Title
Improvements in Density Functional Theory Calculations, Analysis, and Applications

Author
Veccham Krishna Prasad, Srimukh Prasad

Publication Date
2020

Peer reviewed|Thesis/dissertation

Improvements in Density Functional Theory Calculations, Analysis, and Applications

by

Srimukh Prasad Veccham Krishna Prasad

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin P. Head-Gordon, Chair
Professor K. Birgitta Whaley
Professor Kranthi K. Mandadapu

Fall 2020

Improvements in Density Functional Theory Calculations, Analysis, and Applications

Abstract

Improvements in Density Functional Theory Calculations, Analysis, and Applications

by

Srimukh Prasad Veccham Krishna Prasad

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Martin P. Head-Gordon, Chair

Density functional theory (DFT) is the most widely used quantum chemistry method. This dissertation explores and advances different aspects of DFT. The computational cost of DFT scales as the third power of the system size. In Chapter 2, we introduce an embedded many-body expansion, called the polarized many-body expansion, which removes the computational bottleneck steps in DFT calculations by exploiting the pairwise additivity of some components of interaction energy in molecular clusters. Energy decomposition analysis fragments interaction energy into physico-chemically meaningful components like permanent electrostatics, dispersion, polarization, and charge transfer. In Chapter 3, we introduce a non-perturbative approach to decompose charge transfer energy into pairwise additive terms. This approach improves upon the previous perturbative approach and can be applied to chemical complexes with both weak and strong charge transfer interactions. Different density functional approximations (DFAs) show varied performance for predicting properties of different chemical systems. In Chapters 4 and 5, we identify DFAs which give good accuracy for predicting the interaction energy of hydrogen with different model binding sites. These DFAs can potentially be used for *in silico* high-throughput screening of materials capable of storing hydrogen at desired conditions.

To Amma and Naana,
for always putting me first.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my PhD advisor Prof. Martin Head-Gordon for his endless academic, scientific, and emotional support that he provided me throughout graduate school. Martin was always kind when I made mistakes, patient when I was slow, and despite this, he was encouraging when I had new ideas. Meeting with Martin on Mondays over the past five years has been a humbling and motivating experience at the same time – I could sense my enthusiasm levels increase after meeting with him. I appreciate his ability to see the silver lining in my projects even in their darkest phases and inspiring me to push forward. Most importantly, I would like to thank him for teaching me how to think and ask the right questions – my two most important takeaways from graduate school. While I am against the concept of idolizing human role models, I somehow find myself making an exception for Martin. Lastly, I would like to thank Martin for supporting me financially throughout graduate school and being very understanding of my international student status in various regards.

I would like to thank Dr. Joonho Lee for teaching me everything I know about the implementational details of electronic structure theory. His infectious enthusiasm for science motivated me to work long hours tirelessly. I would like to thank him for always being available to talk about science at any given time. Joonho, with his almost-superhuman programming and debugging ability, helped me get started with programming in Q-Chem and troubleshoot multiple highly non-trivial programming bugs. I am also grateful to him for sharing his core idea behind PolBE and working with me on it. Also, Joonho has been my longest and closest friend throughout graduate school and I have enjoyed hanging out with him outside of work as well.

I would like to thank Dr. Luke Bertels for being the best officemate I could have asked for. I am truly grateful for his kindness and generosity, which has earned him the sobriquet "St. Luke." in our research group. His quick-witted banter made my long work hours even more enjoyable. I am grateful to Matthias Loipersberger for going through the ups and downs of graduate school with me and for sharing my enthusiasm for understanding different aspects of programming in Q-Chem and reading SCF-MI papers for the n^{th} time. I am grateful to Dr. Yuezhi Mao for always being available to help me out and answer questions. His deep knowledge of the theoretical and computational aspects of EDA helped me get started quickly on multiple projects. Most importantly, I will always be grateful for all the help he offered in preparing me for my qualifying exam. I would like to Dr. Narbe Mardirossian for telling me that I could achieve anything if I worked hard enough and leading by example. I cherish the lively work atmosphere he had created in the group and am grateful to him for including me in it. I am also very grateful for his help with navigating through my post-graduation plans. I would like to thank Diptarka Hait for always enthusiastically discussing scientific problems and for all the late-night company in Pitzer Center. Discussing Indian politics with him introduced me to perspectives that I was previously unaware of. I am thankful to Dr. Ehud Tsivion for being my first mentor in graduate school and teaching me about the hydrogen storage problem. I would like to thank Prof. Christopher Stein for

helping me navigate difficult times despite knowing me for only a short time at that point. With his depth of knowledge and humility, Christopher has been a real inspiration in the latter years of my time at Berkeley. I am also grateful to Dr. Alexander Zech for educating me about various programming paradigms and helping me think about them.

