111–116, 2005.

[28] Burley, S. K. and Petsko, G. A., "Aromatic-aromatic interaction: A mechanism of protein structure stabilization," *Science*, vol. 229, pp. 23–28, 1985.

[29] Burns, L. A., Vázquez-Mayagoitia, Á., Sumpter, B. G., and Sherrill, C. D., "Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals," *J. Chem. Phys.*, vol. 134, p. 084107, 2011.

[30] Bytautas, L., Ivanic, J., and Ruedenberg, K., "Split-localized orbitals can yield stronger configuration interaction convergence than natural orbitals," *J. Chem. Phys.*, vol. 119, no. 16, pp. 8217–8224, 2003.

[31] Bytautas, L. and Ruedenberg, K., "Correlation energy extrapolation by intrinsic scaling. I. Method and application to the neon atom," *J. Chem. Phys.*, vol. 121, no. 22, pp. 10905–10918, 2004.

[32] Bytautas, L. and Ruedenberg, K., "Correlation energy extrapolation by intrinsic scaling. II. The water and nitrogen molecule," *J. Chem. Phys.*, vol. 121, no. 22, pp. 10919–10934, 2004.

[33] Bytautas, L. and Ruedenberg, K., "Correlation energy extrapolation by intrinsic scaling. III. Compact wave functions," *J. Chem. Phys.*, vol. 121, pp. 10852–10862, 2004.

[34] Černý, J. and Hobza, P., "The X3LYP extended density functional accurately describes H-bonding but fails completely for stacking," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 1624–1626, 2005.

[35] Chai, J. and Head-Gordon, M., "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections," *Phys. Chem. Chem. Phys.*, vol. 10, p. 6615, 2008.

[36] Chen, W. and Gordon, M. S., "Energy decomposition analyses for many-body interaction and applications to water complexes," *J. Phys. Chem.*, vol. 100, pp. 14316–14328, 1996.

[37] Chong, Y. S., Carroll, W. R., Burns, W. G., Smith, M. D., and Shimizu, K. D., "A high-barrier molecular balance for studying face-to-face arene-arene interactions in the solid state and in solution," *Chem. Eur. J.*, vol. 15, pp. 9117–9126, 2009.

[38] Collings, J. C., Roscoe, K. P., Robins, E. G., Batsanov, A. S., Stimson, L. M., Howard, J. A. K., Clark, S. J., and Marder, T. B., "Arene-perfluoroarene interactions in crystal engineering 8: structures of 1:1 complexes of hexafluorobenzene with fused-ring polyaromatic hydrocarbons," *New J. Chem.*, vol. 26, p. 1740, 2002.

[39] Cozzi, F., Cinquini, M., Annuziata, R., Dwyer, T., and Siegel, J. S., "Polar/π interactions between stacked aryls in 1,8-diarylnaphthalenes," *J. Am. Chem. Soc.*, vol. 114, no. 14, pp. 5729–5733, 1992.

[40] Cramer, C. J. and Truhlar, D. G., "Implicit solvation models: Equilibria, structure, spectra, and dynamics," *Chem. Rev.*, vol. 99, pp. 2161–2200, 1999.

[41] Crawford, T. D., Sherrill, C. D., Valeev, E. F., Fermann, J. T., King, R. A., Leininger, M. L., Brown, S. T., Janssen, C. L., Seidl, E. T., Kenny, J. P., and Allen, W. D., "PSI3: An open-source *ab initio* electronic structure package," *J. Comput. Chem.*, vol. 28, pp. 1610–1616, 2007.

[42] Davidson, E. R., "Properties and uses of natural orbitals," *Rev. Mod. Phys.*, vol. 44, pp. 451–464, 1972.

[43] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C., and Lundqvist, B. I., "van der Waals density functional for general geometries," *Phys. Rev. Lett.*, vol. 92, no. 24, p. 246401, 2004.

[44] Distasio, R. A. and Head-Gordon, M., "Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies," *Mol. Phys.*, vol. 105, pp. 1073–1083, 2007.

[45] Doerksen, R. J. and Thakkar, A. J., "Polarizabilities of heteroaromatic molecules: Azines revisited," *Int. J. Quantum Chem. Symp.*, vol. 30, p. 1633, 1996.

[46] Dongwook, K., Lee, E. C., Kim, K. S., and Tarakeshwar, P., "Cation-π-anion interaction: A theoretical investigation of the role of induction," *J. Phys. Chem. A*, vol. 111, pp. 7980–7986, 2007.

[47] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "Applicability of LCAO-X-alpha methods to molecules containing transition-metal atoms - nickel atom and nickel hydride," *Int. J. Quantum Chem. Symp.*, vol. 11, p. 81, 1977.

[48] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "On some approximations in applications of $X\alpha$ theory," *J. Chem. Phys.*, vol. 71, pp. 3396–3402, 1979.

[49] Dunning, T. H., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, pp. 1007–1023, 1989.

[50] Edmiston, C. and Krauss, M., "Configuration-interaction calculation of H_3 and H_2," *J. Chem. Phys.*, vol. 42, pp. 1119–1120, 1965.

[51] Elsohly, A. M. and Tschumper, G. S., "Comparison of polarization consistent and correlation consistent basis sets for noncovalent interactions," *Int. J. Quantum Chem.*, vol. 109, p. 91, 2009.

[52] Elstner, M., Hobza, P., Frauenheim, T., Suhai, S., and Kaxiras, E., "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment," *J. Chem. Phys.*, vol. 114, no. 12, pp. 5149–5155, 2001.

[53] Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Kennedy, M. R., Sherrill, C. D., and Merz, K. M., "Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes," *J. Chem. Theory Comput.*, vol. 7, pp. 790–797, 2011.

[54] Feller, D., "The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water," *J. Chem. Phys.*, vol. 98, pp. 7059–7071, 1993.

[55] Feyereisen, M., Fitzgerald, G., and Komornicki, A., "Use of approximate integrals in ab initio theory. an application in MP2 calculations," *Chem. Phys. Lett.*, vol. 208, pp. 359–363, 1993.

[56] Flocke, N. and Bartlett, R. J., "A natural linear scaling coupled-cluster method," *J. Chem. Phys.*, vol. 121, p. 10935, 2004.

[57] Friesner, R. A., Murphy, R. B., Beachy, M. D., Ringnalda, M. N., Pollard, W. T., Dunietz, B. D., and Cao, Y., "Correlated ab initio electronic structure calculations for large molecules," *J. Phys. Chem. A*, vol. 103, pp. 1913–1928, 1999.

