UNIVERSITY OF CALIFORNIA
RIVERSIDE

Fast and Accurate Electronic Structure Methods for Predicting Two- and Three-Body Noncovalent Interactions.

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Yuanhang Huang

June 2015

Dissertation Committee:

Professor Gregory J. O. Beran, Chairperson
Professor Chia-en A. Chang
Professor Jingsong Zhang

The Dissertation of Yuanhang Huang is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisor Prof. Beran who helped me to realize my dream to do scientific research. With his support and guidance, I gradually built a solid background in theoretical chemistry and found the interesting areas I want to explore in the future. He gave me many interesting projects to work on. Instead of providing the answers directly, he always encouraged me to think about the problems independently and shared his thought to guide me proceeding in the right direction. Prof. Beran always conducts himself as a gentleman and sets a good example for his students. Not only a good listener, he is ready for helpful discussions whenever I get stuck. More than just scientific research, he also helped me to improve the abilities in presentations, programming, data analysis, formal writings, etc. I will be forever grateful for the wonderful five years' study and research time with Prof. Beran.

I also want to thank the other members in the Beran group: Shuhao Wen, Yin Luo, Kaushik Nanda, Kelly Theel, Joshua Hartman, Yonaton Heit, Jessica McKinley and Dominique Nocito. Besides discussing interesting scientific problems, they helped me a lot so I can adjust to the new environment and start my work quickly. I am also appreciate the help from Dr. Alston Misquitta on the use of CamCASP and the interesting data shared by Prof. David Sherrill's group. I am glad to have the chance to collaborate with Dr. Yihan Shao on MC-MP2C project, Matthew Goldey and Prof. Martin Head-Gordon on attenuated MP2C project, Dr. Jan Rezac and Prof. Pavel Hobza on three-body interaction benchmarking project. The financial supports from U.S. National Science Foundation (CHE-1112568 and CHE-1362465) and supercomputer time from XSEDE (TG-CHE110064) are gratefully

acknowledged.

I would like to thank Prof. Qunxiang Li, Shulai Lei and Ke Xu in USTC, who guided me to start research in quantum chemistry. I also feel grateful for all the teachers and friends I have ever met. They taught me about the wonderful world, shaped my mind and character, brought me a lot of fun, helped me and encouraged me be stronger to enjoy any challenges.

Finally, thank my parents for all the years' constant support and love. No matter when and where, I know they are always with me.

To Mengfei

ABSTRACT OF THE DISSERTATION

Fast and Accurate Electronic Structure Methods for Predicting Two- and Three-Body Noncovalent Interactions.

by

Yuanhang Huang

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, June 2015
Professor Gregory J. O. Beran, Chairperson

Noncovalent interactions are ubiquitous in chemistry. As a source of stabilization, they play an important role in many interesting chemical processes, such as protein folding, molecular recognition, molecular self-assembly, physical adsorption, etc. Accurate energy predictions from first principles on many-body systems like molecular crystals requires electronic structure methods able to describe various types of noncovalent interactions like hydrogen bonding, electrostatic, induction, and van der Waals dispersion across different intramolecular conformations and intermolecular arrangements with high and uniform accuracy. Besides accuracy, computational efficiency should be considered for practical applications. Here, fast and accurate electronic methods are developed to treat both two-body and three-body noncovalent interactions.

For two-body interactions, the MP2C method developed by Pitonak and Hesselmann proves to be a reliable method with affordable computational cost. To improve the computational efficiency of MP2C dispersion correction, we propose the use of monomer-

centered basis sets instead of dimer-centered ones. For an individual dimer, this change accelerates the dispersion correction several-fold. For molecular crystals, 100-fold speedups for dispersion correction calculation are achieved by utilizing translational symmetry. To improve the computational efficiency of the MP2 part in MP2C method, we demonstrate that one can avoid calculating the unnecessary long-range MP2 correlations by attenuating the Coulomb operator, allowing the dispersion correction to handle the long-range interactions inexpensively. Utilizing excellent fortuitous cancellations between finite basis set errors, attenuation errors and correlation errors, further computational savings could be achieved by the use of small basis set to approach complete basis set limit quality results.

For three-body interactions, which are challenging for many widely-used, low-cost electronic structure methods, we propose a straightforward model that corrects conventional MP2 with a damped three-body Axilrod-Teller-Muto dispersion correction. The damping function compensates for the absence of higher-order dispersion contributions and non-additive short-range exchange terms not found in MP2. Examinations on trimer benchmark set and benzene crystal demonstrate the reliability of this model for various applications.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With tremendous progress in computer technology and quantum mechanical theory, people are now able to study large and/or condensed-phase systems from first-principles calculations, which seemed infeasible a few decades back. For chemical systems people are interested in, energy is one of the most fundamental properties, from which thermal and kinetic properties can be derived. Accurate energy prediction is always the main research area in theoretical and computational chemistry, since it determines how reliable one can apply the method to study other properties. Noncovalent interactions play an ubiquitous role in chemical systems, they are the main stabilization source for molecular self-assembly, biorecognition, physical adsorption, etc., for instance. Unlike the electronic energy of single molecule where electrons are highly correlated to form chemical bond, noncovalent interactions are weak correlations between two or more noncontact units. Because the interaction energies are much weaker, even modest changes of the correlation model and basis set can introduce errors that are comparable in magnitude to the energy itself. Accurate energy

prediction of noncovalent interactions has become an important criteria for the quality of a theoretical model [1, 2].

In addition to their importance in intermolecular interactions, noncovalent interactions can also be found between different functional groups in large molecules, like different domains in a protein, for instance. Molecular crystals represent another well-known many-body systems involving a variety of noncovalent interactions. So they are good systems for people to examine the performance of different theoretical methods. To predict molecular crystal lattice energies, we first introduce an effective fragment-based QM/MM model. In the QM part, reliable quantum mechanical methods are necessary for capturing the interactions between different fragments accurately, and this thesis focuses on the development of computationally practical quantum mechanical methods which can be used in such models. Accordingly, the next several sections also introduce the standard electronic structure methods which provide the starting point for our developments here. These methods range from the self-consistent field algorithm based on a single particle approximation to various electron correlation approximations. To make robust energy predictions, the completeness of the basis set used to expand electron densities or molecular orbitals is also a necessary condition. Finally, we will introduce several commonly-used theoretical methods for accurately predicting noncovalent interactions, from symmetry adapted perturbation theory (SAPT) which gives physical picture in understanding different energy component contributions quantitatively, to a multipole moment expanded force field model which can be taken as simplified SAPT, and to supermolecular approach allowing the use of almost all the current quantum chemistry methods.

1.1 Motivation: molecular crystal lattice energy prediction

People often appreciate how magic the world is, that protein folding is directed by an invisible "hand" (noncovalent interactions). This driven force engineers the polypeptide chains into some unique and complex spatial structures. The similar phenomenon can also be observed in molecular crystal growth that the well separated molecules condense into periodic packing motifs, spontaneously at the first glance! Unlike gas and liquid phases, molecular crystals are in solid state with more compact and stable structures, so they can be stored and transported easily in practice. For example, pharmaceutical drugs are generally made into crystals for convenience. However, a lot of molecular crystals have different packing structures or polymorphs with different solubilities and bio-availabilities. There have been several cases where commercial drugs (such as Ritonavir for HIV disease and Rotigotine for Parkinson's disease) were recalled from the market due to the occurrence of insoluble polymorphs, causing a great financial loss.

Theoretical molecular crystal structure prediction from first principles is potentially much cheaper than experimentally synthesizing, purifying drug molecules, and testing polymorphs formed under different conditions, and thus it could be used to identify and avoid polymorphism problems in advance. Under the constraints of current computational affordability and theoretical capability, a great efforts have been made to develop theoretical approaches that are promising to handle these large and complex systems. Among them, fragment based methods provide a lower-cost alternative by partitioning the periodic system into small fragments and dealing with each small fragments individually [3–5]. Compared to traditional techniques which calculate the entire periodic crystal at the same level of

theory, these divide-and-conquer models allow one to treat different fragments with different levels of theory in order to satisfy different cost and accuracy requirements. Fragment approaches also avoid the need to translate the expensive QM calculations into reciprocal space. In other words, this approach leads to very flexible computational models, which can be expressed as

$$E_{total} = E_{1\text{-}body} + \Delta E_{2\text{-}body} + \Delta E_{3\text{-}body} + ... \tag{1.1}$$

where the 1-body and 2-body terms are comparatively small in both fragment size and quantity, while they contribute to most of the total energy. On the contrary, the higher order many-body interaction terms are generally cost-prohibitive to evaluate. To make the lattice energy calculations feasible, many-body terms are usually truncated only with the main contribution terms left in the model. However, these many-body terms cannot be ignored if one wants to achieve high accuracy energy prediction. Instead, we use more accurate and expensive quantum mechanical (QM) methods to deal with 1-body and short-range 2-body interaction terms, while the long-range and many-body interaction terms are approximated with much cheaper and less accurate molecular mechanical (MM) methods. Based on this scheme, the hybrid QM/MM many-body interactions (HMBI) model [3, 6–9] to approximate crystal energy is

$$E_{total}^{HMBI} = E_{total}^{MM} + \sum_i (E_i^{QM} - E_i^{MM}) + \sum_{ij} d_{ij}(R)(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM}) \tag{1.2}$$

where E_i and $\Delta^2 E_{ij}$ are 1-body and 2-body interaction energies, respectively. The MM total energy E_{total}^{MM} can also be expanded as

$$E_{total}^{MM} = E_{1\text{-}body}^{MM} + \Delta E_{2\text{-}body}^{MM} + \Delta E_{many\text{-}body}^{MM} \tag{1.3}$$

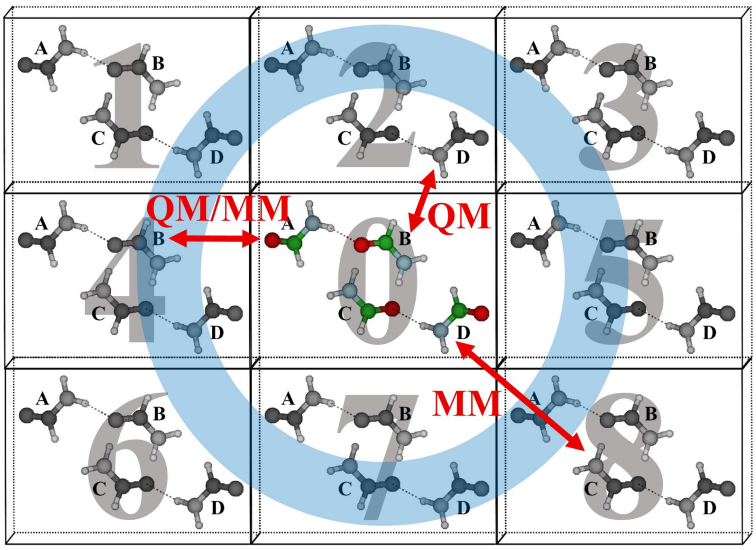

Figure 1.1: Schematic of the HMBI model. Centered on the original unit cell labeled as 0, the periodic crystal is divided into QM region, MM region, and QM/MM buffer region (blue shadow).

In the HMBI model, a multipole expanded *ab initio* force field (AIFF), [3, 9–15] which will be discussed in detail later, is used in MM part. The $d_{ij}(R)$ term in Equation 1.2 ensures a smooth transition in potential energy surface between short-range QM and long-range MM two-body interaction regimes. Its value decays from 1 to 0 as the intermolecular distance R increases from the QM boundary to MM boundary in the buffer area (Figure 1.1).

Fragment methods like HMBI make it more practical to employ high-quality electronic structure methods toward the computation of molecular crystal lattice energies, but accurate prediction of these energies remains challenging because it requires [16] methods that can: (1) resolve ∼1-10 kJ/mol energy differences between polymorphs, (2) balance inter- and intra-molecular interactions, (3) describe different types of intermolecular interactions with high and uniform accuracy, meaning these methods are reliable for a wide variety of systems, (4) be systematically improvable for both correlation methods and basis

sets, and (5) be computationally affordable to study interesting large-sized molecules. Before discussing theoretical methods for noncovalent interactions in detail, the next sections review some basic concepts in electronic structure theory.

1.2 Theoretical background

The quantum mechanical theory developed in last century enables people to study the physical world at the atom and electron levels. Applying the Schrödinger equation to study chemical systems enriches people's understanding of atomic and molecular properties. Limited by the correlation problem that only quite a few simple systems can be solved accurately, most of the time the equation is solved more approximately. The very basic approximations are the Born-Oppenheimer approximation to separate the electronic and nuclear degrees of freedom, and the single particle approximation to partially decouple electrons from each other, which will be discussed first. To improve the mean-field approximation, electron correlation effects should be included, which is the core problem of modern quantum chemistry. Besides correlation, the completeness of the basis set expanding the electron density or molecular orbitals cannot be ignored if one wants to achieve high accuracy.

1.2.1 Single particle approximation

Start with the non-relativistic time-independent Schrödinger equation

$$\hat{H}\Psi(R,r) = E\Psi(R,r) \tag{1.4}$$

where $\hat{H}$ is Hamiltonian, Ψ is wavefunction containing coordinates of nucleus (R) and electron (r). The eigenvalue E obtained by solving the equation is the total energy of the system. The expression of Hamiltonian in atomic unit is

$$\begin{aligned}\hat{H} &= -\sum_I \frac{\hbar^2}{2M_I}\nabla^2_{R_i} - \sum_i \frac{\hbar^2}{2}\nabla^2_{r_i} - \sum_{iI}\frac{Z_I}{|R_I - r_i|} + \frac{1}{2}\sum_{ij,i\neq j}\frac{1}{|r_i - r_j|} + \frac{1}{2}\sum_{IJ,I\neq J}\frac{Z_I Z_J}{|R_I - R_J|} \\ &= \sum_I \hat{T}_I + \sum_i \hat{T}_i + \sum_{iI}\hat{U}_{iI} + \frac{1}{2}\sum_{ij,i\neq j}\hat{J}_{ij} + \frac{1}{2}\sum_{IJ,I\neq J}\hat{J}_{IJ}\end{aligned} \tag{1.5}$$

where the operator of total energy is expressed as the sum of operators for different energy components: kinetic energy of nuclei ($\sum_I \hat{T}_I$, first term on the right), kinetic energy of electrons ($\sum_i \hat{T}_i$, second term on the right), Coulomb interaction between nuclei and electrons ($\sum_{iI}\hat{U}_{iI}$, the third term on the right), Coulomb interaction between electrons ($\frac{1}{2}\sum_{ij,i\neq j}\hat{J}_{ij}$, the fourth), and the last term $\frac{1}{2}\sum_{IJ,I\neq J}\hat{J}_{IJ}$ for nucleus-nucleus Coulomb interaction. M_I is nuclear mass. R_I and r_i stand for nuclear and electronic position, respectively. $\hbar$ is $h/2\pi$ where h is Planck constant, and Z in the equation above is nuclear charge.

Electrons move several orders of magnitude faster than nuclei, which means they can respond to changes in the nuclear positions almost instantaneously. Born-Oppenheimer approximation assumes nuclear coordinates are fixed parameters, so only electronic motions are considered at some certain instant. So the nucleus-nucleus Coulomb interactions $\sum_{IJ,I\neq J} J_{IJ}$ become a constant that can be taken out from Hamiltonian (Equation 1.5). This leaves only terms with related to electrons

$$\hat{H} = \sum_i \hat{T}_i + \sum_{iI}\hat{U}_{iI} + \frac{1}{2}\sum_{ij,i\neq j}\hat{J}_{ij} \tag{1.6}$$

To further simplify the Schrödinger equation, the single particle Hartree-Fock the-

ory is used to approximate the pair-wise Coulomb interaction as the interaction of a selected electron in the mean field generated by all other electrons. Thus molecular orbitals obtained are independent with each other (Hartree product). Since electrons are fermions constrained by the Pauli exclusion principle, the wavefunction Ψ is expressed in form of Slater determinant

$$\Psi^{HF}(r_1, r_2 ... r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(r_1) & \psi_1(r_2) & ... & \psi_1(r_N) \\ \psi_2(r_1) & \psi_2(r_2) & ... & \psi_2(r_N) \\ . & . & ... & . \\ \psi_N(r_1) & \psi_N(r_2) & ... & \psi_N(r_N) \end{vmatrix} \tag{1.7}$$

in which ψ_i is molecular orbital (MO), and r_i stands for electron position. $\sqrt{N!}$ ensures the total wavefunction Ψ^{HF} to be normalized. The total Hartree-Fock energy can be expressed as the expectation value of the Hamiltonian

$$\begin{aligned} E^{HF} &= \langle \Psi^{HF} | \hat{H} | \Psi^{HF} \rangle \\ &= \sum_{iI} \langle \psi_i | \hat{T}_i + \hat{U}_{iI} | \psi_i \rangle + \frac{1}{2} \sum_{i,j} \langle \psi_i \psi_j | \hat{J}_{ij} (1 - \hat{P}_{ij}) | \psi_i \psi_j \rangle \end{aligned} \tag{1.8}$$

where the operator $\hat{P}_{ij}$ is permutation operator, which stems from expanding the Slater determinant to exchange electrons $\hat{P}_{ij}|\psi_j \psi_i\rangle = |\psi_i \psi_j\rangle$ and $\hat{P}_{ii} = 1$. In order to obtain the Hartree-Fock energy, one should know each molecular orbital ψ_i first. The variational principle is applied subject to the constraint that the MOs remain orthonormal,

$$\delta \left[E^{HF} - \sum_i \epsilon_i \Big(\langle \psi_i | \psi_j \rangle - 1 \Big) \right] = 0 \tag{1.9}$$

where the Lagrange multiplier ϵ_i can be understand as orbital energy of electron i. After replacing E^{HF} with the expression in Equation 1.8, and differentiating bra orbital $\langle \psi_i |$, one

will get

$$\langle \delta\psi_i | \hat{T}_i + \hat{U}_i + \sum_j \langle \psi_j | \hat{J}_{ij} - \hat{K}_{ij} | \psi_j \rangle - \epsilon_i | \psi_i \rangle = \langle \delta\psi_i | \hat{f}_i - \epsilon_i | \psi_i \rangle = 0 \tag{1.10}$$

where $\hat{f}_i$ is called Fock operator that $\sum_i \hat{f}_i = \hat{H}^{HF}$ and Coulomb exchange operator is defined as $\hat{K}_{ij} = \hat{J}_{ij}\hat{P}_{ij}$. The variational principle holds true for any molecular orbitals $\langle \psi_i |$, so one can obtain a set of orbital equations as

$$\hat{f}_i | \psi_i \rangle = \epsilon_i | \psi_i \rangle \tag{1.11}$$

with orbital index i ranges from 1 to N. A self-consistent field (SCF) iteration procedure is applied to solve both orbital ψ_i and energy ϵ_i, starting from a full set of initial guess orbitals $\psi'_1, \psi'_2 ... \psi'_N$ built on linear superposition of basis functions $\{\phi_i\}$.

Similar to Hartree-Fock theory in using single particle approximation and self-consistent field iterations, density-functional theory (DFT) provides an alternative view on the many-electron system based on the Hohenberg-Kohn theorem [17] and Kohn-Sham equation [18]. The first Hohenberg-Kohn theorem proves that the ground state electron density $n_i(r)$ is uniquely determined given an external potential $V_{ext}(r)$. The potential might be the nuclear potential or some outer field from the environment. Instead of providing rigorous proof, one can briefly imagine that electron densities n_r fluctuate in the fixed external potential (assume the system is isolated from the environment, the ionic potential will be the only potential. Also, there is no structure vibration under B-O approximation) due to the many-electron Coulomb interactions. The total electronic energy changes with the electron fluctuations, and the energy will head toward to a minimum until the Coulomb repulsions between electrons in the fixed attractive ionic potential are balanced at certain

state (variational principle). The system will not jump to other unstable states because it is constrained within the potential well, unless the external potential changes to create a new global minimum. So the ground state energy is the functional of electron density affected only by the external potential

$$E[n(r)] = F[n(r)] + \int V_{ext}(r)n(r)dr \tag{1.12}$$

where $F[n(r)]$ is an universal functional describing the behavior of electrons in a fixed field. It includes kinetic energy of electrons and electron-electron Coulomb interaction that only depends on electron density, which is common in all systems [19]. Because the complicated electron correlation, the kinetic energy cannot be expressed unless one uses single particle approximation to separate the selected electron from the others

$$T[n(r)] = -\frac{1}{2}\sum_i \int \langle \psi_i(r)|\nabla^2|\psi_i(r)\rangle dr \tag{1.13}$$

Under this approximation, the correlation effect and exchange effect (as expressed in Equation 1.8) also depend only on electron density $n(r)$ that they are grouped into one term $E_{XC}[n(r)]$. The total energy expression could be further expanded as

$$E[n(r)] = T[n(r)] + \frac{1}{2}\int\int \frac{n(r)n(r')}{|r-r'|}drdr' + E_{XC}[n(r)] + \int V_{ext}n(r)dr \tag{1.14}$$

Solving Kohn-Sham equation according to the variational principle,

$$\frac{\delta}{\delta n(r)}\left[E[n(r)] - \epsilon_i \int n(r)dr\right] = 0 \tag{1.15}$$

where normalization constraint on electron density $\int n(r)dr = N$ is applied. The final Kohn-Sham equation for single electron becomes

$$\left[-\frac{1}{2}\nabla_i^2 + V_{ext}(r) + \int \frac{n(r')}{|r-r'|}dr' + V_{XC}(r)\right]\psi_i(r) = \epsilon_i\psi_i(r) \tag{1.16}$$

Because the single particle approximation assumes electron densities are independent with each other, exchange-correlation potential $V_{XC}(r) = \delta E_{XC}[n(r)]/\delta n(r)$ only depends on single electron density (coordinate of only one electron), which means it must be localized as electron density. While this is not true for the real system, as one can see in Equation 1.10 that the Coulomb exchange term K_{ij} involves the coordinates of both electron i and j, instead of single electron. Thus most of the approximated exchange functionals suffer a deficiency of local exchange, introducing an error usually called self-interaction energy. Though the correct form of the potential is currently unknown, some exchange-correlation functionals like local-density approximations (LDA), generalized gradient approximations (GGA), and hybrid functionals have been developed to study highly-correlated systems. The great advantage of this approach with single particle approximation is the low $O(N^3)$ scaling with system size to solve Kohn-Sham equation, making it an efficient and widely used tool, especially for large systems.

1.2.2 Electron correlation

The electron correlation problem is probably one of the most important problems in quantum chemistry. Many methods have been developed to treat electron correlations in order to satisfy the increasing requirements for high accuracy and low computational cost [20,21]. Like the classical many-body correlation problem (currently people still cannot predict the exact trajectory of three interacting ideal particles in isolated systems), even approximate treatments of electron-electron correlation can be computationally demanding. By definition, the correlation energy can be expressed as the difference between the exact

energy and the Hartree-Fock energy

$$E^{Corr} = E^{Exact} - E^{HF} \tag{1.17}$$

There are several main approaches to approximate correlation energy. The first one introduced here is Møller-Plesset perturbation theory. As discussed in last section, the Hartree-Fock Hamiltonian and Fock operator are related as

$$\hat{H}^{HF} = \sum_i \hat{f}_i = \sum_i (\hat{h}_i + V^{HF}(i)) \tag{1.18}$$

where core-hamiltonian operator $\hat{h}_i$ is $\hat{T}_i + \hat{U}_i$, and Hartree-Fock potential $V^{HF}(i)$ includes Coulomb and exchange interactions as $\sum_j (\hat{J}_{ij} - \hat{K}_{ij})$. Introducing perturbation operator $\hat{V}$ as the difference between the exact electron Hamiltonian and Hartree-Fock Hamiltonian (based on single particle approximation)

$$\begin{aligned} \hat{V} &= \hat{H} - \hat{H}^{HF} \\ &= \sum_i (\hat{h}_i + \frac{1}{2} \sum_{ij, i \neq j} \hat{J}_{ij}) - \sum_i (\hat{h}_i + V^{HF}(i)) \\ &= \sum_i \left(\frac{1}{2} \sum_{ij, i \neq j} \hat{J}_{ij} - V^{HF}(i) \right) \end{aligned} \tag{1.19}$$

In Rayleigh-Schrödinger perturbation theory, the unperturbed energy is the expectation value of the unperturbed operator on unperturbed wavefunctions. Møller-Plesset perturbation theory uses Hartree-Fock operator and wavefunctions

$$E^{MP0} = \langle \Psi^{HF} | \hat{H}^{HF} | \Psi^{HF} \rangle = \sum_i \epsilon_i \tag{1.20}$$

where Hartree-Fock Hamiltonian is used instead of exact Hamiltonian as Equation 1.8. So the energy obtained is just the sum of orbital energies. Then we consider the first order

energy correction with perturbation operator

$$\begin{aligned} E^{MP1} &= \langle \Psi^{HF} | \hat{V} | \Psi^{HF} \rangle \\ &= \langle \psi_i \psi_j | \Big(\frac{1}{2} \sum_{ij, i \neq j} (\hat{J}_{ij} - \hat{K}_{ij}) - \sum_{i,j} (\hat{J}_{ij} - \hat{K}_{ij}) \Big) | \psi_i \psi_j \rangle \\ &= -\frac{1}{2} \sum_{i,j} \langle \psi_i \psi_j | (\hat{J}_{ij} - \hat{K}_{ij}) | \psi_i \psi_j \rangle \\ &= -\frac{1}{2} \sum_{i,j} \langle \psi_i \psi_j || \psi_i \psi_j \rangle \end{aligned} \tag{1.21}$$

With the first order energy correction on the unperturbed energy, one will get the Hartree-Fock energy of the whole system

$$E^{HF} = E^{MP0} + E^{MP1} = \sum_i \epsilon_i - \frac{1}{2} \sum_{i,j} \langle \psi_i \psi_j || \psi_i \psi_j \rangle \tag{1.22}$$

where one should subtract from the total orbital energy with the double counting of Coulomb and exchange interactions when index i and j traversing the molecular orbitals independently. The first contribution to the correlation energy appears at second-order in the perturbation series

$$\begin{aligned} E^{MP2} &= -\sum_{i,j \to a,b} \frac{\left| \langle \Psi^{HF} | \hat{V} | \Psi^{HF}_{i,j \to a,b} \rangle \right|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \\ &= -\frac{1}{4} \sum_{i,j \to a,b} \frac{\left| \langle \psi_i \psi_j || \psi_a \psi_b \rangle \right|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \end{aligned} \tag{1.23}$$

where compared to the Equation 1.21, the numerator has electrons exciting from the ground states i and j to excited states a and b. We can see that the Coulomb integration involves four indices, leading to a growth rate of $O(N^4)$. While exchange operator $\hat{K} = \hat{J}\hat{P}$ involves permutation between orbitals, thus it scales as $O(N^5)$.