I had the opportunity to learn many things in addition to quantum chemistry during my graduate school career through various courses. I would like to thank Prof. Lin Lin, Prof. Birgitta Whaley, Prof. Robert Littlejohn, Prof. Eran Rabani, and Prof. Jennifer Listgarten for teaching different interesting courses. I am grateful to Prof. Raghavan Sunoj, my undergraduate and masters research advisor for introducing me to computational chemistry early on. I would like to thank Prof. Jeffrey Long, his group members, and other members of the HyMARC consortium for discussing various aspects of the hydrogen storage problem with me.

Throughout my graduate career in the Martin Head-Gordon research group, I have interacted and learned from many group members. I would like to thank Dr. Paul Horn for his extensive contributions to the EDA functionality in Q-Chem. I am thankful to Dr. Jerome Gonthier and Dr. Eloy Ramos-Cordoba for fun lunchtime conversations. I am grateful to Dr. Daniel Levine for encouraging me to just start somewhere when overwhelmed. I am thankful for all past and current group members including Dr. Mickael Delcey, Dr. Susi Lehtola, Dr. David Small, Prof. Shane Yost, Dr. Jonathan Thirman, Dr. Jonathon Witte, Dr. Alec White, Prof. Aurora Pribram-Jones, Prof. Tamar Stein, Prof. Tim Stauch, Dr. Xintian Feng, Dr. Qinghui Ge, Prof. Zhou Lin, Jonathan Weisberg, Dr. Romit Chakraborty, Dr. Jeroen Van der Mynsbrugge, Dr. Cameron Mackie, Dr. James Shee, Brad Ganoe, Katherine Oosterbaan, Lance Bettinson, Josie Hendrix, Abdulrahman Aldossary, Jonathan Wong, Elliot Rossomme, Kevin Ikeda, Adam Rettig, Leonardo dos Anjos Cunha, Jiashu Liang, Juanes Arias-Martinez and Erum Mansoor.

I would like to thank Jessie Woodcock for proactively taking care of all administrative hurdles. I also gratefully acknowledge continued financial support from UC Berkeley and the HyMARC consortium of the Department of Energy.

My parents have been a constant source of support and encouragement throughout my academic career. I am deeply grateful for their unconditional and absolute encouragement of my academic pursuits, without regard for how my choices would affect them. While my parents are not academically-oriented themselves, they always stressed the importance of education which shaped my thinking and played a key role in all of my life decisions. I am thankful to them for supporting my unconventional choice (by their societal standards) to pursue chemistry and assisted me in every possible way they could. I would like to thank my father (Naana), Krishna Prasad, for dropping everything to just spend time with me when I visited home. I am grateful to my mother (Amma), Sreelakshmi, for teaching me the value of hard work and persistence early on. Sometimes when research took a hard turn, the only thing I would look forward to the entire day was to talk to my parents over the phone in the evening.

I would like to thank my grandfather, Sathyanarayana Gupta, for his steadfast support of my academic pursuits. I am grateful to him for taking interest in my work and attempting to

understand what I do. I would like to thank my grandmother, Indiramma, for her love and persistent reminders to wear a sweater in cold weather. I should thank my uncle, Nagendra Prasad, for being a great shopping buddy and always telling me how proud he is of all my achievements.