[58] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J., "Gaussian 09 Revision B.01." Gaussian Inc. Wallingford CT 2009.

[59] Gauss, J. and Stanton, J. F., "The equilibrium structure of benzene," *J. Phys. Chem. A*, vol. 104, pp. 2865–2868, 2000.

[60] Geng, Y., Takatani, T., Hohenstein, E. G., and Sherrill, C. D., "Accurately characterizing the π-π interaction energies of indole-benzene complexes," *J. Phys. Chem. A*, vol. 114, pp. 3576–3582, 2010.

[61] Geronimo, I., Lee, E. C., Singh, N. J., and Kim, K. S., "How different are electron-rich and electron-deficient pi interactions?," *J. Chem. Theory Comput.*, vol. 6, p. 1931, 2010.

[62] Ghosh, D., Kosenkov, D., Vanovschi, V., Williams, C. F., Herbert, J. M., Gordon, M. S., Schmidt, M. W., Slipchenko, L. V., and Krylov, A. I., "Noncovalent interactions in extended systems described by the effective fragment potential method: Theory and application to nucleobase oligomers," *J. Phys. Chem. A*, vol. 114, pp. 12739–12754, 2010.

[63] Gibbs, G. V., Crawford, T. D., Wallace, A. F., Cox, D. F., Parrish, R. M., Hohenstein, E. G., and Sherrill, C .D., "The role of long-range directed Lewis acid-base van der Waals interactions in the formation of nanoparticle clusters," *J. Phys. Chem. A*, submitted.

[64] Goerigk, L. and Grimme, S., "Efficient and accurate double-hybrid-meta-GGA density functionals-evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions," *J. Chem. Theory Comput.*, vol. 7, p. 291, 2011.

[65] Gráfová, L., Pitoňák, M., Řezáč, J., and Hobza, P., "Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set," *J. Chem. Theory Comput.*, vol. 6, pp. 2365–2376, 2010.

[66] Grev, R. S. and Schaefer, H. F., "Natural orbitals from single and double excitation configuration interaction wave functions: Their use in second-order configuration interaction and wave functions incorporating limited triple and quadruple excitations," *J. Chem. Phys.*, vol. 96, p. 6850, 1992.

[67] Grimme, S., "Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies," *J. Chem. Phys.*, vol. 118, no. 20, pp. 9095–9102, 2003.

[68] Grimme, S., "Accurate description of van der Waals complexes by density functional theory including empirical corrections," *J. Comput. Chem.*, vol. 25, pp. 1463–1473, 2004.

[69] Grimme, S., "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *J. Comput. Chem.*, vol. 27, no. 15, pp. 1787–1799, 2006.

[70] Grimme, S., "Semiempirical hybrid density functional with perturbative second-order correlation," *J. Chem. Phys.*, vol. 124, p. 034108, 2006.

[71] Grimme, S., "Do special π-π stacking interactions really exist?," *Angew. Chem. Int. Ed. Engl.*, vol. 47, pp. 3430–3434, 2008.

[72] Grimme, S., "Density functional theory with london dispersion corrections," *WIREs Comput. Mol. Sci.*, vol. 1, pp. 211–228, 2011.

[73] Grimme, S., Antony, J., Ehrlich, S., and Krieg, H., "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[74] Grimme, S., Antony, J., Schwabe, T., and Mück-Lichtenfeld, C., "Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules," *Org. Biomol. Chem.*, vol. 5, pp. 741–758, 2007.

[75] Hajgato, B., Pham-Tran, N.-. N., Veszpremi, T., and Nguyen, M. T., "PCCP and its isomers: a theoretical study," *Phys. Chem. Chem. Phys*, vol. 3, pp. 5158–5164, 2001.

[76] Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., Koch, H., Olsen, J., and Wilson, A. K., "Basis-set convergence in correlated calculations on Ne, N_2, and H_2O," *Chem. Phys. Lett.*, vol. 286, pp. 243–252, 1998.

[77] Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W., and Olsen, J., "Basis-set convergence of the energy in molecular Hartree–Fock calculations," *Chem. Phys. Lett.*, vol. 302, pp. 437–446, 1999.

[78] Halkier, A., Klopper, W., Helgaker, T., Jørgensen, P., and Taylor, P. R., "Basis set convergence of the interaction energy of hydrogen-bonded complexes," *J. Chem. Phys.*, vol. 111, pp. 9157–9167, 1999.

[79] Hampel, C. and Werner, H.-J., "Local treatment of electron correlation in coupled cluster theory," *J. Chem. Phys.*, vol. 104, pp. 6286–6297, 1996.

[80] Häser, M. and Almlöf, J., "Laplace transform techniques in Møller-Plesset perturbation theory," *J. Chem. Phys.*, vol. 96, p. 489, 1992.

[81] Hättig, C., "Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-zeta basis sets for H to Ar and QZVPP basis sets for Li to Kr," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 59–66, 2005.

[82] Herzberg, G., *Molecular Spectra and Molecular Structure: Electronic Spectra and Electronic Structure of Polyatomic Molecules*, vol. 3. Princeton: Van Nostrand, 1950.

[83] Herzberg, G., *Molecular Spectra and Molecular Structure: Spectra of Diatomic Molecules*, vol. 1. Malabar: R.E. Krieger Pub. Co, 1989.

[84] Hesselmann, A., "Improved supermolecular second order Møller-Plesset intermolecular interaction energies using time-dependent density functional response theory," *J. Chem. Phys.*, vol. 128, p. 144112, 2008.

[85] Hesselmann, A. and Jansen, G., "The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchangecorrelation potential," *Phys. Chem. Chem. Phys.*, vol. 5, p. 5010, 2003.

[86] Hesselmann, A., Jansen, G., and Schütz, M., "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," *J. Chem. Phys.*, vol. 122, p. 014103, 2005.

[87] Hill, J. G. and Platts, J. A., "Spin-component scaling methods for weak and stacking interactions," *J. Chem. Theory Comput.*, vol. 3, pp. 80–85, 2007.

[88] Hill, J. G. and Platts, J. A., "Local electron correlation descriptions of the intermolecular stacking interactions between aromatic intercalators and nucleic acids," *Chem. Phys. Lett.*, vol. 479, pp. 279–283, 2009.

[89] Hill, J. G., Platts, J. A., and Werner, H., "Calculation of intermolecular interactions in the benzene dimer using coupled-cluster and local electron correlation methods," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 4072–4078, 2006.

[90] Hobza, P., Šponer, J., and Reschel, T., "Density-functional theory and molecular clusters," *J. Comp. Chem.*, vol. 16, no. 11, pp. 1315–1325, 1995.