Configuration interaction (CI) provides a way to approach exact correlation. The

CI wavefunction can be expressed as a linear combination of the electronic configurations

$$|\Psi^{CI}\rangle = c_0|\Psi_0^{HF}\rangle + \sum_{ia} c_{i\to a}|\Psi_{i\to a}^{HF}\rangle + \sum_{i<j,a<b} c_{i,j\to a,b}|\Psi_{i,j\to a,b}^{HF}\rangle + \dots \tag{1.24}$$

where one can obtain different configurations by exciting electrons from occupied orbitals (i, j, ...) into virtual orbitals (a, b, ...), and the spatial correlations in the electron motions will be included by coupling different configurations. Generally, full CI calculations are cost prohibitive and can be performed only for a few small systems, while truncated CI suffers a severe problem of size-consistency.

As a reliable approach for benchmarking reference, coupled cluster (CC) theory allows for instantaneous Coulomb repulsion between electrons. Based on the exponential ansatz, the CC wavefunction can be expressed as

$$\begin{aligned} |\Psi^{CC}\rangle &= e^{\hat{T}}|\Psi^{HF}\rangle \\ &= \left(1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \dots\right)|\Psi^{HF}\rangle \end{aligned} \tag{1.25}$$

where the excitation operator $\hat{T}$ is expanded as different levels of electron excitation $\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots$. In practice, the wavefunction expansion is also truncated to the lower excitation levels,

$$\begin{aligned} |\Psi^{CCS}\rangle &= e^{(\hat{T}_1)}|\Psi^{HF}\rangle, \\ |\Psi^{CCSD}\rangle &= e^{(\hat{T}_1+\hat{T}_2)}|\Psi^{HF}\rangle, \\ |\Psi^{CCSDT}\rangle &= e^{(\hat{T}_1+\hat{T}_2+\hat{T}_3)}|\Psi^{HF}\rangle \end{aligned} \tag{1.26}$$

where for the wavefunction of coupled-cluster singles (CCS), all possible single excitations are included in $\hat{T}_1$ operator. Double excitations $\hat{T}_2$ are also included in the coupled-cluster singles and doubles (CCSD) wavefunction, and so on.

1.2.3 Basis set

Molecular orbitals can be expanded in basis function space, which is also called the linear combinations of atomic orbitals (LCAO-MO) method. Though the natural atomic orbitals (AOs) provide an efficient basis for expanding the molecular orbitals, their complicated functional forms are unfavorable for efficient numerical computations. Instead, Slater-type orbitals (STOs) are more often used as basis functions than the real atomic orbitals. STOs have the form

$$\phi^{STO}(r) = Nx^i y^j z^k e^{-\zeta|r-R|} \tag{1.27}$$

where N is the normalization constant, r is position of electron with respect to the atomic nucleus at R, ζ affects the spatial extent of the functions. A constraint $i + j + k = l$ is employed on the exponent, where l is the angular momentum. Though STOs have similar behavior as atomic orbitals, they are still complicated for numerical computations, especially for correlation calculation requiring many-orbital integrations.

A further approximation to atomic orbital is the use of Gaussian-type orbitals (GTOs), which has the form

$$\phi^{GTO}(r) = Nx^i y^j z^k e^{-\zeta|r-R|^2} \tag{1.28}$$

where the main difference compared to STOs is in the exponent as r^2. GTOs are more favorable for effective numerical computations based on Gaussian product theorem that the product of two GTOs on atom A (at R_A) and atom B (at R_B) can be represented with one GTO centered on $R_P = (\alpha R_A + \beta R_B)/(\alpha + \beta)$ as

$$e^{-\alpha|r-R_A|^2} e^{-\beta|r-R_B|^2} = \left(e^{-\alpha\beta/(\alpha+\beta)|R_A-R_B|^2}\right) e^{-(\alpha+\beta)|r-R_P|^2} \tag{1.29}$$

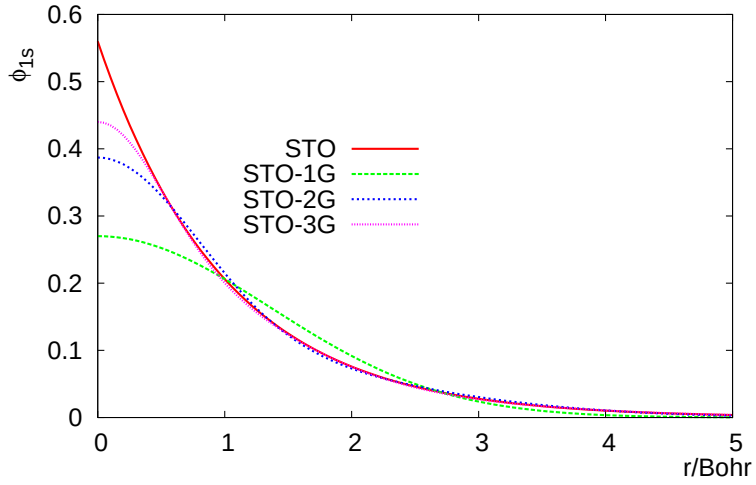

Figure 1.2: Comparison of STO, STO-1G, STO-2G and STO-3G for 1s function. Parameters are taken from Ref. [20].

In practice, several GTOs are linearly combined to form contracted GTOs (or CGTOs for short) to approximate STOs. Such basis sets are often called STO-nG (n is the number of GTOs) as Figure 1.2 shows. If more than one CGTO is used for each atomic orbitals, the basis is referred to as an X-tuple zeta basis. For example, if two basis functions are used per CGTO, it is called a double zeta (DZ), triple zeta (TZ) means there are three basis functions used for one CGTO, and so on. It is widely known that valence electrons are more active than core electrons for chemical properties, like forming or breaking the bond. Generally, fewer CGTOs are used to form core AOs while more CGTOs are used for the valence orbitals, which is called split-valence approach.

In order to describe shifts in the electron density relative to the nuclei (as in polar covalent bonds), one should increase the flexibility of basis functions by including higher order angular momentum functions (they are called "polarization functions"). To describe

delocalized orbital/interactions (like conjugate orbitals), "diffuse functions" with smaller ζ value should be used. Diffuse basis functions exhibit broad spatial extent and saturate the volume surrounding the molecule, which is very important for accurate noncovalent interaction calculations. In the following discussions, Dunning type basis sets [22] are used, which is often written as aug-cc-pVXZ (or aXZ for short). The "aug" prefix indicates that diffuse functions are included, "cc" is the short for correlation consistent, "p" indicates the presence of polarization functions, and "VXZ" describes the number of CGTO basis functions used for each valence AO orbital.

1.3 Methods for noncovalent interactions

A brief introduction to electronic structure methods for noncovalent interactions is necessary before proceeding to the detailed discussions. There are two main ways of modeling intermolecular interactions: symmetry adapted perturbation theory (SAPT) and the supermolecular approach. SAPT [23] is a perturbation approach that treats each isolated molecule as a zeroth order wavefunction, and improves the inter- and intra-molecular correlations with perturbation theory. It provides a physical description of different energy components in noncovalent interactions. To make concept easily understood and comparable to MP2 method, the discussion here is limited to methods which are second-order in perturbation theory [24]. The molecular properties computed from SAPT can be multipole expanded to build an *ab initio* force field in HMBI model.

The supermolecular approach, on the other hand, computes the intermolecular interaction energy as the difference between the supermolecule (e.g. a dimer) and the

constituent monomers. The simplicity of the supermolecular approach makes it adaptable to almost all the existing quantum mechanical methods [1, 2, 25], thus it is popular for practical applications. In order to achieve high accuracy, basis set completeness must be considered as well. Finally, some benchmark test sets will be introduced which can help us to evaluate the performance of different approaches and interpret the physics behind the observations. The next several sections discuss these techniques in general.

1.3.1 Symmetry adapted perturbation theory

Start with polarization series for wavefunctions and energies based on polarization theory (inter-molecular Rayleigh-Schrödinger perturbation theory without exchange effect included) [23]. For dimer AB composed with monomer A and B, the unperturbed Schrödinger equation is

$$\hat{H}_X^{(0)}\Psi_X^{(0)} = E_X^{(0)}\Psi_X^{(0)} \tag{1.30}$$

where X represents either each monomer, or the dimer AB. 0 means there is no perturbation correction (unperturbed). If X stands for dimer AB, the Hamiltonian without including interactions between monomer A and B is $\hat{H}_{AB}^{(0)} = \hat{H}_A^{(0)} + \hat{H}_B^{(0)}$, and the wavefunction without intermolecular exchange effect is just Hartree product $\Psi_{AB}^{(0)} = \Psi_A^{(0)}\Psi_B^{(0)}$. Accordingly, the total energy solved is just the sum of monomer energies $E_{AB}^{(0)} = E_A^{(0)} + E_B^{(0)}$. The molecular orbitals and orbital energies can be computed from HF method or DFT method.

Next, the intermolecular perturbation operator $\hat{V}_{AB}^{pol}$ which does not consider electron exchange between monomers, is added to describe intermolecular polarization interac-

tions. The polarization Schrödinger equation is expressed as

$$\hat{H}^{pol}\Psi = (\hat{H}^{(0)}_{AB} + \hat{V}^{pol}_{AB})\Psi = E^{pol}\Psi \tag{1.31}$$

where Ψ and E^{pol} are the polarization wavefunction and energy, respectively.

Based on R-S perturbation theory, the first-order perturbed polarization correction is

$$E^{(1)}_{pol} = \langle\Psi^{(0)}|\hat{V}^{pol}_{AB}|\Psi^{(0)}\rangle = \langle\Psi^{(0)}_{A}\Psi^{(0)}_{B}|\hat{V}^{pol}_{AB}|\Psi^{(0)}_{A}\Psi^{(0)}_{B}\rangle \tag{1.32}$$

where the physical interpretation is clear: Coulomb interaction between electrons in monomer A and electrons in monomer B. So this term represents the electrostatic interaction $E^{(1)}_{elst}$.

The first order wavefunction correction is expressed as

$$|\Psi^{(1)}\rangle = -\sum_{k\neq 0}\frac{\langle\Psi^{(0)}_{k}|\hat{V}^{pol}_{AB}|\Psi^{(0)}\rangle}{E^{(0)}_{k} - E^{(0)}}|\Psi^{(0)}_{k}\rangle \tag{1.33}$$

where k is the exciting state computed from SCF procedure. With wavefunction correction, the second order polarization correction can be expressed as

$$E^{(2)}_{pol} = \langle\Psi^{(0)}|\hat{V}^{pol}_{AB}|\Psi^{(1)}\rangle \tag{1.34}$$

If one considers the electron excitation of monomer A and monomer B separately, the first order wavefunction $\Psi^{(1)}$ correction represents the induction effect

$$\begin{aligned}|\Psi^{(1)}_{ind}\rangle = &-\sum_{a\neq 0}\frac{\langle\psi^{A}_{a}\psi^{B}_{0}|\hat{V}^{pol}_{AB}|\psi^{A}_{0}\psi^{B}_{0}\rangle}{E^{A}_{a} - E^{A}_{0}}|\psi^{A}_{a}\psi^{B}_{0}\rangle \\ &-\sum_{b\neq 0}\frac{\langle\psi^{A}_{0}\psi^{B}_{b}|\hat{V}^{pol}_{AB}|\psi^{A}_{0}\psi^{B}_{0}\rangle}{E^{B}_{b} - E^{B}_{0}}|\psi^{A}_{0}\psi^{B}_{b}\rangle\end{aligned} \tag{1.35}$$

that either monomer A or monomer B will be polarized. Replacing the wavefunction in

Equation 1.34, one will get the corresponding energy correction

$$E_{ind}^{(2)} = -\sum_{a\neq 0} \frac{|\langle \psi_0^A \psi_0^B | \hat{V}_{AB}^{pol} | \psi_a^A \psi_0^B \rangle|^2}{E_a^A - E_0^A} - \sum_{b\neq 0} \frac{|\langle \psi_0^A \psi_0^B | \hat{V}_{AB}^{pol} | \psi_0^A \psi_b^B \rangle|^2}{E_b^B - E_0^B} \tag{1.36}$$

which can be interpreted as the induction on monomer A by polarizing monomer B as $E_{ind}^{(2)}(A \leftarrow B)$ and the induction on monomer B due to the polarization of monomer A as $E_{ind}^{(2)}(A \rightarrow B)$.

If one considers the electron excitations of monomer A and monomer B simultaneously, the wavefunction correction $\Psi^{(1)}$ represents the dispersion effect

$$|\Psi_{disp}^{(1)}\rangle = -\sum_{a,b\neq 0} \frac{\langle \psi_a^A \psi_b^B | \hat{V}_{AB}^{pol} | \psi_0^A \psi_0^B \rangle}{E_a^A + E_b^B - E_0^A - E_0^B} |\psi_a^A \psi_b^B\rangle \tag{1.37}$$

where both monomer A and monomer B are polarized at the same time. The interaction is called dispersion interaction interpreted as the polarization of one monomer due to the polarization of the other. Similar to induction interactions, replacing the wavefunction correction in Equation 1.34 with the above expression gives the corresponding dispersion energy

$$E_{disp}^{(2)} = -\sum_{a,b\neq 0} \frac{|\langle \psi_0^A \psi_0^B | \hat{V}_{AB}^{pol} | \psi_a^A \psi_b^B \rangle|^2}{E_a^A + E_b^B - E_0^A - E_0^B} \tag{1.38}$$

More detailed discussion of dispersion interaction is the main topic in the following chapters.

As mentioned earlier, the intermolecular perturbation operator $\hat{V}_{AB}^{pol}$ does not include intermolecular exchange. To get the correct interaction energy, one should take account of exchange components corresponding to each polarization term [26]. For first order polarization, the exchange term corresponding to electrostatic interaction $E_{elst}^{(1)}$ is

$$E_{exch}^{(1)} = \frac{\langle \psi_0^A \psi_0^B | \hat{V}_{AB}^{pol} - E_{pol}^{(1)} \hat{P} | \psi_0^A \psi_0^B \rangle}{1 + \langle \psi_0^A \psi_0^B | \hat{P} | \psi_0^A \psi_0^B \rangle} \tag{1.39}$$

where $\hat{P}$ is intermolecular permutation operator which exchanges electrons between monomer A and monomer B. In practice, only single pair of electrons are exchanged at one time. For second order perturbation, the exchange terms corresponding to $E^{(2)}_{ind}$ and $E^{(2)}_{disp}$ are exchange-induction interaction ($E^{(2)}_{ex\text{-}ind}$) and exchange-dispersion interaction ($E^{(2)}_{ex\text{-}disp}$), respectively

$$E^{(2)}_{ex\text{-}ind} = -\langle\Psi^{(0)}|(\hat{V}^{pol}_{AB} - E^{(1)}_{pol})(\hat{P} - \bar{P})|\Psi^{(1)}_{ind}\rangle \tag{1.40}$$

and

$$E^{(2)}_{ex\text{-}disp} = -\langle\Psi^{(0)}|(\hat{V}^{pol}_{AB} - E^{(1)}_{pol})(\hat{P} - \bar{P})|\Psi^{(1)}_{disp}\rangle \tag{1.41}$$

where $\bar{P}$ is the expectation value as $\langle\Psi^{(0)}|\hat{P}|\Psi^{(0)}\rangle$, with the same first order corrected wavefunctions $|\Psi^{(1)}_{ind}\rangle$ and $|\Psi^{(1)}_{disp}\rangle$ have been given before. The total SAPT interaction energy (up to second order) is the combination of polarization series and their corresponding exchange contributions

$$\begin{aligned} E^{(2)}_{SAPT} &= E^{(2)}_{pol} + E^{(2)}_{exchange} \\ &= \left(E^{(1)}_{elst} + E^{(2)}_{ind} + E^{(2)}_{disp}\right) + \left(E^{(1)}_{exch} + E^{(2)}_{ex\text{-}ind} + E^{(2)}_{ex\text{-}disp}\right) \end{aligned} \tag{1.42}$$

To visualize the contributions from different energy components, the potential energy surface (PES) of stacked benzene dimer [27] is provided in Figure 1.3 as an example. We can observe that all the energy components are zero when two molecules are well separated and their contributions increase as the molecules are approaching to each other. Generally, energies of the polarization series are more attractive with negative sign. While the corresponding exchange terms are repulsive. The sum of all these terms gives the total interaction energy. Staring from infinite separation, the molecules attract each other as they come closer. After reaching an equilibrium distance which is the energy minimum, the

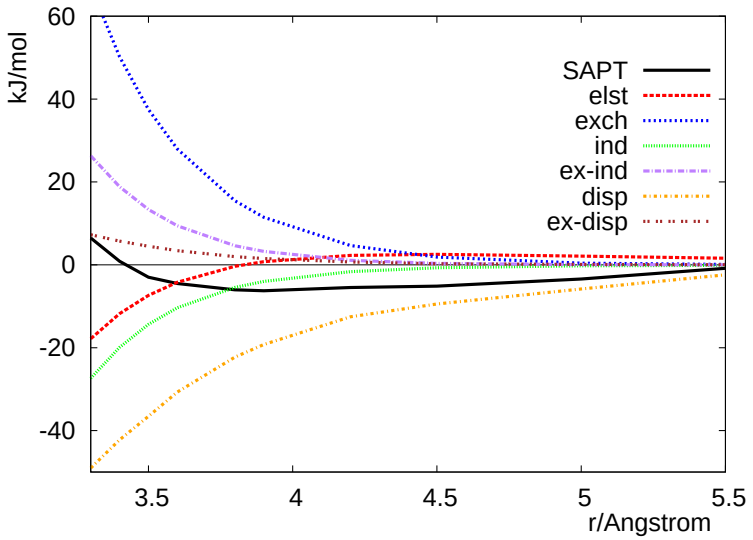

Figure 1.3: Potential energy components computed from symmetry adapted perturbation theory for stacked benzene dimer. The energies change as the intermolecular distance. Structures are taken from [27].

interaction energy becomes less favorable and eventually turns repulsive as the molecules continue approaching. This repulsive effect stems from the fast increase of exchange contributions, which can be interpreted as the strong Coulomb repulsion between electron densities (or orbital overlapping) from the two molecules.

1.3.2 Polarizable multipole *ab initio* force field

As a simple approximation to SAPT, the *ab initio* force field [16] just includes the cheaper polarization series computed from distributed polarizable multipoles. The exchange effect is approximated by using damping functions which decay from 1 to 0 to turn off the attractive terms as the atoms approach. The electrostatic interaction energy between atom

A and B is given by

$$E_{elst} \leftarrow \sum_{tu} Q_t^A T_{tu} Q_u^B \tag{1.43}$$

where the T_{tu} matrix includes the distance- and orientation-dependent contributions for the interaction of two different spherical-tensor multipole moment components Q_t and Q_u. To evaluate the induction contribution, one first finds the induced multipole moments according to

$$\Delta Q_t^A = -\sum_{t'u} \alpha_{tt'}^A T_{t'u} \left(Q_u^B + \Delta Q_u^B \right) \tag{1.44}$$

where $\alpha_{tt'}^A$ is the static polarizability tensor on atom A and ΔQ is an induced multipole moment. Clearly, the induced multipole moment on atom A depends on the induced multipole moment on atom B, so this process is done self-consistently until the induced multipoles converge. So the induction energy contribution between atoms A and B is

$$E_{ind} \leftarrow \sum_{tu} \Delta Q_t^A T_{tu} Q_u^B + Q_t^A T_{tu} \Delta Q_u^B \tag{1.45}$$

The force field dispersion is evaluated via Casimir-Polder integration over frequency-dependent polarizabilities (FDPs):

$$E_{disp} \leftarrow \sum_{tu} \sum_{t'u'} T_{tu} T_{t'u'} \int_0^\infty \alpha_{tt'}^A(\omega) \alpha_{uu'}^B(\omega) d\omega \tag{1.46}$$

The integration is evaluated via numerical quadrature over imaginary frequency ω. The above expression corresponds to an anisotropic model for atom-atom dispersion. However, to a fairly good approximation, one can approximate this with a simple isotropic dispersion model (i.e., the one only averages over the diagonal dipole-dipole and quadrupole-quadrupole elements of the frequency-dependent polarizability). In that case, the dispersion

model reduces to the standard C_6, C_8, etc., terms divided by the interatomic distance R to the corresponding power,

$$E_{disp} \leftarrow \frac{C_6^{AB}}{R_{AB}^6} + \frac{C_8^{AB}}{R_{AB}^8} + \cdots \tag{1.47}$$

The isotropic dispersion coefficients C_n are obtained from the Casimir-Polder integration over the appropriate elements of the isotropic frequency-dependent polarizabilities $\bar{\alpha}$:

$$C_n^{AB} \leftarrow \int_0^\infty \bar{\alpha}^A(\omega)\bar{\alpha}^B(\omega)d\omega \tag{1.48}$$

1.3.3 Supermolecular approach

The supermolecular approach describes molecular interactions by calculating the energy difference between the whole system and the constituent units. For example, the interaction energy in a two-body system is given by

$$E_{int}^{AB} = E^{AB} - E^A - E^B \tag{1.49}$$

where AB represents the dimer, and A, B stand for each monomer. This approach is widely used for practical calculations, because almost all of the existing quantum chemical methods can be applied on this model. Unlike SAPT, one does not need to take extra efforts to describe the complicated exchange terms explicitly, since the exchange effect has been naturally included in the supermolecular orbitals.

A lot of quantum mechanical methods have been developed in order to predict noncovalent interactions accurately and effectively [25, 28]. Some mainstream methods frequently used today is listed below, and their performance on benchmark test set will be provided in the following section.

- **CCSD(T)**: It is very accurate and is often considered as the "gold standard" for benchmark calculations. However, its high computational cost that grows $O(N^7)$ with system size N makes it impractical for many applications, especially for large systems.

- **MP2** [29]: It is accurate for polarization interactions but it often overestimates van der Waals dispersion interactions. It exhibits more affordable $O(N^5)$ computational cost.

- **SCS-MP2**: Spin-component scaled MP2 with empirical spin scaling factors for same-spin (ss) energy component and opposite-spin (os) energy component. Some optimal parameters are SCS-MP2(ss=0.33,os=1.20) [30], SOS-MP2(ss=0.00,os=1.30) [31] and SCS(MI)-MP2(ss=1.75,os=0.17) [32]. These methods are much improved than the conventional MP2, especially for dispersion interactions. Besides SOS-MP2 which scales as $O(N^4)$ by ignoring the more expensive same-spin term (with Coulomb exchange scaling as $O(N^5)$), the other methods have the same size scaling as conventional MP2.

- **MP2.5**: MP2.5 [33] is an average of the MP2 and MP3 interaction energies. It achieves a high accuracy by cancelling the overestimation error in MP2 and underestimation error in MP3. The size scaling rate is the same as MP3 with $O(N^6)$.

- **MP2C**: After recognizing the main error source in MP2 comes from the uncoupled Hartree-Fock (UCHF) dispersion, "coupled" MP2 (MP2C) method [34, 35] replaces the UCHF dispersion found in MP2 with more accurate dispersion based on cou-

pled Kohn-Sham (CKS) theory, which substantially reduces the correlation errors. This correction does not significantly change the computational cost relative to the conventional MP2.

- **DFT-D**: In conventional Kohn-Sham DFT, the approximate density functionals depend only on the local electron density, making them unable to describe long-range correlation interactions like dispersion. DFT-D corrects this deficiency by including the dispersion terms computed with multipole expansion model [36–39]. If one uses empirical dispersion coefficients, the dispersion calculation is almost free and the whole calculation has the same scaling rate as DFT with $O(N^3)$.

1.3.4 Finite basis set error

Theoretically, the basis space to expand molecular orbitals is infinite. In practice, the angular momentum is truncated and a finite number of functions are used to approximate molecular orbitals. Though finite-basis set errors cannot be avoided in practical applications, several approaches have been developed to reduce such errors.

First, for interaction energies computed from supermolecular approach, one should take care of the basis set superposition error (BSSE) which stems from the use of finite basis set. Due to the basis set incompleteness, when building molecular orbitals for the supermolecule, electrons in one monomer often "borrow" basis functions from the neighboring molecules to compensate for basis set incompleteness and to achieve a lower energy. On the other hand, these basis functions are not present for the isolated monomer calculations, so no such basis set borrowing/energy lowering can occur. This leads to an artificial at-

traction between the molecules, and this BSSE does not cancel out readily. Generally, the counterpoise (CP) correction is used to reduce BSSE. The BSSE can be approximated by comparing the difference in the monomer energies with and without the presence of the basis functions from the neighboring molecule(s):

$$E^A_{BSSE} = E^A(AB) - E^A(A) \tag{1.50}$$

and

$$E^B_{BSSE} = E^B(AB) - E^B(B) \tag{1.51}$$

where the symbols in the bracket tells where the basis functions come from. For example, $E^A(AB)$ uses basis functions from both monomer A and monomer B to build orbitals of monomer A. While $E^A(A)$ only includes basis functions from monomer A. Subtracting BSSEs from Equation 1.49, one will obtain the interaction energy with BSSE significantly reduced

$$E^{AB}_{int}(CP) = E^{AB}(AB) - E^A(AB) - E^B(AB) \tag{1.52}$$

In the following chapters, we will use counterpoise correction as default unless otherwise specified as "no-CP".

Second, even after correcting for BSSE, finite basis set error still exists. As shown in Figure 1.4, this error is significant relative to the magnitude of the interaction energies. To reduce the finite basis set error, basis set extrapolation will be utilized to approximate the complete basis set (CBS) limit. Since SCF and correlation energies have different basis set convergence behaviors, they will be extrapolated separately. In the following chapters, two point aTZ-aQZ extrapolation [40, 41] is used to estimate energies in the CBS limit

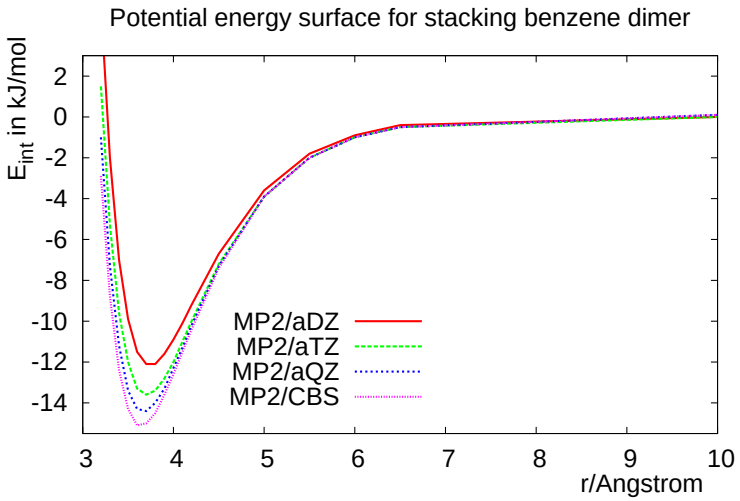

Figure 1.4: Potential energy curves of stacked benzene dimer with MP2/aXZ. CBS limit is estimated with aTZ-aQZ extrapolation. Structures are taken from [27].

$$E_{HF}^{\infty} = E_{HF}^{Q} + \frac{E_{HF}^{Q} - E_{HF}^{T}}{e^{1.54(4-3)} - 1} \tag{1.53}$$

and

$$E_{Corr}^{\infty} = \frac{4^3 E_{Corr}^{Q} - 3^3 E_{Corr}^{T}}{4^3 - 3^3} \tag{1.54}$$

where E^{∞} means energy estimated in CBS limit. E^{T} and E^{Q} represent energies computed with aug-cc-pVTZ and aug-cc-pVQZ basis set, respectively.