I am grateful to Anup Kumar for always taking the initiative to ensure that we stay in touch. Sharing my graduate school experiences with Anup Kumar and Ankush Das and listening to theirs has helped me put many things in perspective. I would like to thank my roommates Archit Gupta, Prateek Shah, and Niladri Chatterji for always picking me up when I was down. I am also grateful for their superior cooking skills and patience while I improved mine. It has always been fun to reminisce and partially relive our undergraduate days at IIT Bombay in Berkeley. I am thankful to Prof. Christopher Stein, Dr. Jagna Witek, Matthias Loipersberger, Dr. Sophia Betzler, and Tanya Krishnakumar for keeping me sane in my last year of graduate school. I would like to thank all members of BAS for periodically encouraging me to take a break from work to hang out. I am grateful to Nisheet Patel and Saaz Sakrikar for keeping in touch throughout graduate school. I would like to thank Krishnakumar Ramamurti, Bharathi Iyer, and Anita Krishnakumar for always welcoming me to their home for holidays.

Research is hard and has its ups and downs. Ideas and analyses do not always work on the first attempt. Through all of these victories and setbacks, both small and large, I had the unwavering support of my girlfriend and partner, Tanya Krishnakumar. I am sincerely grateful for her love which made graduate school a pleasurable experience. I would like to thank her for believing in me even when I did not myself and helping me overcome my severe imposter syndrome. I am thankful to her for understanding my need to work unconventional hours and always being enthusiastic to talk to me at the end of every single day. While she is not a chemist herself, I appreciate her taking interest in my work and thoroughly proof-reading all of my manuscripts as well as this dissertation. This PhD would not have been possible without Tanya and I consider myself lucky for having her as my partner. I look forward to building our life together in the future.

Chapter 1

Introduction

The wavefunction of a system is a mathematical object that completely defines its state at a quantum mechanical level. Application of suitable operators to the wavefunction can, in principle, give us any property of the system. The time-dependent Schrödinger equation, proposed by physicist Erwin Schrödinger in 1926, can be solved to recover the wavefunction ($|\Psi\rangle$) of the system:

$$\hat{H}\,|\Psi\rangle = i\hbar\frac{\partial}{\partial t}\,|\Psi\rangle \tag{1.1}$$

where $\hat{H}$ is the Hamiltonian operator of the system and t is time. The Hamiltonian operator is the quantum mechanical operator for the total energy of the system. While the equation governing the wavefunction and its time evolution have been known for around a century, its application to chemically relevant systems like atoms and molecules leads to mathematically intractable equations. Quantum chemistry is a subdomain of chemistry that attempts to solve the Schrödinger equation for chemical systems by employing suitable mathematical approximations and using it compute properties of systems. These approximations can be very different in nature and each of them is characterized by its computational cost and accuracy for predicting different properties. In this thesis, we have explored the accuracy different approximations for computing specific chemical properties. We have also attempted to push the trade-off between accuracy and computational cost by formulating a new method that makes chemically-motivated mathematical approximations. In addition, we have attempted to bridge the gap between computational calculations and experiments by formulating techniques to compute chemically-meaningful quantities, thus making calculations more interpretable.

1.1 Solving the Electronic Schrödinger Equation

In this section, we will briefly discuss different classes of approximations for solving the Schrödinger equation. As we are concerned with static properties in this work, we will

discuss Eq. (1.1) in a time-independent context:

$$\hat{H}\,|\Psi\rangle = E\,|\Psi\rangle \tag{1.2}$$

where E is the total energy of the system. When applied to chemical systems, which usually consist of nuclei and electrons, the total non-relativistic Hamiltonian can be written as a sum of kinetic energy ($\hat{T}$) and potential energy ($\hat{V}$) contributions of electrons and nuclei as shown Eq. (1.3) in atomic units:

$$\hat{H} = \hat{T}_{\mathrm{e}} + \hat{T}_{\mathrm{n}} + \hat{V}_{\mathrm{nn}} + \hat{V}_{\mathrm{ne}} + \hat{V}_{\mathrm{ee}} \tag{1.3}$$

$$= \left(-\sum_{i}^{N_{\mathrm{elec}}} \frac{1}{2}\hat{\boldsymbol{\nabla}}_i^2 \right) + \left(-\sum_{A}^{N_{\mathrm{nuc}}} \frac{1}{2m_A}\hat{\boldsymbol{\nabla}}_A^2 \right) + \left(\sum_{A}^{N_{\mathrm{nuc}}} \sum_{B>A}^{N_{\mathrm{nuc}}} \frac{Z_A Z_B}{|\hat{\mathbf{R}}_A - \hat{\mathbf{R}}_B|} \right) + \tag{1.4}$$