[91] Hohenstein, E. G., Chill, S. T., and Sherrill, C. D., "Assessment of the performance of the M05-2x and M06-2x exchange-correlation functionals for noncovalent interactions in biomolecules," *J. Chem. Theory Comput.*, vol. 4, pp. 1996–2000, 2008.

[92] Hohenstein, E. G. and Sherrill, C. D., "Effects of heteroatoms on aromatic π-π interactions: Benzene-pyridine and pyridine dimer," *J. Phys. Chem. A*, vol. 113, pp. 878–886, 2009.

[93] Hohenstein, E. G. and Sherrill, C. D., "Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes," *J. Chem. Phys.*, vol. 132, p. 184111, 2010.

[94] Hohenstein, E. G. and Sherrill, C. D., "Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 133, p. 014101, 2010.

[95] Hohenstein, E. G. and Sherrill, C. D., "Efficient evaluation of triple excitations in symmetry-adapted perturbation theory via second-order Møller-Plesset perturbation theory natural orbitals," *J. Chem. Phys.*, vol. 133, p. 104107, 2010.

[96] Hohenstein, E. G. and Sherrill, C. D., "Wavefunction methods for noncovalent interactions," *WIREs Comput. Mol. Sci.*, in press.

[97] Hohenstein, E. G., Duan, J., and Sherrill, C. D., "Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers," *J. Am. Chem. Soc.*, submitted.

[98] Hohenstein, E. G., Jaeger, H. M., Carrell, E. J., Tschumper, G. S., and Sherrill, C. D., "Accurate interaction energies for problematic dispersion-bound complexes: homogeneous dimers of NCCN, P_2 and PCCP," *J. Chem. Theory Comput.*, submitted.

[99] Hohenstein, E. G., Parrish, R. M., and Sherrill, C. D., "Large-scale symmetry-adapted perturbation theory computations via density fitting and Laplace transformation techniques," manuscript in preparation.

[100] Hohenstein, E. G. and Sherrill, C. D., manuscript in preparation.

[101] Hopkins, B. W. and Tschumper, G. S., "Ab initio studies of $\pi \cdots \pi$ interactions: The effects of quadruple excitations," *J. Phys. Chem. A*, vol. 108, no. 15, pp. 2941–2948, 2004.

[102] Hunter, C. A. and Sanders, J. K. M., "The nature of π-π Interactions," *J. Am. Chem. Soc.*, vol. 112, no. 14, pp. 5525–5534, 1990.

[103] Innes, K. K., Ross, I. G., and Moomaw, W. R., "Electronic states of azabenzenes and azanaphthalenes - a revised and extended critical-review," *J. Mol. Spectrosc.*, vol. 132, p. 492, 1988.

[104] Jaeger, H. M., Hohenstein, E. G., Sherrill, C. D., and Schaefer, H. F., "Structures of protonated benzene dimer and intermolecular interaction analysis via symmetry-adapted perturbation theory," *Comput. Theor. Chem.*, accepted.

[105] Jaffe, R. L. and Smith, G. D., "A quantum chemistry study of benzene dimer," *J. Chem. Phys.*, vol. 105, pp. 2780–2788, 1996.

[106] "Jaguar 5.5," Schrodinger, LLC, Portland, Oregon, 2003.

[107] Janowski, T. and Pulay, P., "High accuracy benchmark calculations on the benzene dimer potential energy surface," *Chem. Phys. Lett.*, vol. 447, pp. 27–32, 2007.

[108] Jensen, H. J. Aa., Jørgensen, P., Ågren, H., and Olsen, J., "Second-order Møller-Plesset perturbation theory as a configuration and orbital generator in multiconfiguration self-consistent field calculations," *J. Chem. Phys.*, vol. 88, pp. 3834–3839, 1988.

[109] Jeziorski, B., Bulski, M., and Piela, L., "1st-order perturbation treatment of short-range repulsion in a system of many closed shell atoms or moleucles," *Int. J. Quant. Chem.*, vol. 10, p. 281, 1976.

[110] Jeziorski, B., Moszynski, R., Ratkiewicz, A., Rybak, S., Szalewicz, K., and Williams, H. L., "SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies," in *Methods and Techniques in Computational Chemistry: METECC94* (Clementi, E., ed.), vol. B (Medium-Size Systems), p. 79, STEF: Cagliari, 1993.

[111] Jeziorski, B., Moszynski, R., and Szalewicz, K., "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem. Rev.*, vol. 94, pp. 1887–1930, 1994.

[112] Johnson, E. R. and Becke, A. D., "A post-Hartree-Fock model of intermolecular interactions," *J. Chem. Phys.*, vol. 123, p. 024101, 2005.

[113] Johnson, E. R., Becke, A. D., Sherrill, C. D., and DiLabio, G. A., "Oscillations in meta-generalized-gradient approximation potential energy surfaces for dispersion-bound complexes," *J. Chem. Phys.*, vol. 131, p. 034111, 2009.

[114] Johnson, E. R., Wolkow, R. A., and DiLabio, G. A., "Application of 25 density functionals to dispersion-bound homonuclear dimers," *Chem. Phys. Lett.*, vol. 394, pp. 334–338, 2004.

[115] Jurečka, P., Černý, J., Hobza, P., and Salahub, D. R., "Density functional theory augmented with an empirical dispersion term. interaction energies and geometries of 80 noncovalent complexes compared with *ab initio* quantum mechanics calculations," *J. Comput. Chem.*, vol. 28, pp. 555–569, 2007.

[116] Jurečka, P. and Hobza, P., "True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine-cytosine, adenine-thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment," *J. Am. Chem. Soc.*, vol. 125, pp. 15608–15613, 2003.

[117] Jurečka, P., Šponer, J., Černý, J., and Hobza, P., "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 1985–1993, 2006.

[118] Kendall, R. A., Dunning, T. H., and Harrison, R. J., "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, pp. 6796–6806, 1992.

[119] Kim, D., Hu, S., Tarakeshwar, P., Kim, K. S., and Lisy, J. M., "Cation-π interactions: A theoretical investigation of the interaction of metallic and organic cations with alkenes, arenes, and heteroarenes," *J. Phys. Chem. A*, vol. 107, p. 1228, 2003.

[120] Kim, K. S., Tarakeshwar, P., and Lee, J. Y., "Molecular clusters of π-Systems: Theoretical studies of structures, spectra, and origin of interaction energies," *Chem. Rev.*, vol. 100, no. 11, pp. 4145–4185, 2000.