1.3.5 Benchmark sets for methods evaluations

An excellent theoretical method for noncovalent interactions should be both efficient and accurate enough for various type of interactions. The benchmark dimer test sets like S22 [42, 43], S22x5 [44] (scaling S22 set into four non-equilibrium separations), S66 [45], S66a8 [46] (rotating relative angles from S66 set into eight non-equilibrium ge-

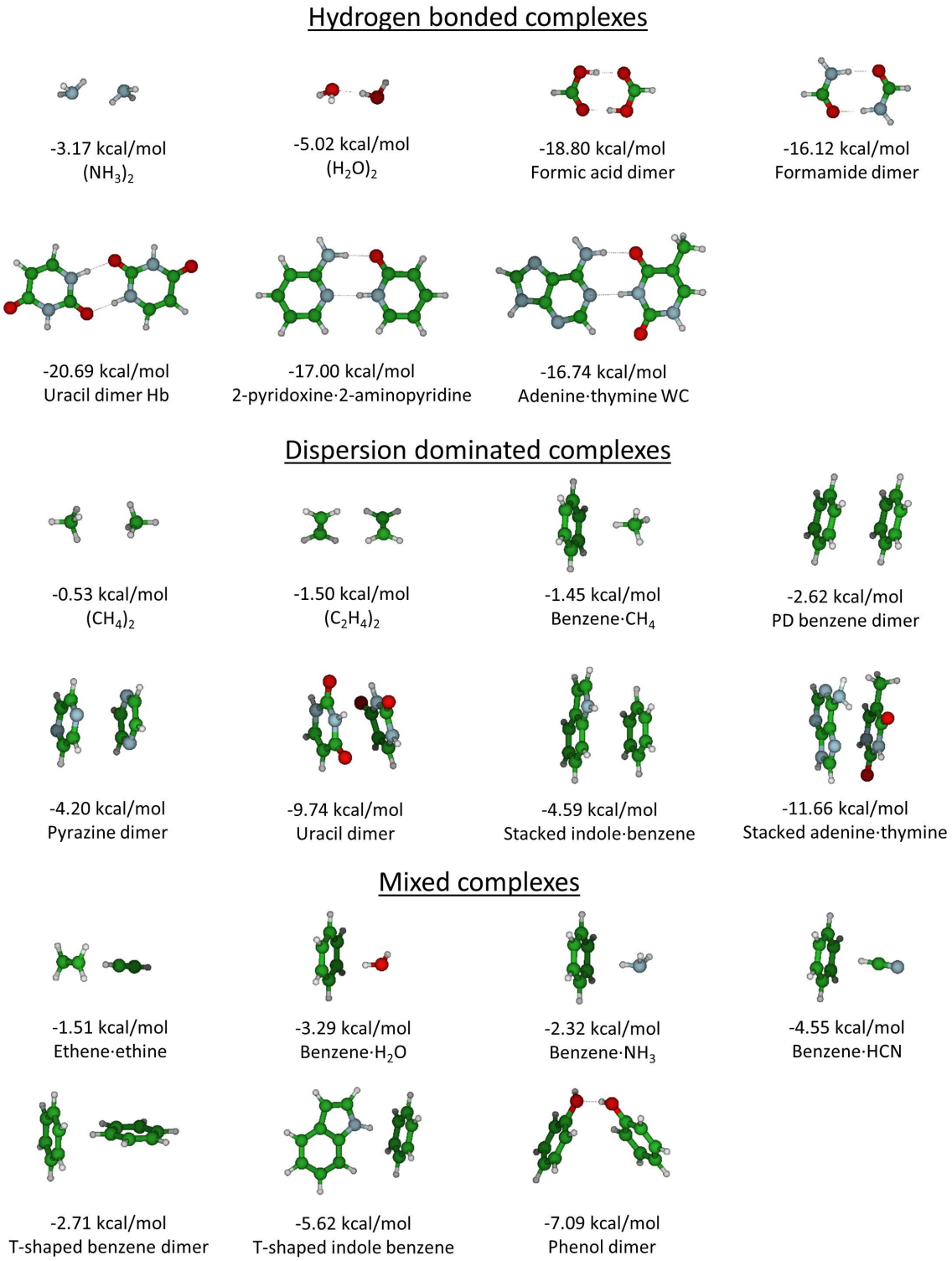

Figure 1.5: S22 dimer set with interaction energies computed from CCSD(T)/CBS [43].

ometries), L7 [47] and S12L [48, 49] provide representative systems of various interactions, which can help to evaluate the performance of different theoretical models. Meanwhile, the reference results (usually computed from CCSD(T)/CBS) are provided for comparison and error analysis. Here we just introduce the S22 test set, but some other benchmark sets will also be used in the following chapters. The S22 test set is divided into three groups based on different types of interactions as Figure 1.5 shows. The first group is hydrogen bond dominant species, the second one is dispersion dominant and the last is mixed with electrostatic and dispersion with similar magnitude.

To give a flavor on how the methods discussed previously perform on the S22 test set, root-mean-square (RMS) errors with respect to CCSD(T) results are provided in Table 1.1 for comparison [1]. From the table, one can draw the conclusions that MP2 performs well on polarization interactions, but it has large error in predicting dispersion interactions. SCS-MP2 decreases the overestimation error in dispersion by scaling the correlation components with a smaller prefactor. The empirical parameters introduced for correlation, however, disrupt the balance for describing other interactions. As we can see from the table, the RMS error of the hydrogen bonded species increases to 0.54 kcal/mol.

Table 1.1: RMS error (kcal/mol) of different methods on S22 test set [1].

methods	S22	H-bonded	Dispersion-bonded	mixed
MP2	0.94	0.27	1.24	0.37
SCS-MP2	0.58	0.54	0.60	0.17
SCS(MI)-MP2	0.26	0.31	0.28	0.22
MP2.5	0.22	0.07	0.32	0.14
MP2C	0.18	0.19	0.10	0.13
SAPT(DFT)	0.49	0.47	0.45	0.30
B3LYP-D	0.82	0.35	0.48	0.67

MP2.5 depending on error cancellations reduces the RMS errors significantly. MP2C only fixes the correlation problem in dispersion, so it is still accurate for polarization interactions. SAPT(DFT) is affected by a lot of factors: the SCF model selected to compute MOs, the type of exchange-correlation kernels (ALDAX, ADAX+CHF, and ALDA, for instance), the level of perturbation correction to improve inter- and intra-molecular correlations, the size basis set used, etc. Table 1.1 shows that this method performs uniformly well for different interactions. Corrected with empirical dispersion, DFT-D method performs well for dispersion interactions, but there is still no improvement on induction interactions.

1.4 Outline of the Dissertation

In the next chapter, we will consider the computational efficiency of the MP2C method which approaches near CCSD(T) accuracy with an affordable computational cost. With fortuitous basis set error cancellations, it is possible to approach dimer-centered (DC) MP2C dispersion correction using a smaller monomer-centered (MC) one. Applying the MC algorithm to an individual dimer provides several-fold speed-ups. More significantly, in the context of fragment-based molecular crystal studies, combining the new algorithm and translational symmetry of these periodic systems reduces the computational cost of dispersion corrections by two orders of magnitude, making the dispersion correction calculation almost free. Most contents of this chapter are reprinted from "Accelerating MP2C dispersion corrections for dimers and molecular crystals", *J. Chem. Phys.*, **138**, 224112 (2013).

In Chapter 3, we focus on the MP2 part of the MP2C method, which is the main

computational bottleneck. Here, we demonstrate that one can avoid calculating the long-range MP2 correlation by attenuating the Coulomb operator, allowing dispersion correction to handle the long-range interactions inexpensively. With relatively modest Coulomb attenuation, one obtains the results close enough to those computed from the conventional MP2C method. In this case, the finite basis set error in the description of dispersion is the main error source. One can reduce the finite-basis error by extrapolating dispersion energy to the CBS limit. Alternatively, with more aggressive attenuation, one can remove just enough short-range repulsive exchange-dispersion to compensate for the finite basis set error. Accordingly, it is possible to approach complete basis set limit quality results with only a small basis set, resulting in substantial computational savings. The existing explicitly correlated MP2C (MP2C-F12) method is also discussed, which also approaches CBS limit accuracy with small basis set. Reliabilities of these methods are examined on different benchmark dimer sets and molecular crystals. Most contents of this chapter are reprinted with the permission from "Achieving High-Accuracy Intermolecular Interactions by Combining Coulomb-Attenuated Second-Order Møller-Plesset Perturbation Theory with Coupled Kohn-Sham Dispersion", *J. Chem. Theory Comput.*, **10**, 2054-2063 (2014).

Finally, in Chapter 4, we consider the three-body interactions, which are nonnegligible contributions to the total lattice energy. However, most low scaling methods exhibit problems in describing three-body interactions. Considering the correlation correction on conventional three-body MP2 method, we utilize the short-range damped Axilrod-Teller-Muto (ATM) dispersion. The damping compensates for the absence of higher-order dispersion contribution (MP4 and above) and non-additive short-range exchange terms not found

in MP2. This model proves to be reliable from the examinations on benchmarking trimer sets and molecular crystals. This work is a reprint of "Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order Møller-Plesset perturbation theory", submitted.

To avoid confusion when making energy comparison, one should notice the different energy units used in this thesis: Chapter 2 and Chapter 3 use kJ/mol (1 Hartree = 2625.5002 kJ/mol), while Chapter 4 uses kcal/mol (1 kcal/mol = 4.184 kJ/mol).

Chapter 2

Improving computational efficiency of MP2C dispersion correction for two-body interactions

2.1 Introduction

Unlike most commonly used DFT functionals, MP2 naturally captures dispersion interactions at affordable computational cost. However, MP2 suffers the well-known deficiency that it frequently overestimates dispersion energies. Furthermore, because the strength of dispersion interactions depends strongly on the orientations and distances of the molecules, errors in the MP2 interaction energies are often uneven across the potential energy surface [44, 45].

Instead of employing large-basis CCSD(T) calculations, which are cost-prohibitive

for many chemically interesting systems, practical calculations are typically limited to methods which scale $O(N^5)$ or less with system size N. Inexpensive MP2 corrections that improve the description of dispersion interactions are therefore particularly important.

Several such MP2 dispersion correction models exist. Spin-component scaling of the MP2 correlation energy contributions can significantly improve MP2 interaction energies [30–32], but such simple empirical approaches sometimes perform poorly, as in oxalyl dihydrazide crystals [50], for instance. Dispersion-weighted MP2 adjusts the scaling depending on the relative sizes of the HF and MP2 interaction energies [51]. Other empirically corrected methods such as MP2.5 [33], spin-component-scaled CCSD [52,53], or dispersion-weighted CCSD methods [54] provide more reliable intermolecular interaction energies, but they scale with $O(N^6)$, which is too expensive for practical use.

Alternatively, since the main error source in MP2 comes from the uncoupled Hartree-Fock (UCHF) treatment of dispersion [55,56], several methods have been proposed to replace the UCHF dispersion energy with a more accurate one computed using coupled Hartree-Fock (CHF) [57], coupled Kohn-Sham (CKS) [34, 35], or high-quality dispersion coefficients obtained from other sources [58]. Of these later approaches, the MP2C method of Pitonak and Hesselmann [34, 35], which corrects MP2 based on the difference between the CKS and UCHF dispersion energies, has proved to be very successful. MP2C is non-empirical (aside from any empiricism present in the density functional/kernels used in the CKS calculation) and agrees well with high-level benchmark calculations for a wide variety of intermolecular interactions [59–64].

Although the MP2C dispersion correction formally scales only $O(N^4)$ versus $O(N^5)$

for MP2, evaluating the dispersion correction still comprises a sizable fraction of the overall computational time in practical calculations. For instance, the dispersion calculation needs to be integrated over several imaginary frequencies (usually 10 grid points for numerical integration), which increases the prefactor to the $O(N^4)$ scaling.

In this chapter, the cost of evaluating the MP2C dispersion correction is substantially reduced by exploiting two features. First, for individual dimer calculation, density-density response function (used for dispersion calculation) are computed in a monomer-centered (MC) basis instead of the original dimer-centered (DC) one, which reduces the computational cost several-fold. Although the accurate computation of the dispersion energy requires a very large basis set [65], the energy difference between the CKS and UCHF dispersion energies proved to be much less sensitive to the basis set, and the MP2C correction obtained in MC basis is nearly identical to the one obtained in DC basis set. Second, for fragment-based models to treat molecular crystals, additional larger savings are achieved by recognizing that, in an MC basis, the UCHF and CKS response functions are molecular properties that are transferable to all symmetry-equivalent molecules in such periodic systems. One simply computes the CKS/UCHF response functions with MC basis set for the handful monomers in the asymmetric central unit cell, then the intermolecular dispersion correction can be evaluated for each of the dozens or more dimer interactions in the crystal at trivial cost. Overall, this approach reduces the cost of MP2C dispersion correction in a molecular crystal by two orders of magnitude, making it virtually cost-free compared to the underlying MP2 calculations, without sacrificing significant accuracy.

2.2 Theory

2.2.1 MP2C dispersion correction

Start with the UCHF dispersion in MP2, which has the form as

$$E_{disp}^{UCHF} = -\sum_{a,b\neq 0} \frac{|\langle \psi_0^A \psi_0^B | r_{AB}^{-1} | \psi_a^A \psi_b^B \rangle|^2}{E_a^A + E_b^B - E_0^A - E_0^B} \tag{2.1}$$

where molecular orbitals and orbital energies are computed from Hartree-Fock theory. After Casimir-Polder transformation, the dispersion expression can be represented with integration over imaginary frequencies ω

$$E_{disp}^{UCHF} = -\frac{1}{2\pi}\int_0^\infty d\omega \Big(\int\int\int\int \alpha^A(r_A, r'_A|\omega)\alpha^B(r_B, r'_B|\omega) \frac{dr_A dr_B}{|r_A - r_B|} \frac{dr'_A dr'_B}{|r'_A - r'_B|} \Big) \tag{2.2}$$

where α^A and α^B represents molecular frequency-dependent density susceptibilities (FDDSs) [66, 67] for monomer A and monomer B respectively. Then the uncoupled (UC) FDDS for one molecule has the form

$$\begin{aligned} \alpha_0^X(r, r'|\omega) &= 4\sum_{ia} \frac{(\epsilon_a - \epsilon_i)}{(\epsilon_a - \epsilon_i)^2 - \omega^2} \psi_i(r)\psi_a(r)\psi_i(r')\psi_a(r') \\ &= \sum_{ia} C_0^{ia}(\omega)\psi_i(r)\psi_a(r)\psi_i(r')\psi_a(r') \end{aligned} \tag{2.3}$$

where i, j and a, b denote occupied and virtual orbitals, respectively. ψ and ϵ represent molecular orbital and orbital energy of monomer X (either A or B). $C_0^{ia}(\omega)$ is the expansion coefficient of uncoupled FDDS involving exciting states and imaginary frequencies. To improve the computational efficiency, density fitting or resolution of the identity technique is used by introducing auxiliary functions (represented with P and Q) to represent products of molecular orbitals $|ia\rangle$,

$$\psi_i(r)\psi_a(r) = |ia\rangle = D_{ia,P}|P\rangle \tag{2.4}$$

where $|P\rangle$ is an auxiliary basis function, and $D_{ia,P}$ is the density fitting coefficient indicating the weight of auxiliary function $|P\rangle$ in the expression of $|ia\rangle$. Multiplying the above equation by $S^{-1}\langle Q|$ on both sides, one obtains the density fitting coefficient

$$D_{ia,P} = \sum_{Q} (S^{-1})_{PQ} T_{ia,Q} \tag{2.5}$$

in which, S_{PQ} is two-index Coulomb integral $\langle P|r_{12}^{-1}|Q\rangle$, and $T_{ia,Q}$ is three-index Coulomb integral $\langle ia|r_{12}^{-1}|Q\rangle$ coupling between AO basis functions and auxiliary functions. Applying the density fitting algorithm, the new UCHF dispersion expression can be rewritten as

$$\begin{aligned} \alpha_0^X(r, r'|\omega) &= \sum_{PQ} \Big(\sum_{ia} D_{ia,P} C_0^{ia}(\omega) D_{ia,Q} \Big) |P\rangle |Q\rangle \\ &= \sum_{PQ} \chi_0^X(\omega) |P\rangle |Q\rangle \end{aligned} \tag{2.6}$$

where $\chi_0^X(\omega)$ is called the uncoupled propagator (or response function) for monomer X, and the dispersion energy with density fitting can be expressed by coupling the propagators with intermolecular Coulomb operator

$$E_{disp}^{UCHF} = -\frac{1}{2\pi} \int_0^\infty Tr\Big[\chi_0^A(\omega) J^{AB} (\chi_0^B(\omega))^T (J^{AB})^T\Big] d\omega \tag{2.7}$$

where the Coulomb integral $J^{AB} = \langle P|r_{12}^{-1}|Q\rangle$ couples $\chi_0^A(\omega)$ and $\chi_0^B(\omega)$. The trace of the matrix resulting from the matrix multiplies is evaluated at each frequency grid point used in the numerical integration by quadrature.

To obtain the CKS dispersion, one first calculates the uncoupled Kohn-Sham propagators with orbitals computed from local Hartree-Fock (LHF). Then for each frequency ω, the coupled Kohn-Sham propagator is obtained by solving Dyson typed equation iteratively

$$\chi_{coup}(\omega) = \chi_0(\omega) + \chi_0(\omega) W \big[S - S\chi_0(\omega) W\big]^{-1} S\chi_0(\omega) \tag{2.8}$$

where

$$W_{PQ} = S_{PQ} + \langle P|f_{xc}|Q\rangle \tag{2.9}$$

and $\langle P|f_{xc}|Q\rangle$ [66–68, 35] are integrals of the exchange-correlation kernel in the auxiliary basis set. In MP2C, the adiabatic LDA exchange-only (ALDAX) kernel is used

$$f_{xc} = -C_x\rho^{-2/3} \tag{2.10}$$

where C_x is Slater-Dirac exchange for LDA with the value $-\frac{4}{3}(\frac{3}{\pi})^{1/3}$, and ρ is for electron density. With the CKS propagator computed, the CKS dispersion energy can be obtained as

$$E_{disp}^{CKS} = -\frac{1}{2\pi}\int_0^\infty Tr\left[\chi_{coup}^A(\omega)J^{AB}\left(\chi_{coup}^B(\omega)\right)^T\left(J^{AB}\right)^T\right]d\omega \tag{2.11}$$

which has the similar form to UCHF dispersion except the propagators are computed from CKS algorithm.

2.2.2 Monomer-Centered dispersion correction

In the conventional MP2C dispersion correction, a dimer-centered (DC) basis set is used, which means that the basis functions of monomer B are present when doing the SCF calculation for monomer A, and vice versa. Realizing that only the difference in the CKS and UCHF dispersion energies, $\Delta E_{disp}^{MP2C} = E_{disp}^{CKS} - E_{disp}^{UCHF}$ (rather than the individual CKS or UCHF dispersion energies) are important in correcting MP2 interaction energies, and the basis set errors in the CKS and UCHF dispersion are similar and will largely cancel, we propose to improve the computational efficiency of the dispersion correction by using a monomer-centered (MC) basis set. This means that when calculating the propagator of

one molecule, only basis functions from that molecule will be used. With smaller number of basis functions (for a single homo-dimer, the size of basis functions will be cut in half), the cost of bottleneck steps can be reduced by a factor of $\sim 2^3 = 8$.

The energy differences between using MC and DC basis set have been examined extensively in the symmetry-adapted perturbation theory (SAPT) literature. For instance, Williams and co-workers [65] found that the first-order polarization and exchange terms are well-described in an MC basis set, while the second-order induction terms require a DC basis, and the dispersion energy converges slowly with respect to basis set. The dispersion energy in a given MC basis is typically much further from convergence than the dispersion energy in the corresponding DC basis. However, better MC basis results can be obtained if the basis set is enhanced by reoptimizing the basis function exponents for dispersion, augmenting the basis with additional high-angular momentum basis functions, adding mid-bond basis functions, and/or adding a few basis functions to the region of the interacting partner molecule (but fewer than would be present in the full DC basis). All of these strategies can provide MC dispersion energies that are very close to DC ones with fewer overall basis functions [67, 69].

The MP2C correction, however, is computed as the difference between CKS and UCHF dispersion, which proves much less sensitive to the basis set than the absolute dispersion energy as Figure 2.1 shows. The fast convergence behavior stems from the finite basis set error cancellation between CKS and UCHF dispersions.

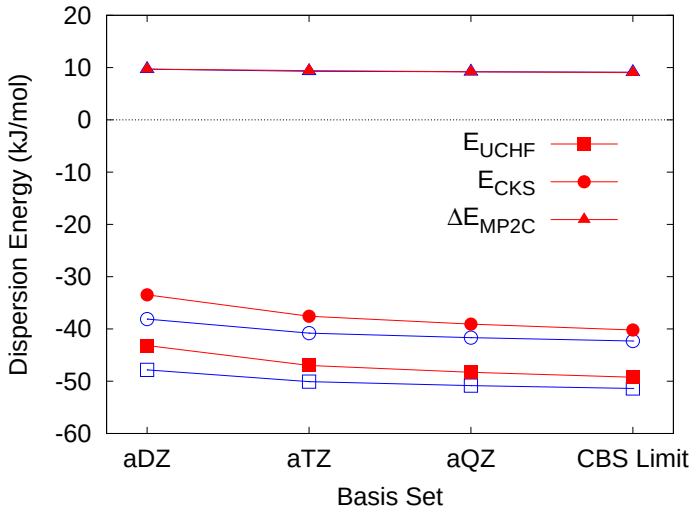

Figure 2.1: Comparison of E_{disp}^{UCHF}, E_{disp}^{CKS} and their difference ΔE_{MP2C} as a function of basis set for the parallel-displaced benzene dimer. Although the individual MC (red) and DC (blue) dispersion energies converge slowly with basis set, the MC and DC MP2C dispersion corrections are nearly identical.

2.2.3 Dispersion correction in HMBI model

As noted above, the largest computational savings from using the MC basis will be obtained when exploiting the symmetry in a molecular crystal. The hybrid many-body interaction (HMBI) model [7, 9, 16], for instance, decomposes the total energy of a crystal using many-body expansion. It treats individual monomers in the unit cell and their short-range pairwise interactions with other molecules using quantum mechanics (QM), while the long-range pairwise and many-body interactions involving three or more molecules are approximated using a polarizable MM force field whose parameters are determined on the fly from *ab initio* [8, 9] calculations. The computational bottleneck in applying a fragment approach like HMBI to a molecular crystal is the need to evaluate dozens or more dimer interactions quantum mechanically.

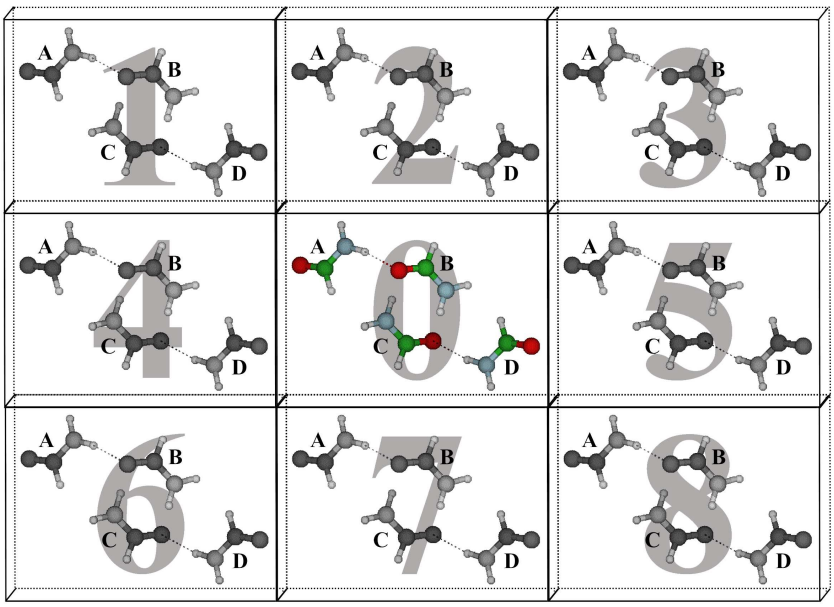

Figure 2.2: 9 unit cells of molecular crystal with 4 monomers (labeled as A,B,C and D) in each, the central unit cell is 0. The propagators computed with MC basis set in image unit cells 1-8 are the same as propagators of molecules with the same geometry and space orientation in the central unit cell.

Figure 2.2 shows 9 unit cells in a molecular crystal where the two-body interactions need to be treated quantum mechanically. When MP2C method is used, the dispersion correction must be computed for each individual dimer. In the original DC MP2C algorithm, for dimer 0A-2C (monomer A in unit cell 0 and monomer C in unit cell 2) as an example, the propagator of monomer 0A needs basis functions in monomer 2C and vice versa. So the propagators must be computed separately for each dimers as required in MP2C method. In the MC algorithm, however, the propagator calculation just depends on the individual molecule, irrespective of which other molecule they happen to be interacting with in a given dimer system. The molecular property of propagator enables the exploitation of crystal periodicity and symmetry to reduce the cost substantially. It means,

one just needs to compute propagators in the central unit cell (0A, 0B, 0C and 0D), while for other monomers in the image unit cells (1-8 as in the figure), they can always find the corresponding monomers with the same geometry and orientation in the central unit cell. When computing the dispersion for dimer 0A-2C, one can just use propagators of 0A and 0C already computed (since the propagator of 2C is identical to that of 0C). The only effort is computing two-index Coulomb integration $J^{0A\text{-}2C} = \langle P^{0A}|r^{-1}|Q^{2C}\rangle$, which has a lower scaling rate $O(N^2)$. Finally, dispersion energies are obtained from Casimir-Polder integration as Equation 2.7 and 2.11.