$$\left(-\sum_{i}^{N_{\mathrm{elec}}} \sum_{A}^{N_{\mathrm{nuc}}} \frac{Z_A}{|\hat{\mathbf{r}}_i - \hat{\mathbf{R}}_A|} \right) + \left(\sum_{i}^{N_{\mathrm{elec}}} \sum_{j>i}^{N_{\mathrm{elec}}} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|} \right)$$

where $\hat{T}_{\mathrm{e}}$ and $\hat{T}_{\mathrm{n}}$ are the kinetic energy operators of the electrons and nuclei respectively, $\hat{V}_{\mathrm{ee}}$ and $\hat{V}_{\mathrm{nn}}$ are the potential energy operators of the electrons and nuclei respectively, and $\hat{V}_{\mathrm{ne}}$ is the potential energy operator of the electron-nuclear interaction. N_{elec} is the number of electrons and N_{nuc} is the number of nuclei in the system. Z_A and Z_B are the charges on nuclei A and B located at positions $\hat{\mathbf{R}}_A$ and $\hat{\mathbf{R}}_B$ with masses m_A and m_B respectively. $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{r}}_j$ are the positions of electrons i and j respectively.

As the atomic nuclei are at least 1800 times more massive that electrons, electrons move much faster than nuclei. The Born-Oppenheimer approximation, which considers the motion of electrons in a fixed field of nuclei, can be invoked to further simplify Eq. (1.3) and decouple the electronic and nuclear motion. With the Born-Oppenheimer approximation, the total Hamiltonian $\hat{H}$ can be written as a sum of electronic and nuclear parts, and subsequently, the total wavefunction $\Psi(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\})$ factorizes into the electronic ($\Psi_e(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\})$) and nuclear components ($\Phi(\{\hat{\mathbf{R}}_A\})$) as shown in Eq. (1.5).

$$\Psi_e(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\}) = \Psi_e(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\})\Phi(\{\hat{\mathbf{R}}_A\}) \tag{1.5}$$

Throughout the rest of this section, we will discuss methods and approximations used to solve the electronic part of the Schrödinger equation shown in Eq. (1.6):

$$\left(-\sum_{i}^{N_{\mathrm{elec}}} \frac{1}{2}\hat{\boldsymbol{\nabla}}_i^2 - \sum_{i}^{N_{\mathrm{elec}}} \sum_{A}^{N_{\mathrm{nuc}}} \frac{Z_A}{|\hat{\mathbf{r}}_i - \hat{\mathbf{R}}_A|} + \sum_{i}^{N_{\mathrm{elec}}} \sum_{j>i}^{N_{\mathrm{elec}}} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|} \right) \Psi(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\}) = E\Psi_e(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\}) \tag{1.6}$$

Eq. (1.6) is an eigenvalue equation which typically has many eigenfunctions $\Psi_e(\{\hat{\mathbf{R}}_A\}, \{\hat{\mathbf{r}}_i\})$ and eigenvalues E. The eigenfunction with the lowest eigenvalue represents the electronic ground state of the system and other higher eigenvalues and their corresponding eigenfunctions represent excited states of the system.

Hartree-Fock Theory

The Hartree-Fock (HF) approximation is a mean-field approximation to the $3N_{\text{elec}}$ dimensional electronic Schrödinger equation which considers that each electron moves in a mean field generated by the remaining $(N_{\text{elec}} - 1)$ electrons. The HF solution to the electronic Schrödinger equation is equivalent to the molecular orbital picture of atoms and molecules that one studies in a general chemistry class as a freshman. As electrons are fermions, electronic wavefunctions should be anti-symmetric with respect to interchange of any two electrons. One way to ensure this is to write electronic wavefunctions as a Slater determinant of orthogonal molecular orbitals $\chi_i(\mathbf{x}_j)$ where $\mathbf{x}_j$ is the collective position and spin coordinate of electron j.