[121] King, R. A., "On the accuracy of spin-component-scaled perturbation theory (SCS-MP2) for the potential energy surface of the ethylene dimer," *Mol. Phys.*, vol. 107, pp. 789–795, 2009.

[122] Kitaura, K. and Morokuma, K., "New energy decomposition scheme for molecular-interactions within Hartree-Fock approximation," *Int. J. Quantum Chem.*, vol. 10, pp. 325–340, 1976.

[123] Klopper, W., Lüthi, H. P., Brupbacher, T., and Bauder, A., "Ab initio computations close to the one-particle basis set limit on the weakly bound van der Waals complexes benzene-neon and benzene-argon," *J. Chem. Phys.*, vol. 101, no. 11, pp. 9747–9754, 1994.

[124] Klopper, W., Noga, J., Koch, H., and Helgaker, T., "Multiple basis sets in calculations of triples corrections in coupled-cluster theory," *Theor. Chem. Acc.*, vol. 97, pp. 164–176, 1997.

[125] Koch, H., de Meras, A. S., and Pedersen, T. B., "Reduced scaling in electronic structure calculations using Cholesky decompositions," *J. Chem. Phys.*, vol. 118, pp. 9481–9484, 2003.

[126] Kolář, M., Kubař, T., and Hobza, P., "Sequence-dependent configurational entropy change of DNA upon intercalation," *J. Phys. Chem. B*, vol. 114, pp. 13446–13454, 2010.

[127] Korona, T., Hesselmann, A., and Dodziuk, H., "Symmetry-adapted perturbation theory applied to endohedral fullerene complexes: A stability study of $H_2@C_{60}$ and $2H_2@C_{60}$," *J. Chem. Theory Comput.*, vol. 5, pp. 1585–1596, 2009.

[128] Kristyán, S. and Pulay, P., "Can (semi)local density functional theory account for the London dispersion forces?," *Chem. Phys. Lett.*, vol. 229, pp. 175–180, 1994.

[129] Kumar, A., Elstner, M., and Suhai, S., "SCC-DFTB-D study of intercalating carcinogens: Benzo(a)pyrene and its metabolites complexed with the G-C base pair," *Int. J. Quantum Chem.*, vol. 95, pp. 44–59, 2003.

[130] Kurita, N. and Sekino, H., "Ab initio and dft studies for accurate description of van der Waals interaction between rare-gas atoms," *Int. J. Quant. Chem.*, vol. 91, no. 3, pp. 355–362, 2003.

[131] Landau, A., Khistyaev, K., Dolgikh, S., and Krylov, A. I., "Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism," *J. Chem. Phys.*, vol. 132, p. 014109, 2010.

[132] Langner, K. M., Kedzierski, P., Sokalski, W. A., and Leszczynski, J., "Physical nature of ethidium and proflavine interactions with nucleic acid bases in the intercalation plane," *J. Phys. Chem. B*, vol. 110, pp. 9720–9727, 2006.

[133] Lee, E. C., Hong, B. H., Lee, J. Y., Kim, J. C., Kim, D., Kim, Y., Tarakeshwar, P., and Kim, K. S., "Substituent effects on edge-to-face aromatic interactions," *J. Am. Chem. Soc.*, vol. 127, pp. 4530–4537, 2005.

[134] Lee, E. C., Kim, D., Jurečka, P., Tarakeshwar, P., Hobza, P., and Kim, K. S., "Understanding of assembly phenomena by aromatic-aromatic interactions: Benzene dimer and the substituted systems," *J. Phys. Chem. A*, vol. 111, pp. 3446–3457, 2007.

[135] Li, S., Cooper, V. R., Thonhauser, T., Lundqvist, B. I., and Langreth, D. C., "Stacking interactions and DNA intercalation," *J. Phys. Chem. B*, vol. 113, pp. 11166–11172, 2009.

[136] Liedl, K. R., "Dangers of counterpoise corrected hypersurfaces. Advantages of basis set superposition improvement," *J. Chem. Phys.*, vol. 108, pp. 3199–3204, 1998.

[137] Löwdin, P. O., "Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction," *Phys. Rev.*, vol. 97, p. 1474, 1955.

[138] Löwdin, P.-O. and Shull, H., "Natural orbitals in the quantum theory of two-electron systems," *Phys. Rev.*, vol. 101, pp. 1730–1739, 1956.

[139] Martinez, T. J. and Carter, E. A., "Pseudospectral methods applied to the electron correlation problem," in *Modern Electronic Structure Theory* (Yarkony, D. R., ed.), vol. 2 of *Advanced Series in Physical Chemistry*, pp. 1132–1165, Singapore: World Scientific, 1995.

[140] Megiel, E., Kasprzycka-Guttman, T., Jagielska, A., and Wroblewska, L., "A theoretical and experimental N-14 NMR study of association of pyridine," *J. Mol. Struct.*, vol. 569, pp. 111–119, 2001.

[141] Mei, X. and Wolf, C., "Highly congested nondistorted diheteroarylnaphthalenes: Model compounds for the investigation of intramolecular pi-stacking interactions," *J. Org. Chem.*, vol. 70, p. 2299, 2005.

[142] Meyer, E. A., Castellano, R. K., and Diederich, F., "Interactions with aromatic rings in chemical and biological recognition," *Angew. Chem., Int. Ed. Engl.*, vol. 42, no. 11, pp. 1210–1250, 2003.

[143] Meyer, W. *Int. J. Quantum Chem. Symp.*, vol. 5, p. 341, 1971.

[144] Meyer, W., "PNO-CI studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane," *J. Chem. Phys.*, vol. 58, pp. 1017–1035, 1973.

[145] Mignon, P., Loverix, S., Proft, F. D., and Geerlings, P., "Influence of stacking on hydrogen bonding: Quantum chemical study on pyridine-benzene model complexes," *J. Phys. Chem. A*, vol. 108, p. 6038, 2004.

[146] Mishra, B. K. and Sathyamurthy, N., "π-π interaction in pyridine," *J. Phys. Chem. A*, vol. 109, pp. 6–8, 2005.

[147] Misquitta, A. J., Jeziorski, B., and Szalewicz, K., "Dispersion energy from density-functional theory description of monomers," *Phys. Rev. Lett.*, vol. 91, p. 033201, 2003.

[148] Misquitta, A. J., Podeszwa, R., Jeziorski, B., and Szalewicz, K., "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," *J. Chem. Phys.*, vol. 123, p. 214103, 2005.