Similar ideas have been used previously when mapping out potential energy surfaces for the interaction between two rigid molecules using SAPT(CKS), for instance. In that context, one can compute the molecular propagator once and then evaluate the interaction energies at many different intermolecular distances and orientations cheaply [67]. In a molecular crystal, the application of periodic boundary conditions and/or space-group symmetry imposes constraints that make the propagators transferable from one monomer to other geometry-identical ones.

2.3 Results and discussion

To demonstrate the performance of the monomer-centered basis MP2C algorithm, we first examine the basis set behavior for dimer interactions. Then we examine its performance on benchmark data set for noncovalent interactions. Finally, we apply it to several different molecular crystals and compare the timings.

2.3.1 Comparison between MC and DC algorithms

To see how well MC approaches DC algorithm, MP2C dispersion correction with different basis sets are tested on S22 benchmark test set [42, 43]. As Figure 2.3 shows, the correlation is extremely good across a range of basis sets. The root-mean-square deviations between the MC and DC algorithms are only 0.25 kJ/mol (aug-cc-pVDZ), 0.14 kJ/mol (aug-cc-pVTZ), 0.13 kJ/mol (aug-cc-pVQZ), and 0.14 kJ/mol (CBS limit).

In Table 2.1, MP2 interaction energies are computed in the CBS limit to avoid finite basis set error introduced in MP2 level. Therefore, one just need to focus on the post-MP2 correlation compared to CCSD(T) benchmark. Without post-MP2 correlation correction, the RMS error is 5.73 kJ/mol with the large errors mainly from the dispersion dominant complexes. The conventional dimer centered MP2C dispersion correction substantially improves the performance of MP2 by replacing the inaccurate UCHF dispersion lacking enough intramolecular correlation with the more accurate CKS dispersion. As discussed before, the finite basis set errors of MP2C dispersion correction converge fast due to the error cancellations between CKS dispersion and UCHF dispersion, with similar RMS errors that 0.87 kJ/mol for aug-cc-pVDZ, 0.73 kJ/mol for aug-cc-pVTZ, 0.71 kJ/mol for aug-cc-pVQZ, and 0.71 kJ/mol for CBS limit. In the following tests, we will just use the medium sized aug-cc-pVTZ basis set for dispersion correction, which will not lead to clear difference.

The error cancellation can also be found between CKS and UCHF dispersions when they are computed with MC basis set. For some of the complexes, errors with MC algorithm are smaller than DC algorithm, while for others, MC results are slightly worse.

Table 2.1: Interaction energies (kJ/mol) computed from MP2, MP2C/DC, MP2C/MC and CCSD(T) on S22 test set. For MP2C method, dispersion correction with different basis set is added to MP2/CBS interaction energies.

	MP2	$+\Delta MP2C/DC$				$+\Delta MP2C/MC$				CCSD(T)
structure	CBS	aDZ	aTZ	aQZ	CBS	aDZ	aTZ	aQZ	CBS	CBS
$(NH_3)_2$	-13.22	-13.43	-13.72	-13.81	-13.89	-13.32	-13.60	-13.72	-13.81	-13.26
$(H_2O)_2$	-20.84	-20.88	-21.17	-21.26	-21.34	-20.75	-21.09	-21.17	-21.22	-21.00
Formic acid dimer	-77.70	-77.32	-77.99	-78.29	-78.49	-76.85	-77.71	-77.94	-78.11	-78.66
Formamide dimer	-66.27	-66.35	-66.98	-67.19	-67.36	-65.88	-66.72	-66.93	-67.09	-67.45
Uracil dimer HB	-85.40	-84.61	-85.19	-85.36	-85.48	-84.29	-85.13	-85.26	-85.35	-86.57
2-pyridoxine·2-aminopyridine	-72.68	-71.42	-72.01	-72.22	-72.34	-70.91	-71.94	-72.04	-72.10	-71.13
Adenine·thymine WC	-69.20	-68.45	-69.03	-69.24	-69.41	-68.08	-69.09	-69.14	-69.17	-70.04
$(CH_4)_2$	-2.05	-2.26	-2.30	-2.34	-2.38	-2.25	-2.27	-2.30	-2.32	-2.22
$(C_2H_4)_2$	-6.61	-6.65	-6.78	-6.82	-6.86	-6.60	-6.72	-6.79	-6.84	-6.28
Benzene·CH_4	-7.57	-6.31	-6.44	-6.48	-6.53	-6.15	-6.37	-6.46	-6.53	-6.07
PD benzene dimer	-20.75	-11.04	-11.46	-11.55	-11.59	-11.03	-11.32	-11.54	-11.71	-10.96
Pyrazine dimer	-28.91	-18.28	-18.37	-18.53	-18.66	-18.43	-18.47	-18.55	-18.60	-17.57
Stacked uracil dimer	-46.44	-39.29	-39.58	-39.83	-40.00	-39.21	-39.36	-39.69	-39.93	-40.75
Stacked indole·benzene	-33.85	-18.87	-19.33	-19.50	-19.62	-18.82	-19.24	-19.53	-19.74	-19.20
Stacked adenine·thymine	-62.05	-47.28	-47.62	-47.91	-48.12	-47.15	-47.38	-47.73	-47.97	-48.79
Ethene·ethine	-6.99	-6.66	-6.74	-6.74	-6.74	-6.57	-6.67	-6.71	-6.74	-6.32
Benzene·H_2O	-14.81	-13.60	-13.85	-13.89	-13.93	-13.47	-13.70	-13.80	-13.87	-13.77
Benzene·NH_3	-11.13	-9.71	-9.83	-9.87	-9.92	-9.54	-9.75	-9.83	-9.88	-9.71
Benzene·HCN	-21.59	-19.37	-19.58	-19.62	-19.66	-19.07	-19.45	-19.55	-19.63	-19.04
T-shaped benzene dimer	-15.19	-11.93	-12.09	-12.09	-12.09	-11.65	-12.00	-12.09	-12.15	-11.34
T-shaped indole·benzene	-29.16	-24.56	-24.81	-24.89	-24.94	-24.25	-24.70	-24.79	-24.86	-23.51
Phenol dimer	-32.47	-29.96	-30.34	-30.46	-30.54	-29.78	-30.25	-30.36	-30.45	-29.66
RMS error	5.73	0.87	0.73	0.71	0.71	1.00	0.76	0.73	0.72	

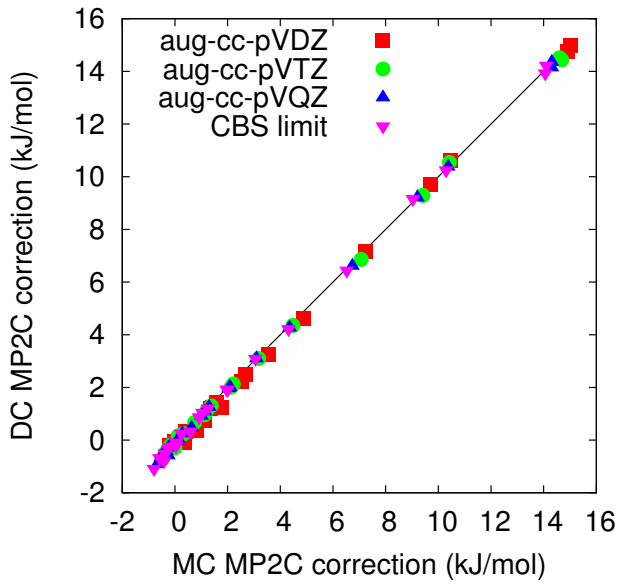

Figure 2.3: Correlation between MC and DC MP2C correction for the S22 test set.

Statistically, the differences in the RMS error between the MC and DC algorithms are almost negligible. Specifically, the RMS of MC MP2C is 0.13 kJ/mol worse than the RMS error of DC approach when using aug-cc-pVDZ basis set, 0.03 kJ/mol worse in the aug-cc-pVTZ case, 0.02 kJ/mol worse in aug-cc-pVQZ basis set, and only 0.01 kJ/mol difference in CBS limit.

2.3.2 Performance on benchmark dimer sets

Molecular crystal packing is governed by a broad spectrum of intermolecular interactions occurring at a variety of distances and orientations. Accordingly, accurate crystal structure modeling requires small, uniform errors across the potential energy surface. MP2C performs well in this regard [1, 44, 35], and the data presented below indicates that using the MC basis MP2C algorithm does not negatively impact this behavior.

Table 2.2: RMS error (kcal/mol) for MP2 and MP2C relative to CCSD(T)/CBS-limit benchmarks on the S22x5 and S66a8 test sets.

	MP2	MC MP2C	DC MP2C
S22x5 set			
$0.9R_e$	8.76	1.53	1.51
$1.0R_e$	5.69	0.80	0.73
$1.2R_e$	2.21	0.48	0.52
$1.5R_e$	0.77	0.28	0.25
$2.0R_e$	0.22	0.17	0.13
Overall	4.79	0.81	0.79
S66a8 set	1.77	1.06	

To test the distance and orientation behavior of MP2C, MC approach is used for the S22x5 [44] and S66a8 [46] test suites. The S22x5 set scales the S22 dimers from the equilibrium separation R_e into shorter and longer intermolecular distances: $0.9R_e$, $1.0R_e$ (traditional S22), $1.2R_e$, $1.5R_e$ and $2.0R_e$. The S66a8 set contains a mixture of 66 dimers at 8 different relative angles generated by rotating either molecules by $\pm 30^\circ$ in the two directions orthogonal to the principal plane of the dimer.

Table 2.2 lists the RMS errors for both test sets. As expected, the MP2C method significantly improves upon the uncorrected MP2 results. For the S22x5 set, the improvements are most dramatic for the separations in the vicinity of the equilibrium intermolecular distance ($0.9R_e$, $1.0R_e$ and $1.2R_e$), where the dispersion interactions are strongest. Overall, the MC MP2C correction reduces the S22x5 RMS errors from 4.79 kJ/mol to 0.81 kJ/mol, which is just slightly worse than the DC correction with 0.79 kJ/mol. MP2 performs better for the S66a8 test set, while MP2C dispersion correction is not as significantly as S22x5. Because aug-cc-pVTZ basis set is used for MP2 level of calculation where the finite basis

set errors are nonnegligible, otherwise the RMS error would be even smaller. The improvement is fairly uniform across the eight different angular arrangements, so the results are not decomposed further into different orientations. These benchmark dimer tests indicate that the MC basis MP2C algorithm performs well across a variety of intermolecular separations and orientations, making it well-suited for molecular crystal studies.

2.3.3 Performance on molecular crystals

While the MC algorithm provides appreciable computational savings in a single dimer calculation, the savings in a molecular crystal are much greater. Therefore, it is important to test how well this algorithm performs for molecular crystal systems.

First, the lattice energies of seven small-molecule molecular crystals are computed with MP2C method in MC and DC basis set, and then compared against earlier estimated complete-basis-set CCSD(T) benchmarks [7, 9] in Table 2.3. As discussed in the original works, those benchmark lattice energies provide 1-2 kJ/mol agreement with experiment, which is probably within the experimental errors. Compared to CCSD(T), MP2 performs well for the hydrogen-bonded crystals, but it significantly overestimates the lattice energy

Table 2.3: Lattice energy for seven small-molecule molecular crystals (kJ/mol)

	MP2	DC MP2C	MC MP2C	CCSD(T)	Experiment
Ice	59.8	60.5	60.2	60.1	59
Formamide	78.7	79.5	79.0	80.5	82±0.3
Acetamide	80.6	80.6	80.8	80.4	86±2
Imidazole	102.0	92.6	92.9	88.6	91±4
Benzene	61.7	49.0	49.0	50.9	52±3
NH_3	39.4	41.1	40.7	40.9	39
CO_2	31.0	27.7	27.7	31.3	31

in imidazole and benzene, both of which exhibit strong dispersion interactions.

Here, we find that MP2C also performs well relative to CCSD(T) for these crystals. For the hydrogen-bonded crystals (ice, formamide, acetamide, and ammonia), the MP2C correction is small and provides only a minor improvement to the already accurate MP2 lattice energies. For benzene and imidazole, where MP2 overestimates the lattice energies by 10.8 kJ/mol (21%) and 13.4 kJ/mol (15%), respectively. MP2C reduces the errors to about 1.9 kJ/mol (4%) and 4.0 kJ/mol (4%). Interestingly, for carbon dioxide, the MP2 and CCSD(T) lattice energies differ by only 0.3 kJ/mol, and the MP2C correction increases the error to 3.6 kJ/mol (12%). Symmetry adapted perturbation theory calculation suggests that interactions between CO_2 molecules at the shorter ranges found in the crystal involve significant contributions from a variety of intermolecular interactions (electrostatic, induction, exchange repulsion, and dispersion) [70]. Perhaps the close agreement between MP2 and CCSD(T) in CO_2 reflects a fortuitous cancellation of errors among these different terms, and correcting the dispersion with MP2C disrupts this error cancellation. Again, MC algorithm predicts the lattice energies close enough to those computed with DC basis set, with minor difference within 5 kJ/mol.

Next, we examine the performance of the new MP2C algorithm for two poly-

Table 2.4: Lattice energies for the two known polymorphs of aspirin in the aug-cc-pVTZ basis (kJ/mol).

	MP2	SCS(MI)-MP2	MC MP2C	Experiment
Form I	132.1	135.6	116.1	115
Form II	132.0	135.5	116.3	
$\Delta E_{I \rightarrow II}$	0.1	0.1	-0.1	

morphic crystals: aspirin and oxalyl dihydrazide. For aspirin, the two polymorphs have very similar crystal packing, and we previously showed that the two polymorphs are virtually degenerate [71], which helps to explain the disordered mixtures of the two that are often observed experimentally [72]. Both MP2 and SCS(MI)-MP2 predict that form I is 0.1 kJ/mol more stable than form II. However, the lattice energies at both levels of theory seem too large compared to the experimental value, as shown in Table 2.4. This experimental lattice energy was obtained by extrapolating the experimental heat of sublimation (ΔH_{sub}(298K)=109.7±0.5 kJ/mol [73]) to 0 K using the standard, simple model, $E_{lattice} = \Delta H_{sub} + 2RT$ [74].

Here, we apply MC MP2C/aug-cc-pVTZ to the same aspirin crystal structures and aug-cc-pVTZ basis set as in the earlier work [71]. Similar to MP2 and SCS(MI)-MP2, MC MP2C predicts that the two crystal polymorphs are virtually degenerate. It does nominally reverse the stability ordering of the two (form II is now preferred by 0.1 kJ/mol instead of form I by 0.1 kJ/mol), though such energy differences are within the expected error bars of the calculations. More importantly, MP2C corrects the apparent overestimation of the aspirin lattice energy. It predicts a value around 116 kJ/mol, compared to the 0 K extrapolated experimental value of 115 kJ/mol [74].

Finally, the five polymorphs of oxalyl dihydrazide provide a particularly challenging example of polymorphism. The polymorphs differ in the degree of intra- versus inter-molecular hydrogen bonding and require a careful electronic structure treatment to obtain reasonable results. Earlier results obtained using force fields and various density functionals were implausible [75]. However, a combination of MP2C, large basis set, and

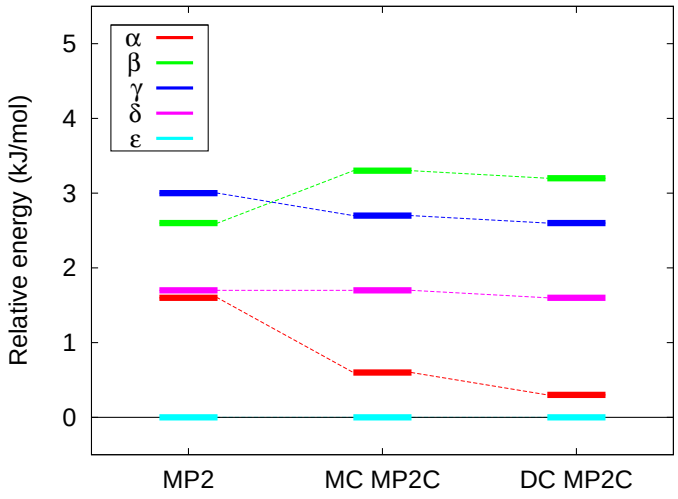

Figure 2.4: Relative polymorph energies for oxalyl dihydrazide. The MP2 ordering is incorrect, while both the MC and DC MP2C algorithms produce identical orderings with very similar energies.

zero-point energy leads to predictions that are consistent with the available experimental data [50]. Experimentally, the α, δ and ϵ polymorphs are thought to be the most stable, while the γ polymorph is less stable and β is the least stable form [76]. Conventional MP2 does not properly order the β and γ polymorphs, but MP2C does.

Here, we compare the results of the MC and DC MP2C algorithms using the same structures and procedures as in the earlier work [50]. Specifically, MP2/CBS-limit lattice energies are combined with the MP2C correction computed in the aug-cc-pVTZ basis. The overall MC lattice energies turn out to be about 1 kJ/mol ($\sim$0.5%) smaller than the DC ones. As shown in Figure 2.4, the differences in the relative polymorph energies are even smaller: the largest difference between the MC and DC relative energies is 0.3 kJ/mol for the α polymorph, while the other MC and DC relative energies differ by only 0.1 kJ/mol.

All these examples highlight the value of using MP2C for predicting molecular

crystal energetics. Although the MC basis algorithm can lead to minor error accumulation in a crystal (e.g., 1 kJ/mol difference between MC and DC lattice energies in the oxalyl dihydrazide versus 0.1-0.2 kJ/mol in the S22 set), the relative energies remain in excellent agreement with the DC algorithm. Overall, MC basis MP2C reliably describes the intermolecular interactions that occur in molecular crystals.

2.3.4 Timings

Having demonstrated the accuracy of the new MP2C algorithm, we now compare timings between the MC and DC algorithms. For an individual dimer MP2C correction, the computational savings are non-trivial. Because one typically computes the MP2C correction in conjunction with a counterpoise-corrected MP2 calculation, the converged HF orbitals used in computing the UCHF and CKS response functions for either an MC or DC basis are already available. Therefore, the MP2C correction timings here exclude the HF times.

For a single dimer of oxalyl dihydrazide in the aug-cc-pVTZ basis, for example, the counterpoise-corrected dual basis RI-MP2 calculation requires $\sim$6 hours on a single core of an Intel Xeon W3520 2.66 GHz processor with 2 GB of RAM. The DC MP2C correction adds another $\sim$1.8 h to the calculation time. By contrast, the MC MP2C correction is roughly five times faster at only 20 min.

The computational savings magnify in molecular crystal calculations because the UCHF and CKS response functions only need to be computed once for each symmetry-unique monomer in the unit cell in the MC basis approach. For the α polymorph of oxalyl dihydrazide, one must perform one monomer and 41 symmetry-unique dimer calculations.

Table 2.5: Approximate counterpoise-corrected MP2C single-point energy timings in hours for crystalline oxalyl dihydrazide (α form) and aspirin (form I). The ΔMP2C timings represent the time for the dispersion correction only; total job times are obtained by adding ΔMP2C and RI-MP2 timings.

	Oxalyl dihydrazide		Aspirin	
	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
Number of monomers	1	1	1	1
Number of dimers	41	41	47	47
RI-MP2	48 h	250 h	425 h	2450 h
DC ΔE_{MP2C}	23 h	72 h	135 h	390 h
MC ΔE_{MP2C}	0.2 h	0.6 h	0.8 h	2.8 h

Performing the counterpoise-corrected, dual-basis RI-MP2/aug-cc-pVTZ calculations requires ∼250 h, and the DC MP2C correction adds an additional 72 h. By contrast, the MC-based MP2C correction for the entire crystal requires less than 40 min, making it two orders of magnitude faster than the DC correction and a trivial expense compared to the underlying RI-MP2 calculations. Of those nearly 40 min, about 10 min are required to calculate the response functions for the one unique monomer, while the remaining almost 30 min are spent computing the dispersion correction for the 41 dimers (at ∼40 s per dimer).

For aspirin form I, one monomer and 47 dimers must be computed with HMBI (again exploiting space-group symmetry). The counterpoise-corrected dual-basis RI-MP2/aug-cc-pVTZ calculation requires around ∼2450 h. With the DC algorithm, the MP2C correction for the entire crystal requires an additional 390 h. By contrast, with the new MC algorithm, the MP2C correction can be computed for all the dimers in the crystal in less than 3 h. In all cases shown in Table 2.5, the MC basis algorithm reduces the MP2C correction cost from ∼15%-35% to less than 1% of the overall job time. Clearly, any further

computational savings in MP2C must be achieved by reducing the cost of the underlying MP2 calculations.

2.4 Conclusions

In conclusion, the MP2C dispersion correction can be computed using a monomer-centered basis instead of a dimer-centered one with virtually no loss in accuracy, thanks to excellent basis set error cancellation. This error cancellation also works for other dispersion energies with different propagators, for example the (CHF - UCHF) dispersion correction in our test also proves that the difference between MC algorithm and DC algorithm is negligible. For simple dimer calculations, this change reduces the costs of evaluating the dispersion correction by several-fold. For fragment-based molecular crystal calculations, however, one can exploit the translational symmetry to reduce the MP2C correction cost by two orders of magnitude, reducing the cost of the dispersion correction from a significant expense to virtually free. Overall, monomer-centered basis MP2C provides an excellent balance between accuracy and efficiency that makes reliable electronic structure calculations feasible for many systems, including chemically interesting molecular crystals. Of course, the approach described here does not alter the costs of the underlying MP2 calculation and the next chapter will address the problem by improving the computational efficiency of MP2 part in the MP2C model.

Chapter 3

Improving computational efficiency of MP2 correlation for two-body interactions

3.1 Introduction

As discussed in last chapter, evaluating the MP2C dispersion correction with a monomer-centered (MC) basis set instead of a dimer-centered (DC) one substantially improves the computational efficiency without compromising accuracy [77]. For a single dimer interaction energy calculation, this simple change reduces the computational time by a factor up to ∼8. On the other hand, one can achieve ∼100-fold speedups in a fragment-based molecular crystal calculation by utilizing translational symmetry in such periodic systems.

Although the MC basis MP2C dispersion correction offers significant speedups, it

is in some sense computationally absurd. When considering the cost of both the RI-MP2 calculation (including Hartree-Fock) and the dispersion correction, the RI-MP2 calculation dominates the computational cost. The RI-MP2 calculation consumes 65-85% of the time when a DC basis correction is used, and more than 90% when the MC basis correction is used. A sizable fraction of that RI-MP2 time is spent computing the UCHF dispersion, only to discard that contribution and replace it with an inexpensive $O(N^4)$ CKS dispersion energy. Could we simply avoid computing those long-range MP2 correlation effects in the first place and thereby drastically reduce the necessary computational effort?

One can analyze the composition of the MP2 interaction energies using the related SAPT2 method. As the potential energy curve of the benzene dimer in Figure 1.3 shows, for example, the dispersion is dominant at larger intermolecular separations and it still contributes strongly at shorter distances where the total molecular interaction increases. Generally, the nonnegligible but comparatively weaker short-range correlation energies require more effort to compute (e.g. exchange-dispersion scales $O(N^5)$ with system size N, versus $O(N^4)$ for the dispersion). Is there any way to handle the cheaper dispersion accurately while treating the expensive short-range correlation more efficiently?

To answer the first question, we utilize Coulomb attenuation to eliminate long-range correlation at the MP2 level. Coulomb attenuation/range-separation provides an effective means for separating short- and long-range Coulomb interactions, which has been used to increase algorithmic efficiency or develop physically improved density functionals, for instance [78–104]. Then, we apply a modified MP2C dispersion correction that removes any residual UCHF dispersion remaining in the attenuated MP2 model and adds the full-

range CKS dispersion energy to capture the long-range electron correlation. This model enables aggressive Coulomb attenuation at the MP2 level, which allows for both substantial computational savings and higher accuracy than that provided by conventional finite basis MP2C. Specifically, benchmark tests indicate that attenuation reduces the MP2C/aug-cc-pVTZ errors roughly in half and gives results approaching complete basis set (CBS) quality at drastically lower cost. We also provide new physical insights into how attenuated MP2 approximations are able to achieve such good performance.

As for the second question, one should first notice that most of the finite basis set errors stem from correlation interaction, instead of Hartree-Fock interaction energy. Dispersion energy is dominant in long-range correlation, which can be seen when two molecules are well separated that the total correlation is almost dispersion, or from the attenuation curve as we will discuss later. To reduce errors in the correlation energy, one should treat dispersion energy as accurately as possible. In the context of MP2C model, the CKS dispersion energy should be computed in the CBS limit to reduce most of the finite basis set errors. While for the more complicated short-range correlation which exhibits less sensitivity to basis set incompleteness, there is a chance to sacrifice a little accuracy for greater efficiency. Under this consideration, one can achieve speedups for the most steeply scaling $O(N^5)$ portion of the calculation by reducing the number of basis functions N. Generally, the finite basis set error varies with the strength of the intermolecular interactions. To ensure the general applicability of the models developed here, we will across a variety of systems. A conflict emerges with this scheme, however. To obtain accurate dispersion without significant basis set errors, one cannot continue utilizing the MC algorithm developed

in last chapter, so we use the DC algorithm in that case instead.

In addition to the methods described above, we also test the performance of explicitly correlated MP2 (MP2-F12) [51, 105] algorithm, which is an alternative method for approaching the CBS limit with small basis sets. In particular, we compare the performance of the new methods developed here to results from MP2C-F12/aDZ and MP2C-F12/aTZ.

3.2 Theory

3.2.1 Attenuated MP2C

In the attenuated MP2 approach of Goldey and Head-Gordon [102,103], one range separates the Coulomb operator according to

$$\frac{1}{r} = \frac{\mathrm{terf}(r; r_0)}{r} + \frac{\mathrm{terfc}(r; r_0)}{r} \tag{3.1}$$

where [106]

$$\mathrm{terf}(r; r_0) = \frac{1}{2}\left\{\mathrm{erf}\left[\frac{r - r_0}{\sqrt{2}r_0}\right] + \mathrm{erf}\left[\frac{r + r_0}{\sqrt{2}r_0}\right]\right\} \tag{3.2}$$

and

$$\mathrm{terfc}(r; r_0) = 1 - \mathrm{terf}(r; r_0) \tag{3.3}$$

The function $\mathrm{terfc}(r; r_0)/r$ describes the short-range Coulomb interaction, while $\mathrm{terf}(r; r_0)/r$ describes the long-range contribution, as Figure 3.1 shows. The user-defined parameter r_0 controls the relative length-scales of the two components, with smaller values of r_0 corresponding to a more rapid decay of the short-range $\mathrm{terfc}(r; r_0)/r$ term. While many different forms of the attenuation function are possible, the $\mathrm{terfc}(r; r_0)/r$ form maintains

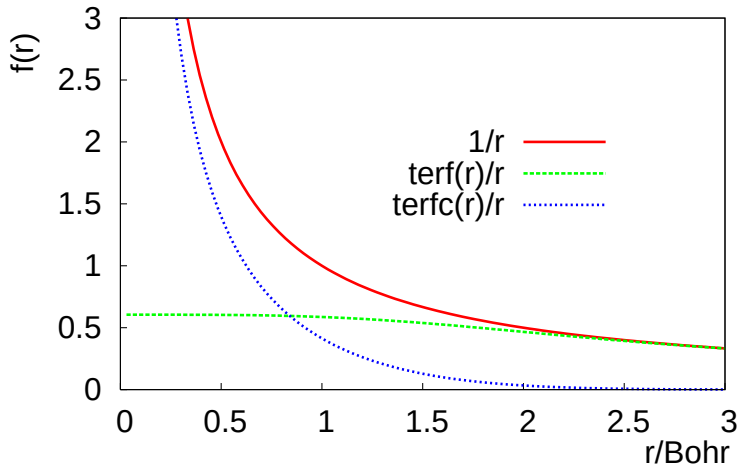

Figure 3.1: Partitioning of Coulomb operator $1/r$ into short-range interaction $\mathrm{terfc}(r;r0)/r$ and long-range interaction $\mathrm{terf}(r;r_0)/r$. The parameter r_0 adjusts the partitioning length-scale.

the correct curvature of the Coulomb operator at short-range [106]. This helps ensure that the short-range correlation is minimally affected by the elimination of the long-range correlation.