[149] Misquitta, A. J. and Szalewicz, K., "Intermolecular forces from asymptotically corrected density functional description of monomers," *Chem. Phys. Lett.*, vol. 357, pp. 301–306, 2002.

[150] Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A., "Predicting absolute binding free energies to a simple model site," *J. Mol. Biol.*, vol. 371, pp. 1118–1134, 2007.

[151] MOLPRO, version 2010.1, a package of ab initio programs, Werner, H.-J., Knowles, P. J., Manby, F. R., Schütz, M., Celani, P., Knizia, G., Korona, T., Lindh, R., Mitrushenkov, A., Rauhut, G., Adler, T. B., Amos, R. D., Bernhardsson, A., Berning, A., Cooper, D. L., Deegan, M. J. O., Dobbyn, A. J., Eckert, F., Goll, E., Hampel, C., Hesselmann, A., Hetzer, G., Hrenar, T., Jansen, G., Köppl, C., Liu, Y., Lloyd, A. W., Mata, R. A., May, A. J., Tarroni, R., Thorsteinsson, T., Wang, M., and Wolf, A., see http://www.molpro.net.

[152] Morokuma, K., "Molecular orbital studies of hydrogen bonds. III. C=O $\cdots$ H-O hydrogen bond in $H_2CO \cdots H_2O$ and $H_2CO \cdots 2H_2O$," *J. Chem. Phys.*, vol. 55, p. 1236, 1971.

[153] Moszynski, R., Heijmen, T. G. A., and Jeziorski, B., "Symmetry-adapted perturbation theory for the calculation of Hartree-Fock interaction energies," *Mol. Phys.*, vol. 88, p. 741, 1996.

[154] Moszynski, R., Jeziorski, B., Rybak, S., Szalewicz, K., and Williams, H. L., "Many-body theory of exchange effects in intermolecular interactions. Density matrix approach and applications to He-F^-, He-HF, H_2-HF, and Ar-H_2 dimers," *J. Chem. Phys.*, vol. 100, pp. 5080–5092, 1994.

[155] Moszynski, R., Jeziorski, B., and Szalewicz, K., "Many-body theory of exchange effects in intermolecular interactions. second-quantization approach and comparison with full configuration interaction results," *J. Chem. Phys.*, vol. 100, pp. 1312–1325, 1994.

[156] Muto, Y., "Force between non-polar molecules," *Proc. Phys. Math. Soc. Jpn.*, vol. 17, p. 629, 1943.

[157] Neese, F., Hansen, A., and Liakos, D. G., "Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis," *J. Chem. Phys.*, vol. 131, p. 064103, 2009.

[158] Neese, F., Wennmohs, F., and Hansen, A., "Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method," *J. Chem. Phys.*, vol. 130, p. 114108, 2009.

[159] Neogrády, P., Pitoňák, M., and Urban, M., "Optimized virtual orbitals for correlated calculations: An alternative approach," *Mol. Phys.*, vol. 103, pp. 2141–2157, 2005.

[160] Parker, T. M., Hohenstein, E. G., Parrish, R. M., Hud, N. V., and Sherrill, C. D., "Energy component analysis and substituent effects of DNA base pair steps as a function of rise and twist," manuscript in preparation.

[161] Parr, R. G. and Yang, W., *Density-Functional Theory of Atoms and Molecules*, vol. 16 of *International Series of Monographs on Chemistry*. New York: Oxford, 1989.

[162] Patkowski, K., Szalewicz, K., and Jeziorski, B., "Third-order interactions in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 125, p. 154107, 2006.

[163] Patkowski, K., Szalewicz, K., and Jeziorski, B., "Orbital relaxation and the third-order induction energy in symmetry-adapted perturbation theory," *Theor. Chem. Acc.*, vol. 127, p. 211, 2010.

[164] Piacenza, M. and Grimme, S., "van der Waals interactions in aromatic systems: Structure and energetics of dimers and trimers of pyridine," *Chem. Phys. Chem.*, vol. 6, p. 1554, 2005.

[165] Pitonak, M. and Hesselmann, A., "Accurate intermolecular interaction energies from a combination of MP2 and TDDFT response theory," *J. Chem. Theory Comput.*, vol. 6, pp. 168–178, 2010.

[166] Pitoňák, M., Holka, F., Neogrády, P., and Urban, M., "Optimized virtual orbitals for correlated calculations: Towards large scale CCSD(T) calculations of molecular dipole moments and polarizabilities," *J. Mol. Struct. THEOCHEM*, vol. 768, p. 79, 2006.

[167] Pitoňák, M., Neogrády, P., Černý, J., Grimme, S., and Hobza, P., "Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD(T) benchmark data," *ChemPhysChem*, vol. 10, pp. 282–289, 2009.

[168] Pitoňák, M., Neogrády, P., Řezáč, J., Jurečka, P., Urban, M., and Hobza, P., "Benzene dimer: High-level wave function and density functional theory calculations," *J. Chem. Theory Comput.*, vol. 4, pp. 1829–1834, 2008.

[169] Pitoňák, M., Řezác, J., and Hobza, P., "Spin-component scaled coupled-clusters singles and doubles optimized towards calculation of noncovalent interactions," *Phys. Chem. Chem. Phys.*, vol. 12, pp. 9611–9614, 2010.

[170] Podeszwa, R., Bukowski, R., and Szalewicz, K., "Density-fitting method in symmetry-adapted perturbation theory based on Kohn-Sham description of monomers," *J. Chem. Theory Comput.*, vol. 2, pp. 400–412, 2006.

[171] Podeszwa, R., Bukowski, R., and Szalewicz, K., "Potential energy surface for the benzene dimer and perturbational analysis of π-π interactions," *J. Phys. Chem. A*, vol. 110, pp. 10345–10354, 2006.

[172] Podeszwa, R., Patkowski, K., and Szalewicz, K., "Improved interaction energy benchmarks for dimers of biological relevance," *Phys. Chem. Chem. Phys.*, vol. 12, pp. 5974–5979, 2010.

[173] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Predicting structure of molecular crystals from first principles," *Phys. Rev. Lett.*, vol. 101, p. 115503, 2008.

[174] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles," *Phys. Chem. Chem. Phys.*, vol. 11, p. 2241, 2009.

[175] Pulay, P. and Hamilton, T. P., "UHF natural orbitals for defining and starting MC-SCF calculations," *J. Chem. Phys.*, vol. 88, no. 8, pp. 4926–4933, 1988.

[176] Raghavachari, K., Trucks, G. W., Pople, J. A., and Head-Gordon, M., "A 5th-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989.