Applying this Coulomb partitioning to MP2 correlation, one discards the long-range correlation by neglecting $\mathrm{terf}(r;r_0)/r$ and replacing the Coulomb operator $1/r$ with $\mathrm{terfc}(r;r_0)/r$ when evaluating the two-electron integrals for the MP2 correlation energy. In the context of a typical resolution of the identity (RI) MP2 implementation (i.e., one that does not fully exploit the sparsity introduced by Coulomb attenuation) [107], this corresponds to constructing the key intermediate B_{ia}^P tensors according to

$$B_{ia}^P = \sum_Q \left(ia \left| \frac{\mathrm{terfc}(r;r_0)}{r} \right| Q \right) \left(Q \left| \frac{\mathrm{terfc}(r;r_0)}{r} \right| P \right)^{-1/2} \tag{3.4}$$

In other words, the attenuated Coulomb operator is employed both as the density-fitting metric and for the Coulomb interaction between the pair of electrons. The MP2 correlation

energy is then computed as usual by taking products of such B tensors and dividing by the appropriate energy denominator.

The MP2C correction, as discussed in last chapter, works by replacing the UCHF treatment of dispersion found in MP2 with an improved one computed at the CKS level. The Coulomb attenuation algorithm removes some, but not all, of the intermolecular dispersion from MP2. Therefore, when restoring the long-range dispersion at the CKS level, one must take care to avoid double-counting the intermediate-range dispersion contribution. Accordingly, in attenuated MP2C, we subtract out whatever residual dispersion remains in attenuated MP2 by computing the UCHF contribution with the same attenuated Coulomb operator, and then we add the full CKS dispersion (without any attenuation)

$$E_{MP2C}(atten) = E_{MP2}(atten) - E_{disp}^{UCHF}(atten) + E_{disp}^{CKS}(full) \tag{3.5}$$

In this scheme, the CKS dispersion can be evaluated using the same algorithm as in the conventional MP2C model. However, the UCHF dispersion algorithm must be modified to incorporate Coulomb attenuation so it will be consistent with attenuated MP2. First, we replace the standard $1/r$ Coulomb operator with $\mathrm{terfc}(r;r_0)/r$ when computing the UCHF propagators

$$\begin{aligned}[\chi_0(\omega)]_{PQ} &= -4\sum_{ia} D_{ia,P} C_0^{ia}(\omega) D_{ia,Q} \\ &= -4\sum_{ia} S^{-1}\left[\frac{(P|\mathrm{terfc}(r;r_0)/r|ia)\epsilon_{ia}(ia|\mathrm{terfc}(r;r_0)/r|Q)}{\epsilon_{ia}^2+\omega^2}\right] S^{-1}\end{aligned} \tag{3.6}$$

where i and a are occupied and virtual molecular orbitals, and P and Q are auxiliary basis functions for density fitting. $D_{ia,P}$ is the attenuated density fitting coefficient. C_0^{ia} is the expansion coefficient for uncoupled FDDS as discussed in last chapter. The quantity

$\epsilon_{ia} = \epsilon_i - \epsilon_a$ is the energy difference between occupied and virtual orbitals computed at the Hartree-Fock level. The S matrix also utilizes the attenuated Coulomb operator, $S = \langle P|\text{terfc}(r;r_0)/r|Q\rangle$. Also, the intermolecular Coulomb integrals should be attenuated as $J^{AB} = (P|\text{terfc}(r;r_0)/r|Q)$, which appears in the Casimir-Polder integration

$$E_{disp}^{UCHF} = -\frac{1}{2\pi}\int_0^\infty Tr\left[\chi_0^A(\omega)J^{AB}(\chi_0^B(\omega))^T(J^{AB})^T\right]d\omega \tag{3.7}$$

Note that all of the necessary MP2C integrals can be performed using either a dimer-centered [35] or monomer-centered basis [77], which we will discuss later.

The final step requires choosing the parameter r_0 to determine how aggressively the Coulomb operator will be attenuated. In the limit as $r_0 \to 0$, $\text{terfc}(r;r_0)/r \to 0$, and one attenuates the MP2 correlation energy completely, leaving only the Hartree-Fock energy plus CKS dispersion. In the limit where $r_0 \to \infty$, $\text{terfc}(r;r_0)/r \to 1/r$, however, restoring back to the full Coulomb operator and conventional MP2C. For small molecule dimers like those in the S66 test set, as long as one does not choose a very small r_0 parameter (i.e. aggressive attenuation), attenuating the propagators will not significantly affect the dispersion energies. For example, attenuating the propagators with $r_0 = 0.9$ Å gives results that are very identical to the nonattenuated dispersion. This means the short-range Coulomb operator can still capture the whole correlation effect regarding these small sized molecules. In practice, the parameter r_0 is chosen empirically, as described in later sections.

3.2.2 Small basis set for short-range correlation

For high accuracy calculations, both complicated correlation methods and large basis sets are required. Both lead to high computational costs that grow steeply with

system size. Coulomb operator partitioning improves the computational efficiency for the correlation calculation, as we will demonstrate below. As for the basis set, further speedups can be achieved by treating the more steeply scaling $O(N^5)$ short-range correlation with a small basis set. While we need to describe the relatively cheaper scaling $O(N^4)$ long-range correlation accurately, since it is more sensitive to basis set. Accordingly, an alternative MP2C model can be expressed as

$$E_{int}^{MP2C/SR-aXZ} = E_{int}^{HF}(CBS) + \left[E_{int}^{MP2}(aXZ) - E_{disp}^{UCHF}(DC, aXZ)\right] + E_{disp}^{CKS}(DC, CBS) \tag{3.8}$$

where the short-range correlation (MP2 correlation without dispersion) is computed with small basis set, and long-range correlation (CKS dispersion) is computed in CBS limit. One should notice that monomer-centered dispersion does not converge to the dimer-centered result, even in the CBS limit (as shown in Figure 2.1). Compared to UCHF dispersion, the CKS algorithm requires extra time to transform from Hartree-Fock orbitals into local Hartree-Fock (LHF) orbitals and to integrate the exchange-correlation kernel. It would be more efficient if the cheaper UCHF dispersion is used to correct for the finite basis set error in correlation, and the MC dispersion correction is used to reduce the correlation error. Then the improved model is

$$\begin{aligned} E_{int}^{MP2C/SR-aXZ} &= E_{int}^{HF}(CBS) + E_{int}^{MP2}(aXZ) \\ &+ \left[E_{disp}^{UCHF}(DC, CBS) - E_{disp}^{UCHF}(DC, aXZ)\right] \\ &+ \left[E_{disp}^{CKS}(MC, CBS) - E_{disp}^{UCHF}(MC, CBS)\right] \end{aligned} \tag{3.9}$$

Compared to MC MP2C/CBS method, the only difference is the finite basis set for short-range correlation. Of course, the terms $E_{int}^{MP2}(aXZ)$, $E_{disp}^{UCHF}(DC, CBS)$ and $E_{disp}^{UCHF}(DC, aXZ)$

can be attenuated to exploit sparsity as attenuated MP2C.

3.3 Results and discussion

3.3.1 Understanding attenuated MP2

First, we examine the behavior of Coulomb attenuation by performing an error analysis of attenuated MP2 model on the S66 dimer test set [45]. To do so, we decompose the attenuated MP2 errors relative to the estimated CBS limit CCSD(T) benchmarks [45], $E^{attMP2}(aTZ, no\text{-}CP) - E^{CCSD(T)}(CBS, CP)$ into contributions arising from the finite basis set, the post-MP2 correlation error, and the error introduced by attenuating the Coulomb operator. For attenuated MP2/aug-cc-pVTZ without CP-correction, for instance, these errors are defined as

Finite basis set error:

$$E^{MP2}(aTZ, no\text{-}CP) - E^{MP2}(CBS, CP) \tag{3.10}$$

Correlation error:

$$E^{MP2}(CBS, CP) - E^{CCSD(T)}(CBS, CP) \tag{3.11}$$

Attenuation error:

$$E^{attMP2}(aTZ, no\text{-}CP) - E^{MP2}(aTZ, no\text{-}CP) \tag{3.12}$$

The total error is the sum of these three contributions. Counterpoise-corrected attenuated MP2/aug-cc-pVTZ can be partitioned analogously, except the counterpoise correction is applied to all terms in the expressions above. For attenuated MP2/aug-cc-pVTZ without

counterpoise correction (Figure 3.2a), basis set superposition error leads to overbinding of the dimer. The inadequate treatment of electron correlation in MP2 also frequently leads to overbinding the dimers, particularly for the dimers where dispersion interactions are important. Figure 3.2a shows that these errors range from a few kJ/mol up to nearly 10 kJ/mol in π-stacking cases like benzene dimer. Together, these two error sources lead to systematic overbinding of the dimers. On the other hand, applying Coulomb attenuation reduces the long-range intermolecular correlation (i.e., attractive long-range dispersion interactions) and weakens the binding. Attenuated MP2 works, therefore, by choosing an appropriate value of r_0 (e.g., $r_0 = 1.35$ Å for the aug-cc-pVTZ basis [103]) such that one attenuates enough of the long-range correlation to cancel out the large errors arising from the finite basis set and the MP2-level treatment of correlation.

Applying a counterpoise correction to attenuated MP2 greatly reduces the basis set error, as shown in Figure 3.2b. It also changes the sign of the basis set error. Counterpoise-corrected aug-cc-pVTZ interaction energies are generally underbound, unlike their noncounterpoise-corrected counterparts. However, MP2 correlation still often overestimates the interaction energies, and the sum of the basis set and correlation errors leads to overbinding for many of the S66 dimers. Once again, one can choose an appropriate degree of Coulomb attenuation to reduce this overbinding. Because the errors that need to be cancelled in this case are generally smaller, one attenuates less aggressively to keep a larger fraction of the long-range correlation. Hence, the optimal r_0 value for counterpoise-corrected MP2/aug-cc-pVTZ is 1.75 Å instead of 1.35 Å for the noncounterpoise-corrected case [103]. There are notable cases like the acetamide dimer, however, where the combined

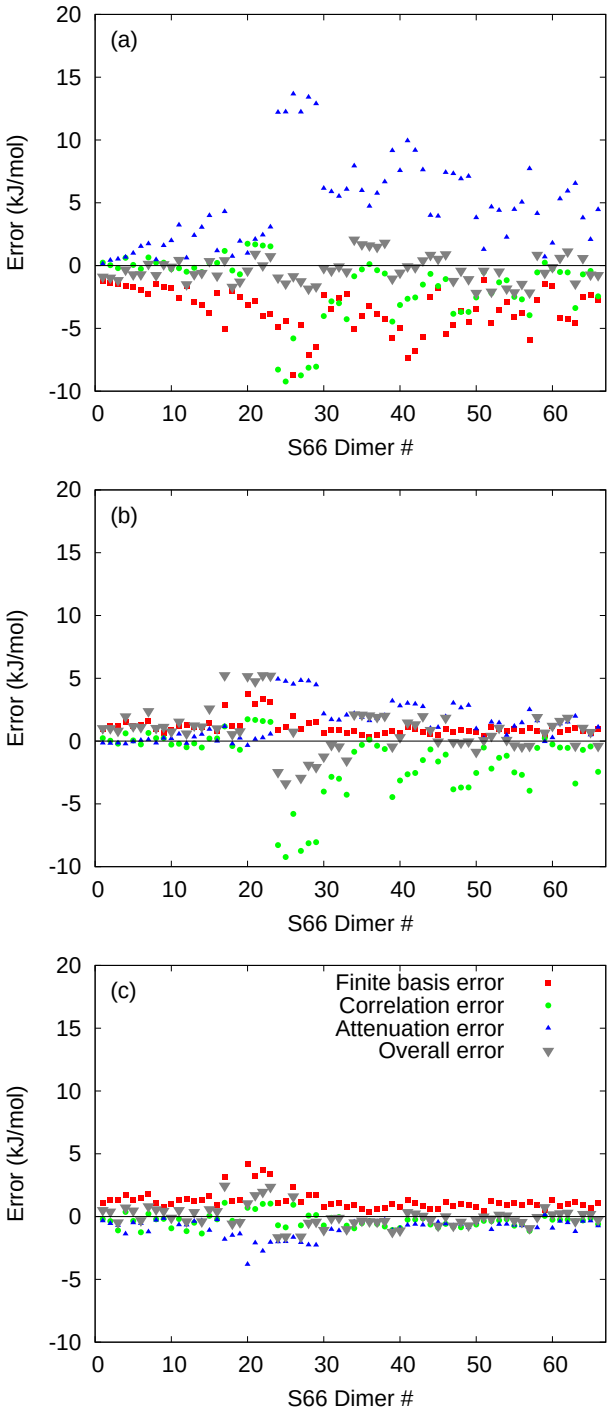

Figure 3.2: Error contributions for each dimer in the S66 set and aug-cc-pVTZ basis set for (a) attenuated MP2 without counterpoise correction ($r_0 = 1.35$ Å), (b) attenuated MP2 with counterpoise correction ($r_0 = 1.75$ Å), and (c) attenuated MP2C with counterpoise correction ($r_0 = 0.9$ Å). The energy-decomposition of the errors is described in the main text.

basis set and MP2 correlation errors lead to underbinding when a counterpoise correction is employed. In such cases, attenuating the long-range attraction actually increases the errors. This explains the earlier finding that attenuated-MP2 actually performs better without a counterpoise correction for the S66 test set [103].

To summarize, attenuated MP2 works for intermolecular interactions by cancelling several different source of errors with opposite signs. One attenuates away enough of the attractive dispersion interaction from MP2 to cancel the overbinding of intermolecular interactions that typically occurs with MP2 in finite basis sets. Because attenuated MP2 relies on the cancellation of the large individual error terms, it can be very sensitive to the choice of r_0, as seen previously.

3.3.2 Attenuated MP2C and the optimal r_0 parameter

Next, we consider attenuated MP2C to reduce the correlation error. A similar energy decomposition can be performed for counterpoise-corrected MP2C in the aug-cc-pVTZ basis.

Finite basis set error:

$$E^{MP2C}(aTZ, CP) - E^{MP2C}(CBS, CP) \tag{3.13}$$

Correlation error:

$$E^{MP2C}(CBS, CP) - E^{CCSD(T)}(CBS, CP) \tag{3.14}$$

Attenuation error:

$$E^{attMP2C}(aTZ, CP) - E^{MP2C}(aTZ, CP) \tag{3.15}$$

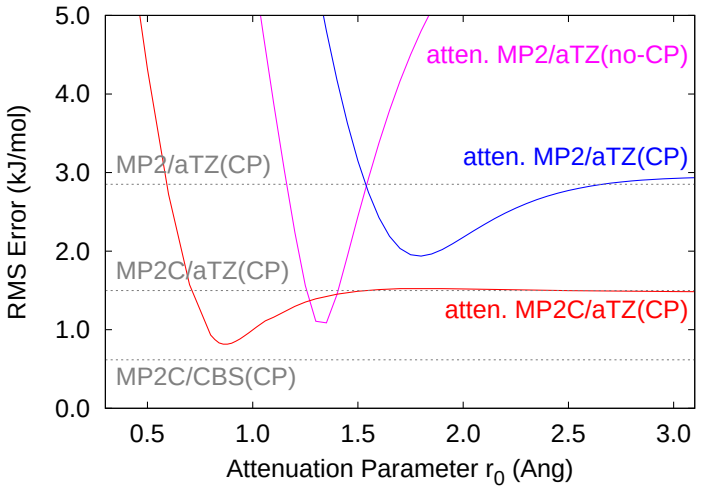

Figure 3.3: Root-mean-square errors (in kJ/mol) for attenuated MP2, attenuated MP2C, standard MP2, and standard MP2C relative to the CCSD(T) benchmarks for the S66 test set. Compared to attenuated MP2, attenuated MP2C achieves higher accuracy, is less sensitive to the value of r_0, and allows for more aggressive attenuation (a smaller r_0 value). Note that attenuated MP2 error does asymptote to the conventional MP2 result for large r_0 values beyond the range plotted here.

as shown in Figure 3.2c. The MP2C dispersion correction is already fairly well converged in the aug-cc-pVTZ basis [77], so the basis set error here is nearly identical to the basis set error for counterpoise-corrected attenuated MP2 (Figure 3.2b). However, the dispersion correction dramatically reduces the correlation error. The MP2C correlation error has no systematic bias and it is typically smaller than the basis set error. Because the basis set error dominates, the sum of these two contributions leads to systematic underbinding of the dimers, which is the opposite of the overbinding seen in the MP2 case.

We now seek an appropriate attenuation parameter r_0 to compensate for these errors in MP2C. Figure 3.3 plots RMS errors for the S66 test set (the same set was used to parameterize attenuated MP2) as a function of r_0. For values of $r_0 > 1.5$ Å, attenuated

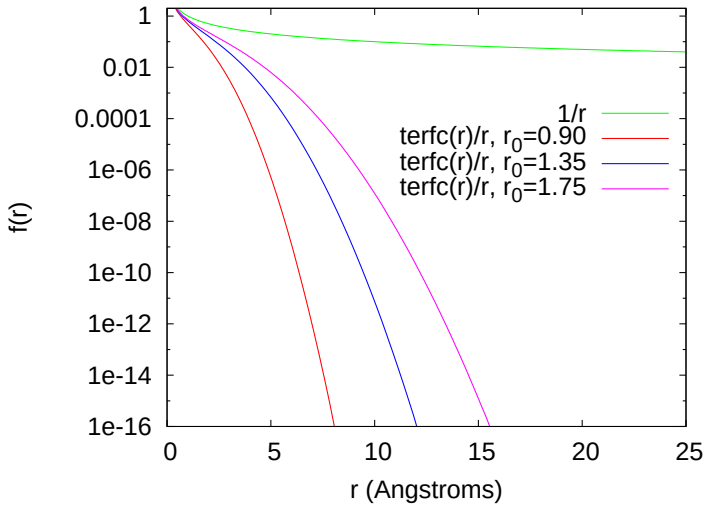

Figure 3.4: Attenuation with the optimal r_0 values for attenuated MP2 (no-CP), attenuated MP2 (CP) and attenuated MP2C dramatically reduces the range of the Coulomb operator, but the more aggressive attenuation ($r_0 = 0.9$ Å) possible for attenuated MP2C leads to a modified Coulomb operator that dies off completely by 7-8 Å.

MP2C results are very close to their nonattenuated counterparts. Decreasing r_0 further reduces the RMS errors until a minimum is reached around $r_0 = 0.9$ Å, with an RMS error of only 0.8 kJ/mol. For r_0 values smaller than 0.9 Å, the RMS errors increase once again.

The 0.8 kJ/mol RMS error for attenuated MP2C/aug-cc-pVTZ at the optimal r_0 value is roughly half the error of conventional MP2C in the same basis, and it approaches the 0.6 kJ/mol accuracy of CBS limit MP2C. The attenuated MP2C accuracy is somewhat higher than the smallest errors achieved by attenuated MP2. Moreover, the attenuated MP2C results are much less sensitive to the choice of the r_0 parameter than attenuated MP2.

Attenuated MP2C also allows for more aggressive Coulomb attenuation than attenuated MP2 ($r_0 = 0.9$ Å versus $r_0 = 1.35$ Å), which should lead to additional computa-

tional savings in an efficient implementation that exploits the sparsity of the two-electron integrals. As shown in Figure 3.4, the attenuated Coulomb operator dies off completely within 7-8 Å when $r_0 = 0.9$ Å, which will significantly increase the sparsity of the electron repulsion integrals compared to using larger r_0 values or the conventional Coulomb operator. Generally speaking, the number of electron repulsion integrals grows quartically with system size. Accounting for overlap sparsity reduces that growth to quadratic, while attenuation of the sort used here will reduce that growth to linear over rather short length scales. Some numerical investigations of this behavior have been reported previously using a slightly different $(\mathrm{erfc}(\omega r)/r)$ form of Coulomb attenuation [108].

Returning to Figure 3.2c, we observe that the error introduced by attenuating MP2C with $r_0 = 0.9$ Å does indeed largely cancel the finite basis and correlation errors to produce the overall high accuracy. However, the sign of the MP2C attenuation error is opposite to that of the MP2 attenuation error, indicating that different physics are involved.

To understand this difference, Figure 3.5 plots the RMS energy change between the attenuated and nonattenuated versions of MP2, MP2C and the UCHF dispersion component for the S66 set as a function of the attenuation parameter. The changes in the attenuated UCHF dispersion energies overlap with the changes in the attenuated MP2 results almost perfectly. In other words, for intermolecular interactions, attenuating medium- and long-range MP2 correlation is essentially equivalent to attenuating the UCHF dispersion. Only at shorter r_0 values, where exchange-dispersion effects become important, do the attenuated UCHF dispersion and attenuated MP2 curves diverge. Note that contributions like intermolecular polarization are handled at the HF level.

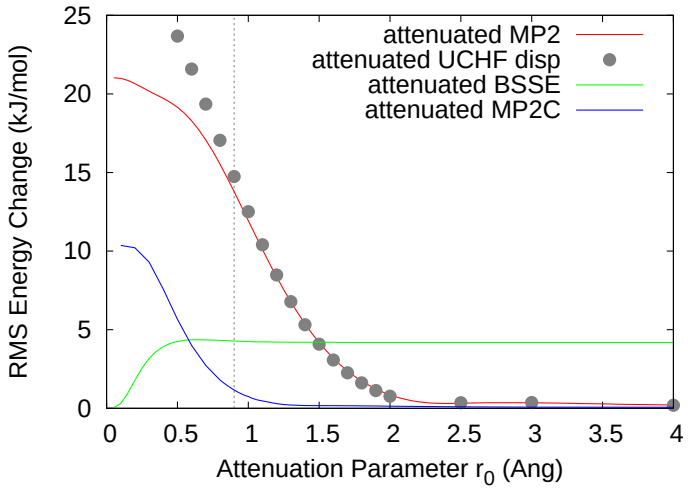

Figure 3.5: Root-mean-square energy change between the attenuated and nonattenuated MP2, UCHF dispersion, and MP2C models in the aug-cc-pVTZ basis. The RMS change in the MP2/aug-cc-pVTZ counterpoise correction due to attenuation is also shown ("attenuated BSSE"). The dashed vertical line indicates $r_0 = 0.9$ Å.

Subtracting the UCHF dispersion energy from the MP2 energy in MP2C, $E_{MP2} - E_{disp}^{UCHF}$, produces an intermolecular "dispersion-free" MP2 model, to which the CKS dispersion is subsequently added. With the UCHF dispersion removed, the intermolecular interactions are nearly independent of the attenuation parameter for $r_0 > 1.5$ Å, as indicated by the attenuated MP2C curve in Figure 3.5 (because the CKS dispersion contribution is not attenuated, it cancels when taking the difference between the attenuated and standard models in Figure 3.5). In other words, Coulomb attenuation eliminates interactions that are discarded anyway when computing the MP2C dispersion correction, which explains why attenuated MP2C gives virtually identical results to conventional MP2C for $r_0 > 1.5$ Å in Figure 3.3.

If one chooses to attenuate more aggressively using r_0 values below 1.5 Å, however,

short-range exchange-dispersion energy begins to be attenuated, as indicated by the rapid increase in the MP2C energy change for small r_0 (Figure 3.5). Unlike dispersion energies, exchange-dispersion interactions are repulsive, and attenuating them increases the strength of the intermolecular binding. This allows one to compensate for the finite basis errors in MP2C/aug-cc-pVTZ and achieve near CBS limit accuracy (Figure 3.3). It explains why the MP2C attenuation error has the opposite sign of the MP2 attenuation error in Figure 3.2, and this difference in sign is exactly what is needed to cancel the basis set error.

It is also important to recognize that one must use a counterpoise correction to obtain accurate results with attenuated MP2C. As shown in Figure 3.5, Coulomb attenuation does not significantly reduce the BSSE (as measured by the size of the counterpoise correction) until $r_0 < 0.5$ Å, which is well below the r_0 range for which accurate attenuated MP2 or MP2C results are obtained. If one omits the counterpoise correction, the combined basis set and correlation errors cause overbinding. One cannot compensate for this overbinding by attenuating the repulsive exchange-dispersion interaction — attenuating MP2C only binds the dimers more strongly. On the one hand, the need for a counterpoise correction does make attenuated MP2C more expensive than attenuated MP2, for which the counterpoise correction is neither necessary nor desirable. On the other hand, the magnitudes of the error cancellations involved in attenuated MP2C are much smaller, which reduces its sensitivity to the value of the parameter r_0 and makes the model more transferable to other systems.

In Chapter 2, we demonstrated that one can obtain significant computational savings by using a monomer-centered basis set to evaluate the dispersion correction instead of a dimer-centered basis set. The same idea also applies for attenuated MP2C. As shown

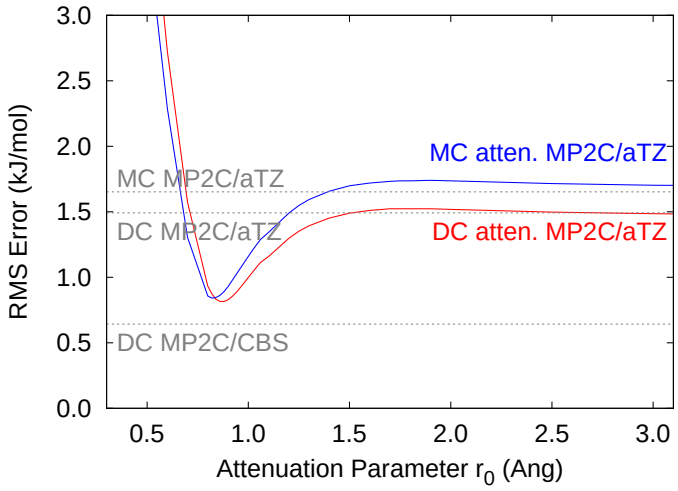

Figure 3.6: Difference between MC and DC attenuated MP2C/aug-cc-pVTZ for the S66 test set as a function of the attenuated parameter r_0. All results employ a counterpoise correction.

in Figure 3.6, using an MC basis instead of a DC one has only a small effect on the overall errors. Switching from a DC to an MC basis increases the RMS error for attenuated MP2C by only 0.1-0.2 kJ/mol on the S66 test set, which is comparable to what is observed for conventional MP2C. The optimal r_0 value for the MC case does shift slightly lower to 0.82 Å, but the difference in RMS error is small. Together, these results suggest that one can still use either an MC or DC basis set when computing the dispersion correction in attenuated MP2C.