[177] Rayón, V. M. and Sordo, J. A., "On the validity of the counterpoise correction for the basis set superposition error including the fragment relaxation terms," *Theor. Chem. Acc.*, vol. 99, pp. 68–70, 1998.

[178] Rendell, A. P. and Lee, T. J., "Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks," *J. Chem. Phys.*, vol. 101, pp. 400–408, 1994.

[179] Riley, K. E., Pitonak, M., Cerny, J., and Hobza, P., "On the structure and geometry of biomolecular binding motifs (hydrogen-bonding, stacking, X-H$\cdots\pi$): WFT and DFT calculations," *J. Chem. Theory Comput.*, vol. 6, pp. 66–80, 2010.

[180] Riley, K. E., Pitoňák, M., Jurečka, P., and Hobza, P., "Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories," *Chem. Rev.*, vol. 110, pp. 5023–5063, 2010.

[181] Ringer, A. L., Figgs, M. S., Sinnokrot, M. O., and Sherrill, C. D., "Aliphatic C–H/π interactions: Methane-benzene, methane-phenol, and methane-indole complexes," *J. Phys. Chem. A*, vol. 110, pp. 10822–10828, 2006.

[182] Ringer, A. L., Senenko, A., and Sherrill, C. D., "Models of S/π interactions in protein structures: Comparison of the H_2S-benzene complex with PDB data," *Protein Sci.*, vol. 16, pp. 2216–2223, 2007.

[183] Ringer, A. L. and Sherrill, C. D., "First principles computation of lattice energies of organic solids: The benzene crystal," *Chem. Eur. J.*, vol. 14, pp. 2542–2547, 2008.

[184] Ringer, A. L. and Sherrill, C. D., "Substituent effects in sandwich configurations of multiply substituted benzene dimers are not solely governed by electrostatic control," *J. Am. Chem. Soc.*, vol. 131, pp. 4574–4575, 2009.

[185] Ringer, A. L., Sinnokrot, M. O., Lively, R. P., and Sherrill, C. D., "The effect of multiple substituents on sandwich and T-shaped π-π interactions," *Chem. Eur. J.*, vol. 12, pp. 3821–3828, 2006.

[186] Roeggen, I. and Wisloff-Nilssen, E., "On the Beebe-Linderberg 2-electron integral approximation," *Chem. Phys. Lett.*, vol. 132, pp. 154–160, 1986.

[187] Ruedenberg, K., Cheung, L. M., and Elbert, S. T., "MCSCF optimization through combined use of natural orbitals and the Brillouin-Levy-Berthier theorem," *Int. J. Quantum Chem.*, vol. 16, pp. 1069–1101, 1979.

[188] Saenger, W., *Principles of Nucleic Acid Structure*. New York: Springer-Verlag, 1984.

[189] Schutz, M., "A new, fast, semi-direct implementation of linear scaling local coupled cluster theory," *Phys. Chem. Chem. Phys.*, vol. 4, no. 16, pp. 3941–3947, 2002.

[190] Schütz, M. and Manby, F. R., "Linear scaling local coupled cluster theory with density fitting. Part I: 4-external integrals," *Phys. Chem. Chem. Phys.*, vol. 5, pp. 3349–3358, 2003.

[191] Schütz, M. and Werner, H.-J., "Low-order scaling local electron correlation methods. IV. linear scaling local coupled-cluster (LCCSD)," *J. Chem. Phys.*, vol. 114, no. 2, pp. 661–681, 2001.

[192] Schwabe, T. and Grimme, S., "Double-hybrid density functionals with long-range dispersion corrections: Higher accuracy and extended applicability," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 3397–3406, 2007.

[193] Scuseria, G. E. and Ayala, P. Y., "Linear scaling coupled cluster and perturbation theories in the atomic orbital basis," *J. Chem. Phys.*, vol. 111, pp. 8330–8343, 1999.

[194] Seo, J., Kim, I., and Lee, Y. S., "π-π interaction energies in monosubstituted-benzene dimers in parallel- and antiparallel-displaced conformations," *Chem. Phys. Lett.*, vol. 474, pp. 101–106, 2009.

[195] Shao, Y., Molnar, L. F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S. T., Gilbert, A. T. B., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P., Jr., R. A. D., Lochan, R. C., Wang, T., Beran, G. J. O., Besley, N. A., Herbert, J. M., Lin, C. Y., Voorhis, T. V., Chien, S. H., Sodt, A., Steele, R. P., Rassolov, V. A., Maslen, P. E., Korambath, P. P., Adamson, R. D., Austin, B., Baker, J., Byrd, E. F. C., Dachsel, H., Doerksen, R. J., Dreuw, A., Dunietz, B. D., Dutoi, A. D., Furlani, T. R., Gwaltney, S. R., Heyden, A., Hirata, S., Hsu, C.-P., Kedziora, G., Khalliulin, R. Z., Klunzinger, P., Lee, A. M., Lee, M. S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E. I., Pieniazek, P. A., Rhee, Y. M., Ritchie, J., Rosta, E., Sherrill, C. D., Simmonett, A. C., Subotnik, J. E., Woodcock, H. L., Zhang, W., Bell, A. T., Chakraborty, A. K., Chipman, D. M., Keil, F. J., Warshel, A., Hehre, W. J., Schaefer, H. F., Kong, J., Krylov, A. I., Gill, P. M. W., and Head-Gordon, M., "Advances in methods and algorithms in a modern quantum chemistry program package," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 3172–3191, 2006.

[196] Sherrill, C. D. and Schaefer, H. F., "Compact variational wave functions incorporating limited triple and quadruple substitutions," *J. Phys. Chem.*, vol. 100, pp. 6069–6075, 1996.

[197] Sherrill, C. D., Sumpter, B. G., Sinnokrot, M. O., Marshall, M. S., Hohenstein, E. G., Walker, R. C., and Gould, I. R., "Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π-π, CH/π, and SH/π interactions," *J. Comput. Chem.*, vol. 30, pp. 2187–2193, 2009.

[198] Sherrill, C. D., Takatani, T., and Hohenstein, E. G., "An assessment of theoretical methods for nonbonded interactions: Comparison to complete basis set limit coupled-cluster potential energy curves for the benzene dimer, the methane dimer, benzene-methane, and benzene-H_2S," *J. Phys. Chem. A*, vol. 113, pp. 10146–10159, 2009.

[199] Shieh, H.-S., Berman, H. M., Dabrow, M., and Neidle, S., "The structure of drug-deoxydinucleoside phosphate complex; generalized conformational behavior of intercalation complexes with RNA and DNA fragments," *Nucleic Acids Res.*, vol. 8, pp. 85–97, 1980.