3.3.3 Small basis set for short-range correlation

As we have observed in Figure 3.5, that the correlations energies are dominated by dispersion energy when $r_0 > 1.5$ Å. For attenuated MP2C, the minimum of RMS curve in Figure 3.3 is somewhat sensitive to the r_0 parameter (though it is less sensitive to r_0 than

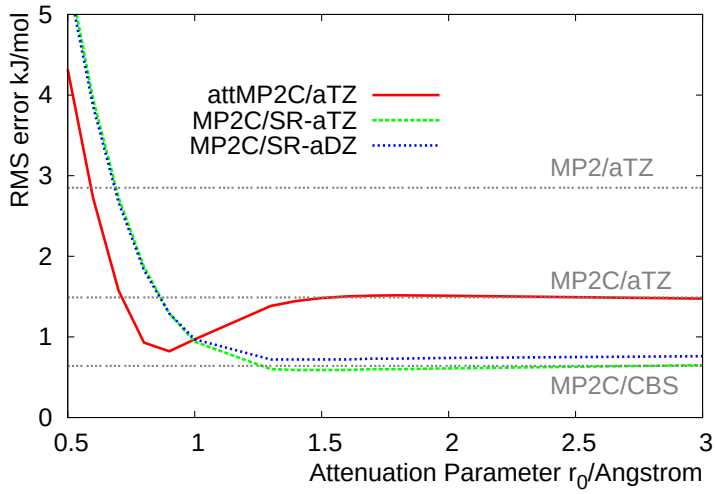

Figure 3.7: Root-mean-square errors (in kJ/mol) as functions of r_0 for attenuated MP2C/aTZ, MP2C/SR-aTZ, MP2C/SR-aDZ and standard MP2C with respect to CCSD(T) benchmarks on S66 test set.

attenuated MP2), and at large r_0 the curve converges to the RMS error of MP2C/aTZ, which exhibits undesirably large finite basis set error. Is it possible to reduce the sensitivity of the attenuated MP2C model to r_0 even further, without sacrificing accuracy? This would increase confidence in applying the method for systems beyond those used to parameterize r_0.

Accordingly, we consider the attenuation behavior of the MP2C/SR-aXZ model in which the short-range (SR) correlation is computed with a small basis set while the HF interaction energy and long-range correlation (dispersion) are extrapolated to CBS limit to reduce most of the finite basis set error. Figure 3.7 compares the RMS errors of these models as a function of the attenuation length-scale parameter r_0. As discussed in the previous section, when the r_0 value is greater than 1.3 Å, the attenuation error is small (see Figure 3.5) and attenuated MP2C gives results that closely match conventional MP2C/aTZ with sizable

finite basis set error. In contrast, the MP2C/SR-aDZ and MP2C/SR-aTZ models give results which are very close to those from MP2C/CBS. These models use a large basis where needed (for the long-range correlation), and a smaller basis where it is less important (short-range correlation). At $r_0 \approx 1.4$ Å, MP2C/SR-aTZ has an even smaller RMS error than MP2C/CBS because of some fortuitous cancellations. MP2C/SR-aDZ performs only ∼0.1 kJ/mol worse than the SR-aTZ case. Note that for the MP2C/SR-aXZ models, attenuating more aggressively ($r_0 < 1.4$ Å) increases the errors, unlike for attenuated MP2C/aTZ. Recall that more aggressive attenuation improves attenuated MP2C/aTZ by removing some exchange-dispersion to compensate for finite basis set errors. In the MP2C/SR-aXZ models, computing HF and the dispersion at the CBS limit eliminates most of the basis set errors, so no such cancellation can occur.

The fact that MP2/SR-aXZ models give excellent results that are insensitive to the r_0 parameters as long as $r_0 > 1.3$ Å has a couple advantages. First, this suggests that the optimal parameter should be transferable to other systems. Second, while our focus is mostly on intermolecular correlation, previous work has shown that intramolecular MP2 correlation behaves best with an attenuation parameter around 1.35 Å [103]. In other words, one can choose a value of r_0 that behaves well for intra- and intermolecular interactions simultaneously. Treating both types of interactions well at the same time is problematic for many affordable electronic structure methods, so the value of this feature should not be overlooked. On the other hand, computing the HF and long-range dispersion corrections in the CBS limit is substantially more expensive than the attenuated MP2C/aTZ model, as will be discussed later.

3.3.4 Error analysis of MP2C-F12/aDZ

The MP2C/SR-aXZ method discussed in the previous section improves the computational efficiency for short-range correlation calculation by using small basis set while introducing only small errors. On the other hand, it requires one to compute accurate dispersion in the CBS limit. Explicitly correlated MP2 (MP2-F12) provides yet another method for approaching the CBS limit with small basis sets [51, 105]. Explicit correlation improves the treatment of the electron-electron cusp condition, which means that fewer basis functions are needed to describe the wavefunctions and allows even relatively small basis sets to approximate the CBS limit. As in the case of conventional MP2, an MP2C-style dispersion correction can also be applied to MP2-F12 to improve the description of long-range correlations,

$$E_{int}^{MP2C\text{-}F12/aXZ} = E_{int}^{MP2\text{-}F12/aXZ} + E_{disp}^{\Delta MP2C} \tag{3.16}$$

where the MP2-F12 is computed with finite basis set aXZ, and MC algorithm for dispersion correction $E_{disp}^{\Delta MP2C}$can also be applied.

To understand the behavior of MP2C-F12, we first decompose the errors for the S66 benchmark set dimers into different sources analogously to what was done for the attenuated MP2C models in Section 3.3.2. The results are plotted in Figure 3.8. Note that the y-axis range in Figure 3.8 is significantly smaller than that in Figure 3.2. First, we consider the finite basis set error in MP2-F12/aDZ — both the HF energy and correlation energy are computed in the aDZ basis set. In Figure 3.8a, we can observe that these errors are cancelled out to some extent, particularly for the dimers (i.e. No.17 and No.20-No.24)

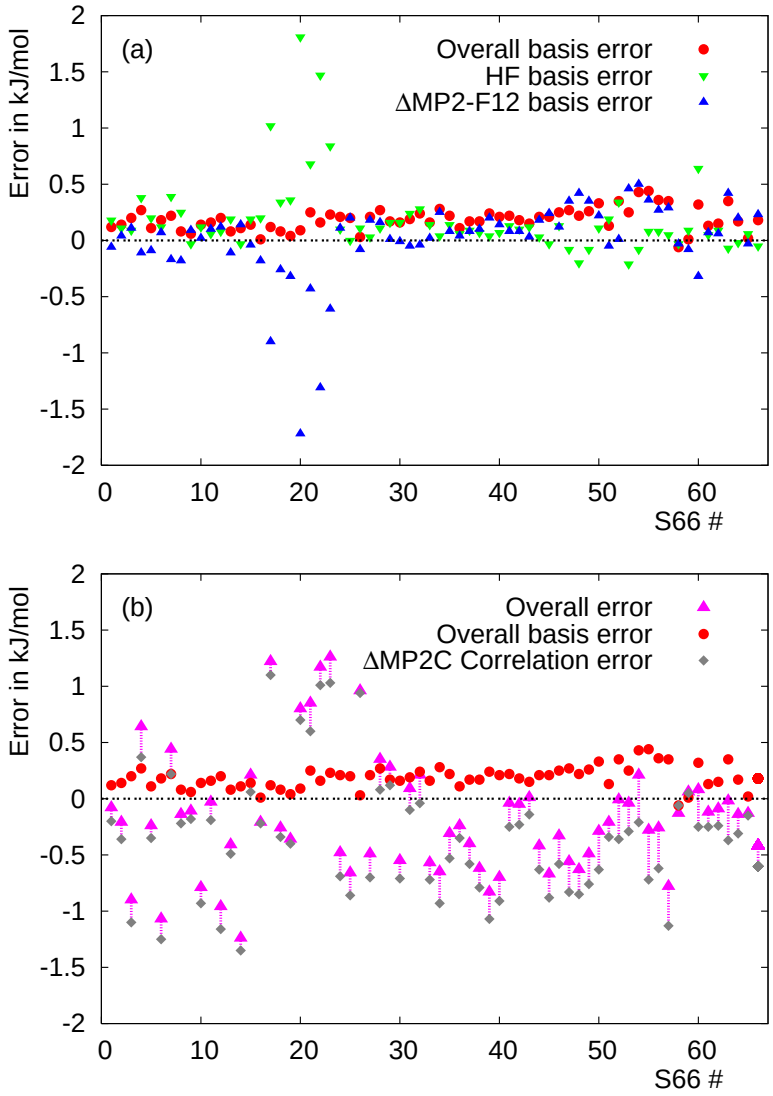

Figure 3.8: (a). HF basis error is the difference between HF/aDZ and HF/CBS, ΔMP2-F12 basis set error is the difference between MP2-F12/aDZ and MP2/CBS, and the overall basis error is the sum these two. (b). With the same overall basis error in (a), ΔMP2C correlation error is the difference between MP2C/CBS and CCSD(T)/CBS.

that exhibit large interaction energies. The net finite basis set errors lead to a modest underbinding. This underbinding will cancel some of the post-MP2C correlation errors (MP2C/CBS - CCSD(T)/CBS) which tend toward overbinding for most species in S66 test set (Figure 3.8b), producing further overall reduction in the RMS error.

Explicitly correlated models are somewhat more expensive than conventional algorithms. Indeed, these models were impractical until relatively recently, when density-fitting techniques which drastically reduce the cost of the integral evaluation became available. Even with density fitting techniques, MP2-F12 calculations are much slower than conventional RI-MP2.

To summarize, the RMS errors of different methods on S66 test set [45] are listed in the Table 3.1. Basis set superposition error (BSSE) is generally large but can be reduced by using a counterpoise correction. The RMS of MP2/aTZ is close to MP2/CBS due to some cancellations between underbinding finite basis set error and overbinding correlation error in UCHF dispersion. Attenuated MP2 without counterpoise correction balances between finite basis set error, attenuation error, and correlation error. The balance breaks and the RMS increases if BSSE is reduced with counterpoise correction. MP2C/CBS corrects most of the correlation errors in MP2/CBS and provides a promising method to approach CCSD(T) accuracy. If smaller basis set aTZ is used, nonnegligible finite basis set error will be introduced. With the attenuation approach, however, one can choose a suitable attenuation parameter that allows the finite basis set error be cancelled by the attractive attenuation error. MP2C/SR-aXZ only introduces finite basis set error in short-range correlation and the error is small enough to be ignored. MP2C-F12/aDZ offers an alternative

Table 3.1: Root-mean-square errors of different methods relative to CCSD(T)/CBS on S66 test set. The optimal r_0 values are used for attenuated methods.

method	basis	BSSE	optimal r_0 (Å)	RMS error (kJ/mol)
MP2	aTZ	no CP	—	6.4
MP2	aTZ	CP	—	2.9
attMP2	aTZ	CP	1.75	2.0
attMP2	aTZ	no CP	1.35	1.1
MP2	CBS limit	CP	—	3.1
MP2C	aTZ	no CP	—	4.1
MP2C	aTZ	CP	—	1.5
attMP2C	aTZ	CP	0.9	0.8
MP2C	CBS limit	CP	—	0.6
MP2C	CBS limit/SR-aTZ	CP	≥ 1.3	0.6
MP2C	CBS limit/SR-aDZ	CP	≥ 1.3	0.7
MP2C-F12	aTZ	CP	—	0.6
MP2C-F12	aDZ	CP	—	0.6

way to reduce finite basis set error with complicated algorithms and excellent error cancellations. One should note, however, that the computational expense of explicitly correlated MP2C-F12 grows somewhat more rapidly as a function of system size (in terms of both CPU time and RAM requirements) than non-F12 methods.

Having understood the basic features of these models and their performance on the S66 benchmark set of dimers, the next sections assess the transferability of these models by examining their performance on other systems. This is particularly important for the models involving an empirical r_0 attenuation parameter, since one wishes to confirm that this parameter is not overly dependent on the chemical system.

3.3.5 Transferability to other dimer interactions: Performance on the S22x5 benchmark set

For the methods involving basis set error or attenuation error in short-range correlation, their performance at different intermolecular separations should be examined because exchange-dispersion strongly depends on intermolecular distance, as shown in the potential energy curve (Figure 1.3). Moreover, the attenuated models are optimized on S66 test set, which means that their performance on other systems and for non-equilibrium geometries is unclear. Thus it is necessary to examine the reliability of different models across the potential energy surface.

The S22x5 test set [44] is created by scaling the equilibrium intermolecular separation $1.0R_e$ to four additional intermolecular distances $0.9R_e$, $1.2R_e$, $1.5R_e$ and $2.0R_e$. In Table 3.2, we can see MP2/CBS gives large errors across the potential energy surface because of the large correlation errors in UCHF dispersion. MP2C/aTZ improves the correlation error but still performs relatively poorly due to the finite basis set error. For attenuated MP2, the one without counterpoise correction performs better than the counterpoise corrected model at the equilibrium distance $1.0R_e$ due to fortuitous error cancellations. On the other hand, it produces much larger errors at $0.9R_e$. The physics of the non-CP attenuated MP2 model is wrong because it completely omits long-range correlation. Moving toward shorter separations than it was parameterized for exposes its limited transferability due to its strong dependence on the value of attenuation parameter r_0.

The attenuated MP2C models properly restore the long-range correlation, which substantially improves their transferability. Attenuated MP2C/aTZ performs just slightly

Table 3.2: RMS error (kJ/mol) of different methods on S22x5 test set. R_e represents the equilibrium molecular separation.

methods	basis	$0.9R_e$	$1.0R_e$	$1.2R_e$	$1.5R_e$	$2.0R_e$
MP2	CBS	8.81	5.72	2.24	0.79	0.23
attMP2	aTZ, no CP	4.17	1.98	1.04	0.94	0.46
attMP2	aTZ	3.95	3.01	1.72	1.50	0.83
MP2C	aTZ	3.87	2.24	0.94	0.39	0.19
attMP2C	aTZ	1.81	1.03	0.56	0.31	0.11
MP2C	CBS	1.51	0.73	0.52	0.25	0.13
MP2C	CBS/SR-aTZ	1.44	0.95	0.53	0.31	0.19
MP2C	CBS/SR-aDZ	1.37	1.02	0.59	0.33	0.19
MP2C-F12	aTZ	1.57	0.69	0.51	0.23	0.12
MP2C-F12	aDZ	1.76	1.01	0.55	0.29	0.14

worse than MP2C/CBS. It improves a lot on conventional MP2C/aTZ by cancelling the basis set errors, especially for $0.9R_e$ geometries where the basis set error is larger than the other separations. For MP2C/SR-aXZ model, the basis set error in short-range correlation is comparatively small across the potential energy surface that the RMS errors are just slightly higher except for $0.9R_0$ distance, surprisingly. This could be explained with some accident cancellations between basis set error (in short-range correlation) and correlation error in MP2C (MP2C- CCSD(T)). The MP2C-F12/aDZ method which exhibits good performance on S66 test set, seems to be not as good as MP2C/CBS on S22x5 set. But still, it reduces most finite basis set errors compared to MP2C/aTZ, systematically. When aTZ basis set is for MP2C-F12, the RMS errors are further reduced to approaching MP2C/CBS.

After this round of examination, attenuated MP2C/aTZ, MP2C/SR-aXZ and MP2C-F12/aXZ methods stand out from the candidates. To further test the reliabilities of these promising methods, molecular crystals will be used as the touchstone because they contains a variety of molecular separations and orientations. In the next section, the

methods will be used to predict lattice energies of seven small molecule crystals. A challenging "final exam" will be taken on ordering the lattice energies of five polymorphs of oxalyl dihydrazide.

3.3.6 Reliability of the models for molecular crystal lattice energies

Table 3.3 reports the lattice energies for the seven molecular crystals [7, 9], which were computed using the hybrid many-body interaction (HMBI) approach with the same geometries used in last chapter. The *ab initio* force field (AIFF) parameters were computed in the aug-cc-pVTZ basis. All the methods listed in the table were used to evaluate the interaction energies between pairs of molecules (two-body terms) in the fragment approach. One-body energy terms were computed uniformly in the CBS limit so that we can just focus on the performance of different treatments of the two-body contributions to the lattice energy.

Conventional MP2 in the CBS limit predicts lattice energies within a few kJ/mol of benchmark CCSD(T) results for crystals where dispersion is not important [9]. In crystals such as benzene or imidazole where dispersion interactions play a significant role, however, MP2 overestimates the lattice energy by 10-15 kJ/mol (∼15-20%). For the seven crystals considered here, it exhibits a root-mean-square error of 6.6 kJ/mol with respect to CCSD(T). Attenuated MP2 with optimal parameter r_0 cancels the correlation error to some extent, especially for benzene and imidazole crystals. Similar to the benchmark dimer case, the RMS error (3.6 kJ/mol) of lattice energies computed with attenuated MP2 is also smaller than the RMS with conventional MP2, although the results for some induction dom-

Table 3.3: Lattice energies (kJ/mol) of seven small-molecule crystals computed with fragment-based hybrid many-body interaction (HMBI) approach. QM one-body energies are computed in the CBS limit, and two-body interactions are computed with the methods listed.

	MP2 CBS	attMP2 aTZ	MP2C CBS	attMP2C aTZ	MP2C-F12 aDZ	MP2C SR-aTZ	MP2C SR-aDZ	CCSD(T) CBS	Experiment
Ice	59.8	61.5	60.5	58.1	59.4	59.7	60.0	60.1	59
Formamide	78.7	75.0	79.5	78.7	78.4	78.5	78.4	80.5	82±0.3
Acetamide	80.6	76.4	80.6	79.4	79.9	81.4	81.2	80.4	86±2
Imidazole	102.0	89.0	92.6	92.3	92.0	91.7	92.0	88.6	91±4
Benzene	61.7	45.9	49.0	48.6	47.9	48.7	48.0	50.9	52±3
NH_3	39.4	38.9	41.1	41.7	40.8	40.8	40.8	40.9	39
CO_2	31.0	27.9	27.7	27.0	28.4	27.4	27.7	31.3	31
RMS error	6.6	3.6	2.2	2.6	2.2	2.2	2.3		

inant crystals become worse. As demonstrated in last chapter, CBS limit MP2C performs much better in predicting the lattice energies of these crystals, with an RMS error of 2.2 kJ/mol.

The fast MP2C methods (attMP2C/aTZ, MP2C-F12/aDZ and MP2C/SR-aXZ) seek to approach the MP2C/CBS accuracy by reducing the finite basis set errors in different ways. Compared to CCSD(T), they have almost the same RMS errors (2.2-2.6 kJ/mol) as MP2C/CBS (2.2 kJ/mol). If we compare the methods to MP2C/CBS to see how well they reduce the finite basis set error, much smaller RMS errors are obtained, with 1.1 kJ/mol for attMP2C/aTZ, 0.8 kJ/mol for MP2C-F12/aDZ, and 0.7 kJ/mol for MP2C/SR-aXZ. Unlike other fast MP2C methods, attenuated MP2C/aTZ has two error sources (finite basis set error and attenuation error), and the error cancellations are more difficult to control. Comparatively, attMP2C/aTZ is less reliable than the others, with a slightly larger RMS error. Overall, the lattice energy predictions prove that all these three models reduce most of the basis set error, and the main error source left is the post-MP2C correlation error.

Finally, we apply these methods to order the lattice energies of five polymorphs from oxalyl dihydrazide, which have proved challenging for density functional methods and many inexpensive wavefunctional methods [50]. MP2C/CBS predicts the relative energy order successfully, as discussed in last chapter. All five polymorphs are predicted to lie within a 4 kJ/mol range. The other fast MP2C methods also predict the same energy ordering as MP2C/CBS, with some tiny difference in energy gaps as shown in Figure 3.9. For the energy gaps between different phases, all the methods predict similar gaps for $\epsilon \rightarrow \alpha$ and $\delta \rightarrow \gamma$. While for $\alpha \rightarrow \delta$ and $\gamma \rightarrow \beta$, energy gaps predicted by different

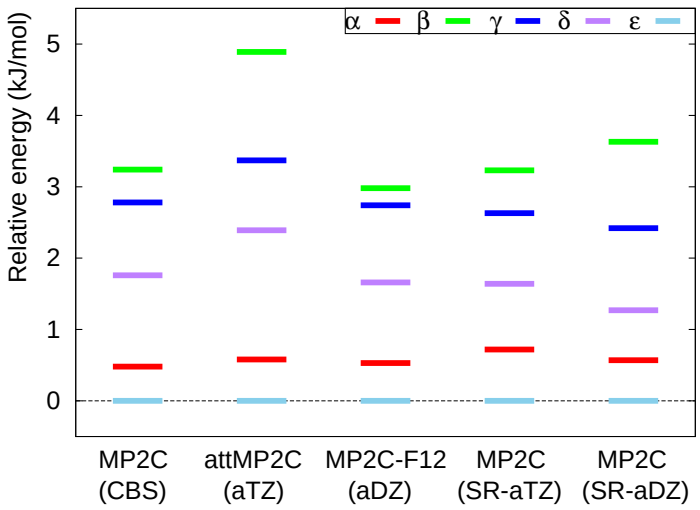

Figure 3.9: Relative polymorph energies for oxalyl dihydrazide computed with MP2C/CBS, attenuated MP2C/aTZ, MP2C-F12/aDZ and MP2C/SR-aDZ methods. All the fast MP2C methods are able to predict the same relative ordering as MP2C/CBS.

methods are not in consistent. Among them, MP2C-F12 and MP2C/SR-aTZ are close enough to MP2C/CBS, indicating only a minor and controllable basis set error introduced. For the whole energy gap, MP2C-F12 predicts the smallest gap, then is MP2C/SR-aTZ which has smaller basis set errors than MP2C/SR-aDZ, and attMP2C/aTZ predicts the largest gap up to 5 kJ/mol. The absolute lattice energy of the reference ϵ phase also shows that attMP2C/aTZ is worse with 190 kJ/mol, far from the others with 181 kJ/mol. From the test, we can conclude that although attMP2C/aTZ predicts the same order of relative energy, it is not as reliable as the other methods because it depends on the cancellations between uncontrollable errors.

3.4 Conclusions

In this chapter, several methods to reduce the finite basis set error and improve the computational efficiency of MP2C interaction energies have been evaluated on different test sets. First, we extended the idea of Coulomb attenuation of the correlation energy using a double error function to MP2C for the treatment of intermolecular interactions. Second, we use a small basis set for short-range correlation and a complete basis set limit for the other energy components. Third, we compare these new models against the existing MP2C-F12 method.

For the attenuated methods, the attenuation strength parameter r_0 was optimized on the S66 test set. The models were then tested on the S22x5 benchmark set and on molecular crystal lattice energies. Several key conclusions regarding the methods have emerged in the process:

(1) Attenuated MP2 compensates for intermolecular overbinding that typically arises from the combination of finite-basis error and missing higher-order correlation effects in MP2 by eliminating an appropriate fraction of the attractive UCHF dispersion. The magnitude of error cancellation involved is fairly large relative to the strengths of the interactions involved.

(2) MP2C eliminates the UCHF dispersion and replaces it with more accurate CKS dispersion. One can therefore attenuate away the long- and medium-range Coulomb interactions in MP2 (i.e, $r_0 \sim 1.5$ Å) with only a small impact on the MP2C interaction energies.

(3) Counterpoise-corrected MP2C/aug-cc-pVTZ typically underbinds dimers due

to the finite-basis set error. Aggressive Coulomb attenuation ($r_0 = 0.9$ Å) allows one to approximately compensate for this underbinding by removing some of the repulsive exchange-dispersion. This enables one to approach CBS limit results using only an aug-cc-pVTZ basis. What is more, one can use either a monomer-centered or dimer-centered basis set for the dispersion correction, with dimer-centered basis set providing marginally better performance. Even though the current implementation of attenuated MP2C does not exploit the increased integral sparsity, it still achieves substantial speed-ups by avoiding the need for larger basis sets. For example, for a conventional two-point TZ-QZ basis set extrapolation to the CBS limit, performing counterpoise-corrected attenuated MP2C interaction energy calculations for the stacked benzene dimer from the S22 test set required about 11 hours (aug-cc-pVTZ) and 43 hours (aug-cc-pVQZ) on a single core of a 2.3 Ghz Intel Xeon E5-2630 processor with 4GB of RAM. For the attenuated version, one can avoid the aug-cc-pVQZ basis calculation entirely, thereby reducing the computational time by 80%.

(4) MP2C/SR-aXZ requires complete-basis set for long-range correlation (dispersion) which contributes to most of the correlation energy, while allows one to compute short-range correlation in small basis set. With finite basis set error in dispersion substantially reduced, this model is more reliable than attenuated MP2C and more flexible in choosing an attenuation strength only if $r_0 \geq 1.3$ Å. MP2C/SR-aXZ needs dimer-centered dispersion computed with TZ-QZ extrapolation to CBS limit, requiring molecular orbitals computed in the aug-cc-pVQZ basis set. Practical timing reveals that although Hartree-Fock scales as only $O(N^3)$, the large pre-factor means that it forms the main computational bottleneck in the sorts of systems considered here. To make this approach more practi-

cal, one needs to improve the computational efficiency of the Hartree-Fock portion of the calculation.

(5) MP2C-F12/aDZ uses an explicitly correlated algorithm for MP2 correlation calculation. Due to excellent fortuitous cancellations between basis set errors in the Hartree-Fock and correlation contributions, even the small aug-cc-pVDZ basis set can be used to approach the accuracy of CBS limit. Consider the same benzene dimer and computational environment as before, it only needs 1.7 hours for MP2-F12/aDZ calculation. However, although MP2-F12 scales $O(N^5)$ like MP2 and MP2C, it has a much larger prefactor and will become more computationally expensive in terms of both CPU time and RAM space.