[200] Singh, N. J., Min, S. K., Kim, D. Y., and Kim, K. S., "Comprehensive energy analysis for various types of π-interaction," *J. Chem. Theory Comput.*, vol. 5, pp. 515–529, 2009.

[201] Sinnokrot, M. O. and Sherrill, C. D., "Highly accurate coupled cluster potential energy curves for benzene dimer: The sandwich, T-shaped, and parallel-displaced configurations," *J. Phys. Chem. A*, vol. 108, no. 46, pp. 10200–10207, 2004.

[202] Sinnokrot, M. O. and Sherrill, C. D., "Substituent effects in π-π interactions: Sandwich and t-shaped configurations," *J. Am. Chem. Soc.*, vol. 126, pp. 7690–7697, 2004.

[203] Sinnokrot, M. O. and Sherrill, C. D., "High-accuracy quantum mechanical studies of π-π interactions in benzene dimers," *J. Phys. Chem. A*, vol. 110, pp. 10656–10668, 2006.

[204] Sinnokrot, M. O., Valeev, E. F., and Sherrill, C. D., "Estimates of the ab initio limit for π-π interactions: The benzene dimer," *J. Am. Chem. Soc.*, vol. 124, pp. 10887–10893, 2002.

[205] Smith, Q. A., Gordon, M. S., and Slipchenko, L. V., "Benzene–pyridine interactions predicted by the effective fragment potential method," *J. Phys. Chem. A*, vol. 115, p. 4598, 2011.

[206] Šponer, J., Leszczynski, J., and Hobza, P., "Nature of nucleic acidbase stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs," *J. Phys. Chem.*, vol. 100, pp. 5590–5596, 1996.

[207] Stevens, W. and Fink, W., "Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer," *Chem. Phys. Lett.*, vol. 139, pp. 15–22, 1987.

[208] Stone, A. J., "Distributed multipole analysis, or how to describe a molecular charge-distribution," *Chem. Phys. Lett.*, vol. 83, pp. 233–239, 1981.

[209] Stone, A. J. and Misquitta, A. J., "Charge-transfer in symmetry-adapted perturbation theory," *Chem. Phys. Lett.*, vol. 473, p. 201, 2009.

[210] Stone, A. J., *The Theory of Intermolecular Forces*. Oxford: Oxford University Press, 1996.

[211] Subotnik, J. E., Sodt, A., and Head-Gordon, M., "A near linear-scaling smooth local coupled cluster algorithm for electronic structure," *J. Chem. Phys.*, vol. 125, p. 074116, 2006.

[212] Szalewicz, K. and Jeziorski, B., "Symmetry-adapted perturbation theory of intermolecular interactions," in *Molecular Interactions: From van der Waals to Strongly Bound Complexes* (Scheiner, S., ed.), p. 3, Wiley, 1997.

[213] Takatani, T., Hohenstein, E. G., Malagoli, M., Marshall, M. S., and Sherrill, C. D., "Basis set consistent revision of the S22 test set of noncovalent interaction energies," *J. Chem. Phys.*, vol. 132, p. 144104, 2010.

[214] Takatani, T., Hohenstein, E. G., and Sherrill, C. D., "Improvement of the coupled-cluster singles and doubles method via scaling same- and opposite-spin components of the double excitation correlation energy," *J. Chem. Phys.*, vol. 128, p. 124111, 2008.

[215] Takatani, T. and Sherrill, C. D., "Performance of spin-component-scaled Møller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 6106–6114, 2007.

[216] Takatsuka, A., Ten-no, S., and Hackbusch, W., "Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller-Plesset perturbation theories," *J. Chem. Phys.*, vol. 129, p. 044112, 2008.

[217] Tarakeshwar, P., Choi, H. S., and Kim, K. S., "Olefinic vs aromatic π-H interaction: A theoretical investigation of the nature of interaction of first-row hydrides with ethene and benzene," *J. Am. Chem. Soc.*, vol. 123, p. 3323, 2001.

[218] Tarakeshwar, P., Choi, H. S., Lee, S. J., Kim, K. S., Ha, T., Jang, J. H., Lee, J. G., and Lee, H., "A theoretical investigation of the nature of the π-H interaction in ethene-H_2O, benzene-H_2O, and benzene-$(H_2O)_2$," *J. Chem. Phys.*, vol. 111, p. 5838, 1999.

[219] Taube, A. G. and Bartlett, R. J., "Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory," *Collect. Czech. Chem. Commun.*, vol. 70, pp. 837–850, 2005.

[220] Tauer, T. P., Derrick, M. E., and Sherrill, C. D., "Estimates of the ab initio limit for sulfur-π interactions: The H_2S-Benzene dimer," *J. Phys. Chem. A*, vol. 109, no. 1, pp. 191–196, 2005.

[221] Thanthiriwatte, K. S., Hohenstein, E. G., Burns, L. A., and Sherrill, C. D., "Assessment of the performance of DFT and DFT-D methods for describing distance dependence of hydrogen-bonded interactions," *J. Chem. Theory Comput.*, vol. 7, pp. 88–96, 2011.

[222] Tomasi, J., Mennucci, B., and Cammi, R., "Quantum mechanical continuum solvation models," *Chem. Rev.*, vol. 105, pp. 2999–3094, 2005.

[223] Tomasi, J. and Persico, M., "Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent," *Chem. Rev.*, vol. 94, pp. 2027–2094, 1994.

[224] Tschumper, G., "Reliable electronic structure computations for weak noncovalent interactions in clusters," in *Reviews in Computational Chemistry, Volume 26* (Lipkowitz, K. and Cundari, T., eds.), pp. 39–89, Wiley, 2009.

[225] Tsuzuki, S. and Lüthi, H. P., "Interaction energies of van der Waals and hydrogen bonded systems calculated using density functional theory: Assessing the PW91 model," *J. Chem. Phys.*, vol. 114, no. 9, pp. 3949–3957, 2001.

[226] Tsuzuki, S., Mikami, M., and Yamada, S., "Origin of attraction, magnitude, and directionality of interactions in benzene complexes with pyridinium cations," *J. Am. Chem. Soc.*, vol. 129, pp. 8656–8662, 2007.

[227] Tsuzuki, S., Uchimaru, T., and Tanabe, K., "Basis set effects on the intermolecular intaraction of hydrocarbon molecules obtained by ab initio molecular orbital method: Evaluation of dispersion energy," *J. Mol. Struct. THEOCHEM*, vol. 307, pp. 107–118, 1994.