(6) All of these fast MP2C methods reduce correlation errors and basis set errors significantly. They perform uniformly well across large regions of the potential energy surface, predict lattice energies of molecular crystals fairly close to conventional MP2C/CBS, and order relative lattice energies of polymorphs from oxalyl dihydrazide successfully. Among them, attenuated MP2C/aTZ gives the largest errors due to its reliance on imperfect error cancellations between finite basis set errors and short-range exchange dispersion. However, it is also much less computationally demanding than the MP2C/SR-aXZ models.

Chapter 4

MP2+ATM model for three body interactions

4.1 Introduction

In recent years, our ability to compute the intermolecular interactions occurring between a pair of molecules has improved dramatically, with major advances in density functional theory (DFT), wave function models, and symmetry adapted perturbation theory (SAPT) [1,2,24,25,28]. The approaches discussed in previous two chapters significantly improves the computational efficiency without sacrificing high accuracy for two-body interaction energy prediction. The treatment of many-body intermolecular interactions, however, remains challenging, since it requires a balanced treatment of polarization, exchange, and dispersion interactions. While fragment-based electronic structure methods facilitate the treatment of complex systems [4,5], computationally efficient approaches for the many-body

intermolecular interactions in individual fragments are needed to make accurate *ab initio* calculations in condensed-phase systems routine.

DFT is often used to model condensed-phase systems, but tests on our 3B-69 benchmark set for three-body intermolecular interactions exposed weaknesses in widely used dispersion-corrected DFT approximations — many density functionals poorly reproduce benchmark coupled cluster singles, doubles, and perturbative triples (CCSD(T)) three-body intermolecular interaction energies [109]. The problem largely stems from delocalization error [110] and exchange error, which is not corrected by adding an Axilrod-Teller-Muto (ATM) three-body dispersion term [111, 112, 109, 113].

Wavefunction methods provide a potential alternative to DFT for modeling these many-body interactions. Natural inclusion of three-body intermolecular dispersion occurs at third order in the supermolecular Møller-Plesset perturbation series [114], and methods like MP2.5 (the average of MP2 and MP3) [33, 115] and spin-component-scaled CCSD [52] for molecular interactions (SCS(MI)-CCSD) provide balanced descriptions of both polarization and dispersion and reproduce CCSD(T) benchmarks faithfully. Unfortunately, their relatively high $O(N^6)$ computational cost with system size N can be prohibitive in condensed-phase applications because the large amount of three-body interactions need to be included.

On the other hand, MP2 provides an excellent description of polarization with only $O(N^5)$ computational cost [109], albeit with no intermolecular three-body dispersion terms [116]. Here, we show that the combination of MP2 with an ATM dispersion term provides an excellent balance between computational cost and accuracy across a range of chemical systems. The dispersion coefficients used to parameterize the ATM dispersion

energies here are non-empirical, though short-range empirical damping is required to obtain good results. We are not the first to propose adding a three-body dispersion correction to MP2 [117–119] but we demonstrate that the $O(N^5)$ cost MP2 + damped ATM approach gives accurate results that are only moderately worse than those from the much more expensive MP2.5 or SCS(MI)-CCSD models.

4.2 Theory

4.2.1 MP2+ATM model for three body interactions

The three-body interaction energies ΔE^3 in supermolecular approach are defined as a difference between the total energy of the trimer and a sum of both monomer energies and all two-body interaction energies ΔE^2 between them. For a trimer consisting of units A, B and C, the three-body energy is:

$$\Delta E^3(ABC) = E(ABC) - E(A) - E(B) - E(C) - \Delta E^2(AB) - \Delta E^2(BC) - \Delta E^2(CA) \quad (4.1)$$

Expanding the interaction energies in terms of monomer and dimer energies, e.g. $\Delta E^2(AB) = E(AB) - E(A) - E(B)$, the three-body energy can be expressed in terms of monomer, dimer, and trimer energies as:

$$\Delta E^3(ABC) = E(ABC) - E(AB) - E(BC) - E(AC) + E(A) + E(B) + E(C) \quad (4.2)$$

To eliminate the basis set superposition error (BSSE) in trimer calculations, an appropriate counterpoise correction should be used. Different methods like DFT, MP2 and CCSD(T) can be applied to the expression above to obtain the corresponding three-body interaction energies.

To improve the correlation in three-body interaction energies computed with MP2, we propose an intermolecular Axilrod-Teller-Muto (ATM) triple-dipole dispersion [111,112] contribution for molecular trimer ABC, which is computed as:

$$E_{ATM} = \sum_{a\in A}\sum_{b\in B}\sum_{c\in C} f_9^{abc} C_9^{abc} \frac{\left(1 + 3cos\hat{a}cos\hat{b}cos\hat{c}\right)}{R_{ab}^3 R_{bc}^3 R_{ac}^3} \tag{4.3}$$

where C_9^{abc} is the dispersion coefficient for atom triplet abc, R_{ij} is the distance between atoms i and j. $\hat{a}$, $\hat{b}$, $\hat{c}$ are the angles of the triangle formed by the three atoms, and f_9 is a distance-dependent damping function that turns off the 3-body dispersion at short-range. The isotropic C_9 dispersion coefficients are obtained from distributed frequency-dependent dipole-dipole polarizabilities $\alpha_{11}(\omega)$ which are calculated from coupled Kohn-Sham (CKS) theory via Casimir-Polder integration [14, 120],

$$C_9^{abc} = \frac{3}{\pi}\int_0^\infty d\omega \alpha_{11}^a(\omega)\alpha_{11}^b(\omega)\alpha_{11}^c(\omega) \tag{4.4}$$

The frequency-dependent polarizabilities were calculated using asymptotically corrected PBE0 [121] with the hybrid ALDAX/coupled Hartree-Fock kernel and the aug-cc-pVTZ basis, as implemented in CamCASP [122]. This approach provides accurate dispersion coefficients [9, 120], but a variety of other approaches for obtaining C_9 coefficients also exist [123–128].

Short-range damping proves critical to the model. The damping function f_9 is written as a product of three 2-body Tang-Toennies damping functions [124, 129–131],

$$f_9^{abc} = f_6^{ab}(R_{ab}) f_6^{bc}(R_{bc}) f_6^{ac}(R_{ac}) \tag{4.5}$$

$$f_6(R,\beta) = 1 - \sum_{k=0}^{n}\left(\frac{(\beta R)^k}{k!}\right) e^{-\beta R} \tag{4.6}$$

where β depends linearly on the sum of typical CCSD atomic van der Waals radii r^{vdW} of atoms a and b [124],

$$\beta = -0.31(r_a^{vdW} + r_b^{vdW}) + 3.43 \text{ Bohr}^{-1} \tag{4.7}$$

Representative CCSD van der Waals radii of 2.63, 3.34, 3.18, and 3.07 Bohr [124] are used for hydrogen, carbon, nitrogen, and oxygen, respectively. No effort was made to optimize for these parameters the systems considered here.

Alternatively, one might use the non-expanded form of this third-order dispersion energy calculated directly from the molecular SAPT CKS propagators, as in the MP2+SDFT model [117] for example. Here, the density-fitted 3-body CKS dispersion is evaluated as,

$$E_{CKS}^{(3)} = -\frac{1}{\pi}\int_0^\infty d\omega Tr\left[\chi^A(\omega)J^{AB}\chi^B(\omega)J^{BC}\chi^C(\omega)J^{CA}\right] \tag{4.8}$$

where the χ are the CKS propagators and the integrals $J^{AB} = \langle P^A|r_{AB}^{-1}|Q^B\rangle$ are computed between auxiliary basis functions on each monomer as we discussed in last two chapters. The response functions $\chi(\omega)$ were computed in a monomer-centered aug-cc-pVTZ basis at ten frequencies using CamCASP.

4.3 Results and discussion

4.3.1 Performance on 3B-69 benchmark trimer set

We first test these models on the recently developed 3B-69 benchmark test set for three-body intermolecular interactions. This set consists of three trimers extracted from each of 23 different molecular crystals. The 69 trimers in this set exhibit a wide variety

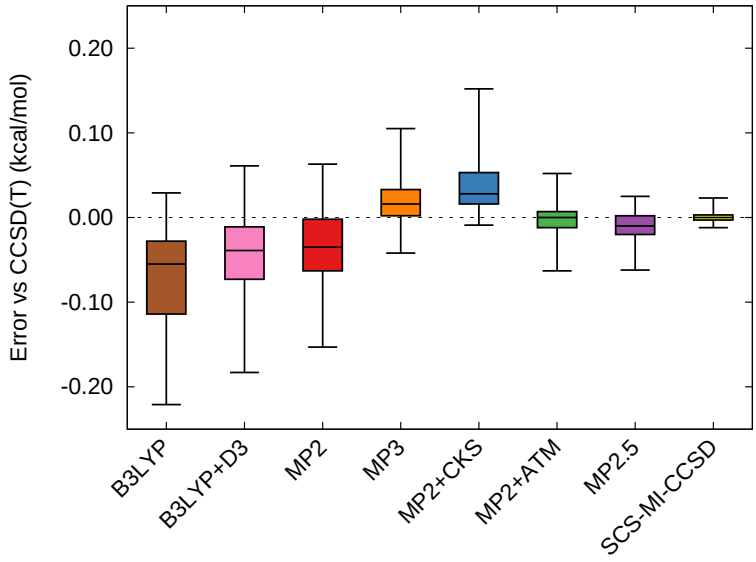

Figure 4.1: Error distribution for various methods on the 3B-69 trimer benchmark test set.

of intermolecular interactions [109], both within the three different geometric arrangements of a given species and across the 23 different species. One trimer might involve favorable three-body polarization, another unfavorable three-body polarization, and the third with polarization and a strongly repulsive three-body dispersion. To perform well across the entire benchmark set, a model must describe each of the various individual types of physical interactions correctly.

Many popular density functional approaches, like B3LYP [109] perform poorly on this test set, even when three-body ATM dispersion corrections like D3 or XDM are applied as Figure 4.1 shows. This poor performance reflects problems with both the treatment of many-body polarization in these functionals and apparent underestimation of the dispersion corrections. MP2 performs moderately better than the DFT models, but it lacks three-body dispersion. Models like MP2.5 and especially SCS(MI)-CCSD perform much better,

reproducing the CCSD(T) benchmark results with high, uniform accuracy.

Figure 4.2 plots the three-body error for each trimer against the estimated three-body dispersion, as computed using the damped ATM model described above. If (1) the ATM model provided a perfect description for the three-body dispersion and (2) that was the only error in the calculations, the errors in these plots would exhibit a slope of -1 and a root-mean-square residual error (RMSE) from the best fit line of 0. Obviously neither criteria is precisely met, but comparison of the MP2 and SCS(MI)-CCSD models indicates that these assumptions are approximately correct. MP2 exhibits a good fit to a line with slope of -0.95, indicating that it describes polarization well but omits three-body dispersion. In contrast, a model correctly describing three-body dispersion would ideally exhibit zero slope and RMSE. Indeed, SCS(MI)-CCSD performs very well on the test set and exhibits slopes and RMS very near to zero in Figure 4.2.

Accordingly, one can improve MP2 by adding the damped ATM three-body dispersion to the model. The slope of the MP2+ATM dispersion in Figure 4.2 is nearly zero (by construction), just as for SCS(MI)-CCSD, though it retains the moderately large RMS residual error of MP2. Figure 4.1 demonstrates that the overall error distribution in the MP2+ATM model is very similar to that of MP2.5, despite its lower N^5 cost. In contrast, both MP3 and SAPT CKS 3-body dispersion overestimate the post-MP2 correlation, producing results that are somewhat worse than the MP2+ATM model. The overestimation is worst in the cases where the dispersion energy is largest, as reflected in the positive slope of +0.37 in Figure 4.2.

Comparison between SAPT and supermolecular MPn models provide valuable in-

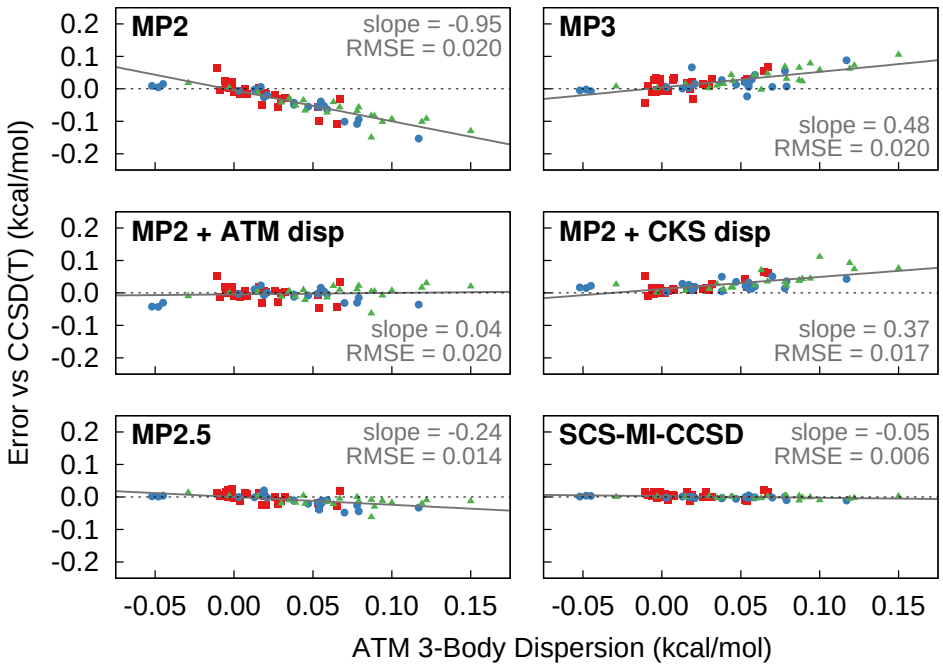

Figure 4.2: Errors in the 3B-69 trimer 3-body energies plotted against the damped ATM 3-body dispersion. The slope of the best-fit line and the RMS residual errors (kJ/mol) from the best fit are given in gray. Red, blue, and green symbols refer to low, medium, and high dispersion structures, respectively.

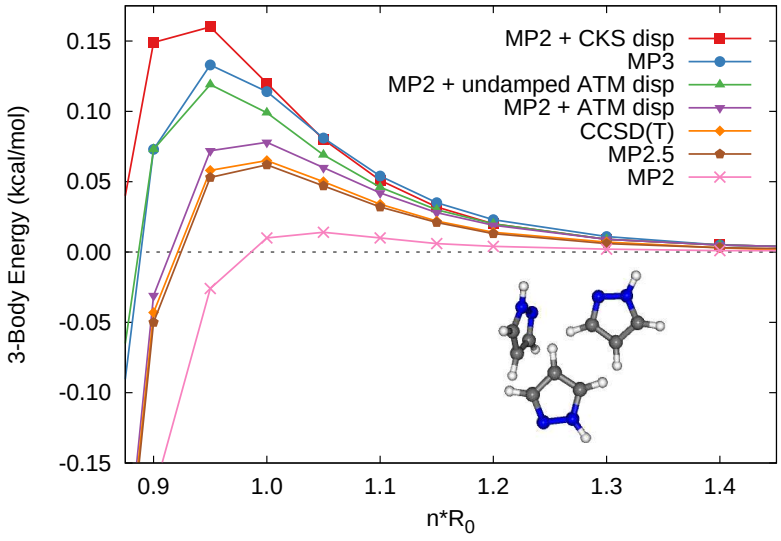

Figure 4.3: 3-body energy for a pyrazole trimer versus the trimer separation

sight [114,117,130]. MP2 lacks not only the (usually repulsive) three-body dispersion which first appears in MP3, but it also neglects a variety of short-range non-additive exchange terms and fourth-order and higher contributions to the non-additive dispersion that are net attractive. Augmenting MP2 with the CKS dispersion term captures one important post-MP2 correction, but neglecting the other 3-body correlation terms causes the model to overestimate the net 3-body contribution, particularly at shorter distance.

Figure 4.3 plots the 3-body energy for a pyrazole trimer (3B-69 structure 12b) as the center-of-mass intermolecular separation is scaled isotropically. At equilibrium distances and beyond, the MP2+CKS model tracks MP3 closely. At shorter separations, however, MP2+CKS lacks the attractive non-additive exchange terms found in MP3, causing it to be too repulsive. An undamped version of MP2+ATM dispersion behaves similarly, though it underestimates the CKS dispersion somewhat. This underestimation likely stems from the

neglect of higher multipolar contributions beyond the triple-dipole dispersion in the ATM model. Applying short-range damping then brings MP2+ATM much closer to CCSD(T) or MP2.5.

4.3.2 Performance on benzene crystal

To test the efficacy of the MP2+ATM model further, we consider the 65 symmetry-unique trimers from the benzene crystal for which benchmark CCSD(T) energies exist [119]. Figure 4.4 demonstrates once again that MP2 exhibits significant errors due to the lack of 3-body dispersion. However, correcting the MP2 dispersion leads to results that approach SCS(MI)-CCSD accuracy. As before, the damped ATM dispersion correction performs much better than the undamped CKS dispersion correction, particularly for the densely packed trimers which have the largest 3-body dispersion, reiterating the importance of short-range damping.

Finally, to put these results in perspective, we consider how accurately one can predict the crystal lattice energy of benzene. State-of-the-art coupled cluster calculations can achieve sub-kJ/mol accuracy [132], but such high-level calculations are infeasible for more complex organic crystals. More practically, what can one achieve using relatively inexpensive MP2-like methods instead of CCSD(T)?

Table 4.1 compares the lattice energy contributions obtained by combining complete basis set (CBS) limit dispersion-corrected MP2C 2-body dimer calculations and the 3-body MP2+ATM dispersion model discussed here against estimated CBS-limit CCSD(T) benchmarks using the structures reported by the Sherrill group [119,133]. Four-body effects

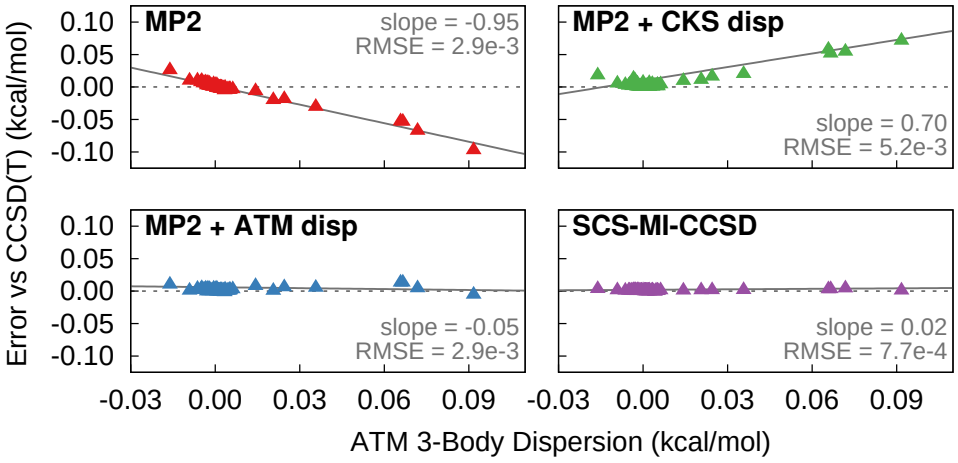

Figure 4.4: Errors in the benzene trimer 3-body energies plotted against the size of the damped ATM dispersion correction

contribute only ∼0.1 kcal/mol in benzene [132] and are neglected here.

MP2C reproduces the sum of the CCSD(T) pairwise interactions to within half a kcal/mol (4%). Replacing the CCSD(T) treatment of short-range trimers with the MP2+ATM model introduces a proportionately large error of 0.18 kcal/mol (20%), but it provides massive computational savings. The two- and three-body errors fortuitously cancel here, producing a net error of only 3%. In contrast, conventional MP2 errors are several-fold larger. In other words, one can model the important intermolecular interac-

Table 4.1: Contributions to the benzene lattice energy (kcal/mol), using the geometry from [117]. In order to be consistent with the convention of a positive lattice energy, the signs of energy contributions are opposite (i.e. positive sign means attractive contribution and negative sign represents repulsive contribution).

	MP2	MP2C/MP2+ATM	CCSD(T)
2-body	16.15	13.64	13.15
3-body	-0.13	-1.07	-0.89
Sum	16.02	12.57	12.26

tions in molecular crystals and other systems very effectively using dispersion-corrected MP2 models at only $O(N^5)$ computational cost.

4.4 Conclusions

To conclude, three-body interactions are more complicated than two-body interactions because one needs to consider the perturbation from a third molecule on the two-body system. DFT-D methods generally have difficulty describing this complicated level of long-range correlation. MP2 lacks enough correlation description on the associative many-body density fluctuation response. Propagators associated three-body CKS dispersion can be used to describe such response effect. This MP2+CKS dispersion (or MP2+SDFT) model slightly improves the conventional MP2, but still not accurate enough due to the missing short-range exchange and higher level of correlation terms. Using multipole expansion (Axilrod-Teller-Muto) model with Tang-Toennies damping function to approximate three-body CKS dispersion proves to be an improved correlation correction on MP2. The examinations on 3B-69 benchmark trimer set, pyrazole potential energy surface and benzene crystal shows that MP2+ATM model is reliable enough for widely application. This simple model with excellent performance stems from the use of damping function that dramatically reduces the correlation errors.

Not just accuracy, efficiency is also achieved at the same time. Three-body interactions with MP2 have been proved to be insensitive to basis set that even aug-cc-pVDZ basis set could be able to achieve the accuracy of CBS limit [119, 134]. What is more, if we consider Coulomb attenuation in the three-body MP2 correlation, in the long-range correla-

tion, the attenuation error is small and changes slowly when r_0 increases. For example, the attenuation RMS error on 3B-69 benchmark trimer set is 0.003 kJ/mol when $r_0 = 4.0$ Å, it increases to 0.005 kJ/mol when $r_0 = 3.0$ Å, 0.010 kJ/mol when $r_0 = 2.0$ Å, and reaches to 0.03 kJ/mol with more aggressive attenuation $r_0 = 1.5$ Å. This attenuation behavior allows us to utilize Coulomb attenuation for MP2 correlation calculation and exploit sparsity to further improve the computational speed.

Chapter 5

Conclusions

In molecular crystal lattice energy predictions with fragment based approach, two- and three-body terms are the main contributions to the total lattice energy. They should be treated with reliable quantum mechanical methods if high accuracy is required. On the other hand, the number of QM terms computed in fragment based models (HMBI model, for example) is generally large, and evaluating these terms is the computational bottleneck. One must improve the efficiency and accuracy with which two- and three-body interactions can be described in order to study chemically interesting molecular crystals.

A conflict exists in that to achieve high accuracy, one should use more complicated model chemistries like CCSD(T) to describe the electron correlation, but such methods are usually cost prohibitive for practical applications. Accordingly, finding new ways to improve the accuracy of less computationally demanding methods has received much attention. DFT-D methods are widely used because of their low $O(N^3)$ scaling with system size and their reliability in describing long-range correlations. If even higher accuracy is required,

one should turn to MP2 based methods with dispersion interaction naturally included and they are also affordable with $O(N^5)$.

It is well known that MP2 typically overestimates dispersion interactions. Decomposing the MP2 interaction energy into different energy contributions with the analogous symmetry adapted perturbation theory model reveals that the main MP2 error source stems from the inaccuracy of uncoupled Hartree-Fock (UCHF) dispersion which lacks enough description of the intra-molecular correlation. Hesselmann and his collaborator developed "coupled" MP2 (MP2C) by replacing UCHF dispersion with more accurate CKS dispersion based on TDDFT (or coupled Kohn-Sham) theory [34, 35]. The applications on a variety of systems prove that MP2C is a reliable method for two-body interactions with the near CCSD(T) accuracy. Our work on this method has focusing on improving the computational efficiency of MP2C without sacrificing its accuracy. Our efforts focused on decreasing the costs by reducing the need for large basis sets.

Our first effort is to consider MP2C dispersion corrections (i.e. $\Delta E_{disp}^{MP2C} = E_{disp}^{CKS} - E_{disp}^{UCHF}$). Generally, the accurate dispersion energy is computed with dimer-centered basis set which needs to include basis functions from the partner molecule when computing molecular orbitals. However, for MP2C dispersion correction, we just require the accuracy in the difference between CKS and UCHF dispersions, instead of the accuracy for each term individually. Fortunately, the basis set errors introduced by using a smaller monomer-centered basis set for CKS and UCHF dispersion cancel each other very well, giving a dispersion correction fairly close to the dimer-centered one. With this scheme, ~ 10 fold speed-up is achieved for the calculation of the dispersion correction in an individual

dimer. More significantly, in the context of fragment-based molecular crystal studies, taking advantage of the translational symmetry in such periodic systems enables one to accelerate the dispersion correction calculation by two orders of magnitude, without compromising accuracy.

Though the dispersion correction is almost free in MC basis set, this step does not form the main computational bottleneck in MP2C method. Our next effort focused on accelerating MP2 calculations. After recognizing that the much effort consumed for UCHF dispersion calculation in MP2 will be discarded and replaced by CKS dispersion, the first scheme utilizes Coulomb attenuation in the calculation of the correlation energy. Accordingly, the short-range attenuated correlation enables one to exploit sparsity in numerical integration. More significantly, with proper choice of the attenuation strength ($r_0 = 0.9$ Å), the attractive attenuation error introduced by short-range correlation cancels the typically repulsive finite basis set error to some extent. It means only with aug-cc-pVTZ basis set, one can avoid the TZ-QZ extrapolation and achieve the accuracy in CBS limit, which realizes several fold speed-ups compared to the conventional MP2C/CBS.

To make attenuated MP2C/aTZ model more flexible for attenuation strength instead of being restricted at only one optimal point $r_0 = 0.9$ Å, dispersion energy can also be computed in the CBS limit. The short-range correlation turns out to be insensitive to basis set, allowing the use of small basis set to improve the computational speed. This MP2C/SR-aXZ models require the dimer-centered dispersion energies computed in the CBS limit. In that case, one cannot take advantage of the monomer-centered algorithm for dispersion correction. As an alternative method, explicitly correlated MP2C (MP2C-F12)

enables one to approach CBS limit with small basis set, depending on the finite basis set error cancellations between HF and correlation. All these methods are proved to be reliable by examinations on a variety of two-body interactions and molecular crystal lattice energy predictions. Thus they can be widely applied for other interesting many-body systems. One should note, however, that in conventional MP2C/CBS, Hartree-Fock energy with large basis set is the main bottleneck, and in MP2-F12 algorithm, correlation calculation requires more CPU time and RAM space even with density-fitting technique implemented. Future efforts should be spent on dealing with these two problems.

For three-body interactions where almost all computationally practical methods exhibit problems for the 3B-69 benchmark trimer set, a new approach for describing these interactions was needed. To account for the associative electron density fluctuation correlation, Axilrod-Teller-Muto triple-dipole dispersion contribution is used for three-body MP2 interaction energy correction. Meanwhile, Tang-Toennis damping function is necessary to cancel out errors in missing higher order correlations. This MP2+ATM method is fast and provides substantial accuracy improvements over MP2. Tests on the 3B-69 benchmark trimer set, pyrazole potential energy surface, and benzene crystal lattice energy demonstrate that the MP2+ATM model is reliable for three-body interactions.

For the future work on noncovalent interactions, there are several problems we are particularly interested in:

- **Intramolecular interactions:** Develop fast and accurate electronic structure methods for intra-molecular interactions. It is challenging for the current low-cost methods to predict exactly the same relative energy order as CCSD(T) on P76 test set [135].

Intra-molecular dispersion correction seems to be a promising starting point, but the question is how to define the correct form of intra-molecular dispersion, i.e. isolating the noncovalent correlations with the covalent correlations.

- **MP2C/CBS for two-body interactions:** We have recognized that the main computational bottleneck in MP2C/CBS approach is the requirement of large basis set for Hartree-Fock calculation. How to improve the computational efficiency for large basis set SCF calculation is our consideration in the future.

- **Multipole expansion of the electrostatic interaction:** AIFF electrostatic interaction has large error in the short-range molecular separations, due to the penetration errors. There are several approaches proposed to reduce such error [136, 137]. We want to examine the reliability of these methods for molecular crystal lattice energy prediction.

- **HMBI model for large systems:** Having achieved significant speedups for MP2C dispersion correction makes us wondering whether there is an efficient way to fully exploit the periodicity of molecular crystals, with most of the effort computing properties of symmetry-unique monomers in the central unit cell. The polarization terms (i.e. electrostatics, induction, and dispersion) for several dozen dimers can be evaluated almost on the fly. The question left is how to include the exchange effect efficiently. One possible way we think is to use damping functions or Coulomb attenuations for the polarization terms.

Bibliography

[1] Kevin E. Riley, Michal Pitonˇcák, Petr Jurecˇcka, and Pavel Hobza. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.*, 110(9):5023–5063, September 2010.

[2] Edward G. Hohenstein and C. David Sherrill. Wavefunction methods for noncovalent interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(2):304–326, March 2012.

[3] Shuhao Wen, Kaushik Nanda, Yuanhang Huang, and Gregory J O Beran. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.*, 14(21):7578–90, June 2012.

[4] Mark S Gordon, Dmitri G Fedorov, Spencer R Pruitt, and Lyudmila V Slipchenko. Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev.*, 112(1):632–72, January 2012.

[5] Krishnan Raghavachari and Arjun Saha. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.*, April 2015.

[6] Gregory J O Beran. Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *J. Chem. Phys.*, 130(16):164115, April 2009.

[7] Gregory J. O. Beran and Kaushik Nanda. Predicting Organic Crystal Lattice Energies with Chemical Accuracy. *J. Phys. Chem. Lett.*, 1(24):3480–3487, December 2010.

[8] Ali Sebetci and Gregory J. O. Beran. Spatially Homogeneous QM/MM for Systems of Interacting Molecules with on-the-Fly ab Initio Force-Field Parametrization. *J. Chem. Theory Comput.*, 6(1):155–167, January 2010.

[9] Shuhao Wen and Gregory J. O. Beran. Accurate Molecular Crystal Lattice Energies from a Fragment QM/MM Approach with On-the-Fly Ab Initio Force Field Parametrization. *J. Chem. Theory Comput.*, 7(11):3733–3742, November 2011.

[10] Anthony Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2013.

[11] Alston J. Misquitta and Anthony J. Stone. Accurate Induction Energies for Small Organic Molecules:? 1. Theory. *J. Chem. Theory Comput.*, 4(1):7–18, January 2008.

[12] Alston J. Misquitta, Anthony J. Stone, and Sarah L. Price. Accurate Induction Energies for Small Organic Molecules. 2. Development and Testing of Distributed Polarizability Models against SAPT(DFT) Energies. *J. Chem. Theory Comput.*, 4(1):19–32, January 2008.

[13] Alston J Misquitta and Anthony J Stone. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.*, 124(2):024111, January 2006.

[14] A. J. Stone and A. J. Misquitta. Atom–atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.*, 26(1):193–222, January 2007.

[15] Anthony J. Stone. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.*, 1(6):1128–1132, November 2005.

[16] Gregory J O Beran, Shuhao Wen, Kaushik Nanda, Yuanhang Huang, and Yonaton Heit. Accurate and robust molecular crystal modeling using fragment-based electronic structure methods. *Top. Curr. Chem.*, 345:59–93, January 2014.

[17] P. Hohenberg. Inhomogeneous Electron Gas. *Phys. Rev.*, 136(3B):B864–B871, November 1964.

[18] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

[19] Efthimios Kaxiras. *Atomic and electronic structure of solids*. Cambridge University Press, 2003.

[20] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[21] Wolfram Koch, Max C Holthausen, and Max C Holthausen. *A chemist's guide to density functional theory*, volume 2. Wiley-Vch Weinheim, 2001.

[22] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007, 1989.

[23] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.*, 94(7):1887–1930, November 1994.

[24] Trent M Parker, Lori A Burns, Robert M Parrish, Alden G Ryno, and C David Sherrill. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.*, 140(9):094106, March 2014.

[25] Lori A Burns, Alvaro Vázquez-Mayagoitia, Bobby G Sumpter, and C David Sherrill. 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.*, 134(8):084107, February 2011.

[26] A Hesselmann, G Jansen, and M Schütz. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: a new efficient method to study intermolecular interaction energies. *J. Chem. Phys.*, 122(1):14103, January 2005.

[27] Mutasem Omar Sinnokrot and C. David Sherrill. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer:? Sandwich, T-Shaped, and Parallel-Displaced Configurations. *J. Phys. Chem. A*, 108(46):10200–10207, November 2004.

[28] Lori A Burns, Michael S Marshall, and C David Sherrill. Appointing silver and bronze standards for noncovalent interactions: a comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches. *J. Chem. Phys.*, 141(23):234111, December 2014.

[29] Florian Weigend, Andreas Köhn, and Christof Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations. *The Journal of chemical physics*, 116(8):3175–3183, 2002.

[30] Mareike Gerenkamp and Stefan Grimme. Spin-component scaled second-order møller–plesset perturbation theory for the calculation of molecular geometries and harmonic vibrational frequencies. *Chemical physics letters*, 392(1):229–235, 2004.

[31] Yousung Jung, Rohini C Lochan, Anthony D Dutoi, and Martin Head-Gordon. Scaled opposite-spin second order Møller-Plesset correlation energy: an economical electronic structure method. *J. Chem. Phys.*, 121(20):9793–802, November 2004.

[32] Robert A Distasio Jr and Martin Head-Gordon. Optimized spin-component scaled second-order møller-plesset perturbation theory for intermolecular interaction energies. *Molecular Physics*, 105(8):1073–1083, 2007.

[33] Michal Pitoňák, Pavel Neogrady, Jiří Černỳ, Stefan Grimme, and Pavel Hobza. Scaled mp3 non-covalent interaction energies agree closely with accurate ccsd (t) benchmark data. *ChemPhysChem*, 10(1):282–289, 2009.

[34] Andreas Hesselmann. Improved supermolecular second order Møller-Plesset intermolecular interaction energies using time-dependent density functional response theory. *J. Chem. Phys.*, 128(14):144112, 2008.

[35] Michal Pitonak and Andreas Heßelmann. Accurate intermolecular interaction energies from a combination of mp2 and tddft response theory. *Journal of Chemical Theory and Computation*, 6(1):168–178, 2009.

[36] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25(12):1463–73, September 2004.

[37] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–99, November 2006.

[38] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615–20, November 2008.

[39] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15):154104, April 2010.

[40] Amir Karton and Jan ML Martin. Comment on:estimating the hartree–fock limit from finite basis set calculations[jensen f (2005) theor chem acc 113: 267]. *Theoretical Chemistry Accounts*, 115(4):330–333, 2006.

[41] Trygve Helgaker, Wim Klopper, Henrik Koch, and Jozef Noga. Basis-set convergence of correlated calculations on water. *The Journal of chemical physics*, 106(23):9639–9646, 1997.

[42] Petr Jurecka, Jirí Sponer, Jirí Cerný, and Pavel Hobza. 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.*, 8(17):1985–93, May 2006.

[43] Tait Takatani, Edward G Hohenstein, Massimo Malagoli, Michael S Marshall, and C David Sherrill. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *J. Chem. Phys.*, 132(14):144104, April 2010.

[44] Lucie Gráfová, Michal Pitonák, Jan Rezac, and Pavel Hobza. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended s22 data set. *Journal of Chemical Theory and Computation*, 6(8):2365–2376, 2010.

[45] Jan Rezáč, Kevin E Riley, and Pavel Hobza. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.*, 7(8):2427–2438, August 2011.

[46] Jan Rezáč, Kevin E. Riley, and Pavel Hobza. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *Journal of Chemical Theory and Computation*, 7(11):3466–3470, November 2011.

[47] Robert Sedlak, Tomasz Janowski, Michal Pito?ák, Jan Rezáč, Peter Pulay, and Pavel Hobza. The accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.*, 9(8):3364–3374, August 2013.

[48] Stefan Grimme. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chemistry*, 18(32):9955–64, August 2012.

[49] Tobias Risthaus and Stefan Grimme. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.*, 9(3):1580–1591, March 2013.

[50] Shuhao Wen and Gregory J. O. Beran. Crystal Polymorphism in Oxalyl Dihydrazide: Is Empirical DFT-D Accurate Enough? *J. Chem. Theory Comput.*, 8(8):2698–2705, August 2012.

[51] Oliver Marchetti and Hans-Joachim Werner. Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted MP2 method. *J. Phys. Chem. A*, 113(43):11580–5, October 2009.

[52] Michal Pitonák, Jan Rezác, and Pavel Hobza. Spin-component scaled coupled-clusters singles and doubles optimized towards calculation of noncovalent interactions. *Phys. Chem. Chem. Phys.*, 12(33):9611–4, September 2010.

[53] Tait Takatani, Edward G Hohenstein, and C David Sherrill. 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.*, 128(12):124111, March 2008.

[54] Michael S. Marshall and C. David Sherrill. Dispersion-Weighted Explicitly Correlated Coupled-Cluster Theory [DW-CCSD(T**)-F12]. *J. Chem. Theory Comput.*, 7(12):3978–3982, December 2011.

[55] G Chałasiński and MM Szcześniak. On the connection between the supermolecular møller-plesset treatment of the interaction energy and the perturbation theory of intermolecular forces. *Molecular Physics*, 63(2):205–224, 1988.

[56] SM Cybulski, G Chal, R Moszyński, et al. On decomposition of second-order mo/ller–plesset supermolecular interaction energy and basis set effects. *The Journal of chemical physics*, 92(7):4357–4363, 1990.

[57] Slawomir M Cybulski and Marion L Lytle. The origin of deficiency of the supermolecule second-order Moller-Plesset approach for evaluating interaction energies. *J. Chem. Phys.*, 127(14):141102, October 2007.

[58] Alexandre Tkatchenko, Robert A DiStasio, Martin Head-Gordon, and Matthias Scheffler. Dispersion-corrected Møller-Plesset second-order perturbation theory. *J. Chem. Phys.*, 131(9):094106, September 2009.

[59] Michael J Deible, Odbadrakh Tuguldur, and Kenneth D Jordan. Theoretical study of the binding energy of a methane molecule in a (H2O)20 dodecahedral cage. *J. Phys. Chem. B*, 118(28):8257–63, July 2014.

[60] Jaroslav Granatier, Michal Pitoák, and Pavel Hobza. Accuracy of Several Wave Function and Density Functional Theory Methods for Description of Noncovalent Interaction of Saturated and Unsaturated Hydrocarbon Dimers. *J. Chem. Theory Comput.*, 8(7):2282–2292, July 2012.

[61] Andreas Hesselmann and Tatiana Korona. On the accuracy of DFT-SAPT, MP2, SCS-MP2, MP2C, and DFT+Disp methods for the interaction energies of endohedral complexes of the C(60) fullerene with a rare gas atom. *Phys. Chem. Chem. Phys.*, 13(2):732–43, January 2011.

[62] Edward G. Hohenstein, Heather M. Jaeger, Emily J. Carrell, Gregory S. Tschumper, and C. David Sherrill. Accurate Interaction Energies for Problematic Dispersion-Bound Complexes: Homogeneous Dimers of NCCN, P 2 , and PCCP. *J. Chem. Theory Comput.*, 7(9):2842–2851, September 2011.

[63] Glen R Jenness, Ozan Karalti, W A Al-Saidi, and Kenneth D Jordan. Evaluation of theoretical approaches for describing the interaction of water with linear acenes. *J. Phys. Chem. A*, 115(23):5955–64, June 2011.

[64] Ozan Karalti, Dario Alfè, Michael J Gillan, and Kenneth D Jordan. Adsorption of a water molecule on the MgO(100) surface as described by cluster and slab models. *Phys. Chem. Chem. Phys.*, 14(21):7846–53, June 2012.

[65] Hayes L. Williams, Eric M. Mas, Krzysztof Szalewicz, and Bogumil Jeziorski. On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. *J. Chem. Phys.*, 103(17):7374, November 1995.

[66] Alston J. Misquitta, Bogumil Jeziorski, and Krzysztof Szalewicz. Dispersion Energy from Density-Functional Theory Description of Monomers. *Phys. Rev. Lett.*, 91(3):033201, July 2003.

[67] Alston J Misquitta, Rafa Podeszwa, Bogumi Jeziorski, and Krzysztof Szalewicz. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.*, 123(21):214103, December 2005.

[68] Andreas Heß elmann and Georg Jansen. Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.*, 367(5-6):778–784, January 2003.

[69] Alston J Misquitta and Krzysztof Szalewicz. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.*, 122(21):214109, June 2005.

[70] Robert Bukowski, Joanna Sadlej, Bogumi Jeziorski, Piotr Jankowski, Krzysztof Szalewicz, Stanisaw A. Kucharski, Hayes L. Williams, and Betsy M. Rice. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *J. Chem. Phys.*, 110(8):3785, February 1999.

[71] Shuhao Wen and Gregory J. O. Beran. Accidental Degeneracy in Crystalline Aspirin: New Insights from High-Level ab Initio Calculations. *Cryst. Growth Des.*, 12(5):2169–2172, May 2012.

[72] Andrew D Bond, Roland Boese, and Gautam R Desiraju. On the polymorphism of aspirin: crystalline aspirin as intergrowths of two "polymorphic" domains. *Angew. Chem. Int. Ed. Engl.*, 46(4):618–22, January 2007.

[73] German L Perlovich, Sergey V Kurkov, Andrey N Kinchin, and Annette Bauer-Brandl. Solvation and hydration characteristics of ibuprofen and acetylsalicylic acid. *AAPS J.*, 6(1):22–30, January 2004.

[74] Tonglei Li and Shaoxin Feng. Empirically augmented density functional theory for predicting lattice energies of aspirin, acetaminophen polymorphs, and ibuprofen homochiral and racemic crystals. *Pharm. Res.*, 23(10):2326–32, October 2006.

[75] Panagiotis G Karamertzanis, Graeme M Day, Gareth W A Welch, John Kendrick, Frank J J Leusen, Marcus A Neumann, and Sarah L Price. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.*, 128(24):244708, June 2008.

[76] Shinbyoung Ahn, Fang Guo, Benson M Kariuki, and Kenneth D M Harris. Abundant polymorphism in a system with multiple hydrogen-bonding opportunities: oxalyl dihydrazide. *J. Am. Chem. Soc.*, 128(26):8441–52, July 2006.

[77] Yuanhang Huang, Yihan Shao, and Gregory J O Beran. Accelerating MP2C dispersion corrections for dimers and molecular crystals. *J. Chem. Phys.*, 138(22):224112, June 2013.

[78] Jeremy P. Dombroski, Stephen W. Taylor, and Peter M. W. Gill. KWIK: Coulomb Energies in O (N) Work. *J. Phys. Chem.*, 100(15):6272–6276, January 1996.

[79] PETER M. W. GILL, ROSS D. ADAMSON, and JOHN A. POPLE. Coulomb-attenuated exchange energy density functionals. *Mol. Phys.*, 88(4):1005–1009, July 1996.

[80] Aaron M. Lee, Stephen W. Taylor, Jeremy P. Dombroski, and Peter M. W. Gill. Optimal partition of the Coulomb operator. *Phys. Rev. A*, 55(4):3233–3235, April 1997.

[81] Peter M.W. Gill. A new expansion of the Coulomb interaction. *Chem. Phys. Lett.*, 270(1-2):193–195, May 1997.

[82] Ross D. Adamson, Jeremy P. Dombroski, and Peter M.W. Gill. Chemistry without Coulomb tails. *Chem. Phys. Lett.*, 254(5-6):329–336, May 1996.

[83] Ross D. Adamson and Peter M.W. Gill. Effects of Coulomb attenuation on chemical properties. *J. Mol. Struct. THEOCHEM*, 398-399:45–54, June 1997.

[84] Itai Panas. Aspects of density functional theory in ab initio quantum chemistry: external correlation for free. *Chem. Phys. Lett.*, 245(2-3):171–177, October 1995.

[85] Itai Panas and Anders Snis. Effective dynamic correlation in multiconfigurational wave-function calculations on atoms and molecules. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)*, 97(1-4):232–239, October 1997.

[86] Andreas Savin and Heinz-Jrgen Flad. Density functionals for the Yukawa electron-electron interaction. *Int. J. Quantum Chem.*, 56(4):327–332, November 1995.

[87] Thierry Leininger, Hermann Stoll, Hans-Joachim Werner, and Andreas Savin. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.*, 275(3-4):151–160, August 1997.

[88] Paola Gori-Giorgi and Andreas Savin. Properties of short-range and long-range correlation energy density functionals from electron-electron coalescence. *Phys. Rev. A*, 73(3):032506, March 2006.

[89] Julien Toulouse, Andreas Savin, and Heinz-Jrgen Flad. Short-range exchange-correlation energy of a uniform electron gas with modified electron-electron interaction. *Int. J. Quantum Chem.*, 100(6):1047–1056, December 2004.

[90] Julien Toulouse, François Colonna, and Andreas Savin. Long-rangeshort-range separation of the electron-electron interaction in density-functional theory. *Phys. Rev. A*, 70(6):062505, December 2004.

[91] Kamal Sharkas, Julien Toulouse, and Andreas Savin. Double-hybrid density-functional theory made rigorous. *J. Chem. Phys.*, 134(6):064113, February 2011.

[92] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchangecorrelation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3):51–57, July 2004.

[93] Michael J G Peach, Aron J Cohen, and David J Tozer. Influence of Coulomb-attenuation on exchange-correlation functional quality. *Phys. Chem. Chem. Phys.*, 8(39):4543–9, October 2006.

[94] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Development of exchange-correlation functionals with minimal many-electron self-interaction error. *J. Chem. Phys.*, 126(19):191109, May 2007.

[95] Hisayoshi Iikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.*, 115(8):3540, August 2001.

[96] Yoshihiro Tawada, Takao Tsuneda, Susumu Yanagisawa, Takeshi Yanai, and Kimihiko Hirao. A long-range-corrected time-dependent density functional theory. *J. Chem. Phys.*, 120(18):8425–33, May 2004.

[97] Benjamin G Janesko, Thomas M Henderson, and Gustavo E Scuseria. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.*, 11(3):443–54, January 2009.

[98] Elon Weintraub, Thomas M. Henderson, and Gustavo E. Scuseria. Long-Range-Corrected Hybrids Based on a New Model Exchange Hole. *J. Chem. Theory Comput.*, 5(4):754–762, April 2009.

[99] Robin Haunschild and Gustavo E Scuseria. Range-separated local hybrids. *J. Chem. Phys.*, 132(22):224106, June 2010.

[100] Jong-Won Song, Daoling Peng, and Kimihiko Hirao. A semiempirical long-range corrected exchange correlation functional including a short-range Gaussian attenuation (LCgau-B97). *J. Comput. Chem.*, 32(15):3269–75, November 2011.

[101] Roberto Peverati and Donald G. Truhlar. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.*, 2(21):2810–2817, November 2011.

[102] Matthew Goldey and Martin Head-Gordon. Attenuating Away the Errors in Inter- and Intramolecular Interactions from Second-Order MøllerPlesset Calculations in the Small Aug-cc-pVDZ Basis Set. *J. Phys. Chem. Lett.*, 3(23):3592–3598, December 2012.

[103] Matthew Goldey, Anthony Dutoi, and Martin Head-Gordon. Attenuated second-order Møller-Plesset perturbation theory: performance within the aug-cc-pVTZ basis. *Phys. Chem. Chem. Phys.*, 15(38):15869–75, October 2013.

[104] Jeng-Da Chai and Martin Head-Gordon. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.*, 128(8):084106, February 2008.

[105] Oliver Marchetti and Hans-Joachim Werner. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys. Chem. Chem. Phys.*, 10(23):3400–9, June 2008.

[106] Anthony D Dutoi and Martin Head-Gordon. A study of the effect of attenuation curvature on molecular correlation energies by introducing an explicit cutoff radius into two-electron integrals. *J. Phys. Chem. A*, 112(10):2110–9, March 2008.

[107] Florian Weigend, Marco Häser, Holger Patzelt, and Reinhart Ahlrichs. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.*, 294(1-3):143–152, September 1998.

[108] Yousung Jung, Alex Sodt, Peter M W Gill, and Martin Head-Gordon. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U. S. A.*, 102(19):6692–7, May 2005.

[109] Jan Rezáč, Yuanhang Huang, Pavel Hobza, and Gregory J O Beran. Benchmark calculations of three-body intermolecular interactions and the performance of low-cost electronic structure methods. *J. Chem. Theory Comput.*, submitted, 2015.

[110] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008.

[111] B. M. Axilrod and E. Teller. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.*, 11(6):299, December 1943.

[112] Yoshio Muto. Force between nonpolar molecules. In *Proc. Phys.-Math Soc. Jpn*, volume 17, pages 629–631, 1943.

[113] Alexandre Tkatchenko and O Anatole von Lilienfeld. Popular Kohn-Sham density functionals strongly overestimate many-body interactions in van der Waals systems. *Physical Review B*, 78(4):045116, 2008.

[114] SM Cybulski, G Chal, R Moszyński, et al. On decomposition of second-order Mo/ller–Plesset supermolecular interaction energy and basis set effects. *The Journal of chemical physics*, 92(7):4357–4363, 1990.

[115] Robert Sedlak, Kevin E Riley, Jan Řezáč, Michal Pitoňák, and Pavel Hobza. MP2.5 and MP2.X: approaching CCSD (T) quality description of noncovalent interaction at the cost of a single CCSD iteration. *ChemPhysChem*, 14(4):698–707, 2013.

[116] Grzegorz Chal, Rick A Kendall, et al. Supermolecular approach to many-body dispersion interactions in weak van der Waals complexes: He, Ne, and Ar trimers. *The Journal of chemical physics*, 101(10):8860–8869, 1994.

[117] Rafał Podeszwa and Krzysztof Szalewicz. Three-body symmetry-adapted perturbation theory based on Kohn-Sham description of the monomers. *The Journal of chemical physics*, 126(19):194101, 2007.

[118] Rafał Podeszwa, Betsy M Rice, and Krzysztof Szalewicz. Predicting structure of molecular crystals from first principles. *Physical review letters*, 101(11):115503, 2008.

[119] Matthew R Kennedy, Ashley Ringer McDonald, A Eugene DePrince III, Michael S Marshall, Rafal Podeszwa, and C David Sherrill. Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *The Journal of chemical physics*, 140(12):121104, 2014.

[120] Alston J Misquitta and Anthony J Stone. Dispersion energies for small organic molecules: first row atoms. *Molecular Physics*, 106(12-13):1631–1643, 2008.

[121] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of chemical physics*, 110(13):6158–6170, 1999.

[122] Alston J Misquitta and Anthony J Stone. CamCASP v5.6 (2011), Accessed Februrary 23 2011.

[123] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics*, 132(15):154104, 2010.

[124] O Anatole von Lilienfeld and Alexandre Tkatchenko. Two-and three-body interatomic dispersion energy contributions to binding in molecules and solids. *The Journal of chemical physics*, 132(23):234109, 2010.

[125] Alexandre Tkatchenko, Robert A DiStasio Jr, Roberto Car, and Matthias Scheffler. Accurate and efficient method for many-body van der Waals interactions. *Physical review letters*, 108(23):236402, 2012.

[126] Robert A DiStasio, O Anatole von Lilienfeld, and Alexandre Tkatchenko. Collective many-body van der Waals interactions in molecular systems. *Proceedings of the National Academy of Sciences*, 109(37):14791–14795, 2012.

[127] A Otero-de-la Roza and Erin R Johnson. Many-body dispersion interactions from the exchange-hole dipole moment model. *The Journal of chemical physics*, 138(5):054103, 2013.

[128] Takeshi Sato and Hiromi Nakai. Density functional method including weak interactions: Dispersion coefficients based on the local response approximation. *The Journal of chemical physics*, 131(22):224104, 2009.

[129] KT Tang and J Peter Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of chemical physics*, 80(8):3726–3741, 1984.

[130] Victor F Lotrich and Krzysztof Szalewicz. Symmetry-adapted perturbation theory of three-body nonadditivity of intermolecular interaction energy. *The Journal of chemical physics*, 106(23):9668–9687, 1997.

[131] Wojciech Cencek, Malgorzata Jeziorska, Omololu Akin-Ojo, and Krzysztof Szalewicz. Three-body contribution to the helium interaction potential. *The Journal of Physical Chemistry A*, 111(44):11311–11319, 2007.

[132] Jun Yang, Weifeng Hu, Denis Usvyat, Devin Matthews, Martin Schütz, and Garnet Kin-Lic Chan. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science*, 345(6197):640–643, 2014.

[133] Ashley L Ringer and C David Sherrill. First principles computation of lattice energies of organic solids: The benzene crystal. *Chemistry-A European Journal*, 14(8):2542–2547, 2008.

[134] M Pitoňák, Pavel Neogrady, and Pavel Hobza. Three-and four-body nonadditivities in nucleic acid tetramers: a ccsd (t) study. *Physical Chemistry Chemical Physics*, 12(6):1369–1378, 2010.

[135] Haydee Valdes, Kristỳna Pluháčková, Michal Pitonák, Jan Řezáč, and Pavel Hobza. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Physical Chemistry Chemical Physics*, 10(19):2747–2757, 2008.

[136] Anthony J Stone. Electrostatic damping functions and the penetration energy. *The Journal of Physical Chemistry A*, 115(25):7017–7027, 2011.

[137] Bo Wang and Donald G Truhlar. Screened electrostatic interactions in molecular mechanics. *Journal of Chemical Theory and Computation*, 10(10):4480–4487, 2014.