[228] Vahtras, O., Almlöf, J., Feyereisen, and M, W., "Integral approximations for LCAO-SCF calculations," *Chem. Phys. Lett.*, vol. 213, pp. 514–518, 1993.

[229] Valiev, M., Bylaska, E., Govind, N., Kowalski, K., Straatsma, T., van Dam, H., Wang, D., Nieplocha, J., Apra, E., Windus, T., and de Jong, W., "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," *Comput. Phys. Commun.*, vol. 181, p. 1477, 2010.

[230] van Duijneveldt, F. B., van Duijneveldt-van de RijdtJ, J. G. C. M., and van Lenthe, J. H., "State of the art in counterpoise theory," *Chem. Rev.*, vol. 94, pp. 1873–1885, 1994.

[231] Vazquez-Mayagoitia, A., Sherrill, C. D., Apra, E., and Sumpter, B. G., "An assessment of density functional methods for potential energy curves of nonbonded interactions: The XYG3 and B97-D approximations," *J. Chem. Theory Comput.*, vol. 6, pp. 727–734, 2010.

[232] von Lilienfeld, O. A., Tavernelli, I., Rothlisberger, U., and Sebastiani, D., "Optimization of effective atom centered potentials for London dispersion forces in density functional theory," *Phys. Rev. Lett.*, vol. 93, no. 15, p. 153004, 2004.

[233] von Lilienfeld, O. A., Tavernelli, I., Rothlisberger, U., and Sebastiani, D., "Performance of optimized atom-centered potentials for weakly bonded systems using density functional theory," *Phys. Rev. B*, vol. 71, p. 195119, 2005.

[234] Wang, W. and Hobza, P., "Theoretical study on the complexes of benzene with isoelectronic nitrogen-containing heterocycles," *Chem. Phys. Chem.*, vol. 9, p. 1003, 2008.

[235] Watt, M., Hardebeck, L. K. E., Kirkpatrick, C. C., and Lewis, M., "Face-to-face arene-arene binding energies: Dominated by dispersion but predicted by electrostatic and dispersion/polarizability substituent constants," *J. Am. Chem. Soc.*, vol. 133, pp. 3854–3862, 2011.

[236] Weigend, F., "A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.*, vol. 4, pp. 4285–4291, 2002.

[237] Weigend, F., Kattannek, M., and Ahlrichs, R., "Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods," *J. Chem. Phys.*, vol. 130, p. 164106, 2009.

[238] Weigend, F., Köhn, A., and Hättig, C., "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," *J. Chem. Phys.*, vol. 116, pp. 3175–3183, 2002.

[239] Werner, H.-J., Manby, F. R., and Knowles, P. J., "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations," *J. Chem. Phys.*, vol. 118, no. 18, pp. 8149–8160, 2003.

[240] Wheeler, S. E. and Houk, K. N., "Substituent effects in the benzene dimer are due to direct interactions of the substituents with the unsubstituted benzene," *J. Am. Chem. Soc.*, vol. 130, pp. 10854–10855, 2008.

[241] Wheeler, S. E. and Houk, K. N., "Origin of substituent effects in edge-to-face aryl-aryl interactions," *Mol. Phys.*, vol. 107, pp. 749–760, 2009.

[242] Wheeler, S. E. and Houk, K. N., "Through-space effects of substituents dominate molecular electrostatic potentials of substituted arenes," *J. Chem. Theory Comput.*, vol. 5, pp. 2301–2312, 2009.

[243] Wheeler, S. E., McNeil, A. J., Mueller, P., Swager, T. M., and Houk, K. N., "Probing substituent effects in aryl-aryl interactions using stereoselective diels alder cycloadditions," *J. Am. Chem. Soc.*, vol. 132, pp. 3304–3311, 2010.

[244] Whitten, J. L., "Coulombic potential-energy integrals and approximations," *J. Chem. Phys.*, vol. 58, pp. 4496–4501, 1973.

[245] Williams, H. L. and Chabalowski, C. F., "Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," *J. Phys. Chem. A*, vol. 105, pp. 646–659, 2001.

[246] Williams, H. L., Szalewicz, K., Jeziorski, B., Moszynski, R., and Rybak, S., "Symmetry-adapted perturbation theory calculation of the Ar-H_2 intermolecular potential energy surface," *J. Chem. Phys.*, vol. 98, pp. 1279–1292, 1993.

[247] Williams, H., Szalewicz, K., Moszynski, R., and Jeziorski, B., "Dispersion energy in the coupled pair approximation with noniterative inclusion of single and triple excitations," *J. Chem. Phys.*, vol. 103, p. 4586, 1995.

[248] Wu, Q. and Yang, W., "Empirical correction to density functional theory for van der Waals interactions," *J. Chem. Phys.*, vol. 116, pp. 515–524, 2002.

[249] Zhang, Y., Xu, X., and Goddard, W. A., "Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics," *Proc. Natl. Acad. Sci. USA*, vol. 106, pp. 4963–4968, 2009.

[250] Zhao, Y., Schultz, N. E., and Truhlar, D. G., "Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions," *J. Chem. Phys.*, vol. 123, p. 161103, 2005.

[251] Zhao, Y., Schultz, N. E., and Truhlar, D. G., "Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions," *J. Chem. Theory Comput.*, vol. 2, pp. 364–382, 2006.

[252] Zhao, Y. and Truhlar, D. G., "Benchmark databases for nonbonded interactions and their use to test density functional theory," *J. Chem. Theory Comput.*, vol. 1, pp. 415–432, 2005.

[253] Zhao, Y. and Truhlar, D. G., "A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions," *J. Chem. Phys.*, vol. 125, p. 194101, 2006.

[254] Zhao, Y. and Truhlar, D. G., "Density functionals for noncovalent interaction energies of biological importance," *J. Chem. Theory Comput.*, vol. 3, pp. 289–300, 2007.

[255] Zimmerli, U., Parrinello, M., and Koumoutsakos, P., "Dispersion corrections to density functionals for water aromatic interactions," *J. Chem. Phys.*, vol. 120, pp. 2693–2699, 2004.

VITA

Edward G. Hohenstein was raised in Toms River, NJ. He graduated from Toms River High School East in 2003. Then, he attended Washington College, a small, liberal arts school in Chestertown, MD on the eastern shore of the Chesapeake bay. Starting in the summer of 2006, he worked in the laboratory of Dr. James R. Locker until his graduation in May of 2007. He graduated *summa cum laude* from Washington College with majors in Chemistry and Mathematics. Shortly thereafter, he began his graduate research with Dr. C. David Sherrill at Georgia Tech.