$$\Phi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \cdots & \chi_1(\mathbf{x}_n) \\ \chi_2(\mathbf{x}_1) & \chi_2(\mathbf{x}_2) & \cdots & \chi_2(\mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_n(\mathbf{x}_1) & \chi_n(\mathbf{x}_2) & \cdots & \chi_n(\mathbf{x}_n) \end{vmatrix} \tag{1.7}$$

Each molecular orbital $\chi_i(\mathbf{x}_j)$ can be decomposed into spatial $\psi_i(\mathbf{r}_j)$ and spin $\sigma_i(\omega_j)$ parts, as shown in Eq. (1.8)

$$\chi_i(\mathbf{x}_j) = \psi_{i,\sigma}(\mathbf{r}_j)\sigma_i(\omega_j) \tag{1.8}$$

Usually, σ_i is chosen to be either an α or β spin orbital which are orthonormal to each other. As these equations are usually solved on a computer, one has to invoke the discretization approximation in the form of writing the spatial part of the molecular orbitals as a linear combination of some basis functions $\phi_\mu(\mathbf{r}_j)$:

$$\psi_{i,\sigma}(\mathbf{r}_j) = \sum_{\mu}^{N} C_{\mu i,\sigma}\phi_\mu(\mathbf{r}_j) \tag{1.9}$$

where $C_{\mu i,\sigma}$ is the coefficient of the i^{th} spatial molecular orbital with spin σ on the μ^{th} basis function. Atom-centered Gaussian basis functions are typically used for molecular systems[1] and plane-wave basis functions are most commonly used for periodic systems.[2] The rest of this work will use atom-centered Gaussian basis functions, sometimes also referred to as atomic orbitals (AOs). The coefficients $C_{\mu i,\sigma}$ are determined using the variational principle, which states that the best set of coefficients for the ground state wavefunction is one that minimizes its energy.

$$E_0 \leq \frac{\langle \Phi(\mathbf{C}) | \hat{H} | \Psi(\mathbf{C}) \rangle}{\langle \Phi(\mathbf{C}) | \Phi(\mathbf{C}) \rangle} = E(\mathbf{C}) \tag{1.10}$$

where E_0 is the exact ground state energy and $E(\mathbf{C})$ is the energy of the single Slater determinant with parameters $\mathbf{C}$. Substituting the Slater determinant in Eq. (1.7) into Eq. (1.10)

and employing Slater-Condon rules[3, 4], we get the Hartree-Fock energy of that Slater determinant:

$$E_{\mathrm{HF}} = \sum_{i=1}^{n} \langle \chi_i | \, \hat{h} \, | \chi_i \rangle + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} [\langle \chi_i \chi_j | \chi_i \chi_j \rangle - \langle \chi_i \chi_j | \chi_j \chi_i \rangle] \tag{1.11}$$

$$\hat{h}(\hat{\mathbf{r}}_i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^{N_{\mathrm{nuc}}} \frac{Z_A}{|\hat{\mathbf{r}}_i - \hat{\mathbf{R}}_A|} \tag{1.12}$$

$$\langle \chi_i | \, \hat{h} \, | \chi_i \rangle = \int \chi_i^*(\hat{\mathbf{x}}_1) \hat{h}(\hat{\mathbf{r}}_1) \chi_i(\hat{\mathbf{x}}_1) d\hat{\mathbf{x}}_1 \tag{1.13}$$

$$\langle \chi_i \chi_j | \chi_i \chi_j \rangle = \int \int \chi_i^*(\hat{\mathbf{x}}_1) \chi_j^*(\hat{\mathbf{x}}_2) \frac{1}{\hat{r}_{12}} \chi_i(\hat{\mathbf{x}}_1) \chi_j(\hat{\mathbf{x}}_2) d\hat{\mathbf{x}}_1 d\hat{\mathbf{x}}_2 \tag{1.14}$$

$$\langle \chi_i \chi_j | \chi_j \chi_i \rangle = \int \int \chi_i^*(\hat{\mathbf{x}}_1) \chi_j^*(\hat{\mathbf{x}}_2) \frac{1}{\hat{r}_{12}} \chi_j(\hat{\mathbf{x}}_1) \chi_i(\hat{\mathbf{x}}_2) d\hat{\mathbf{x}}_1 d\hat{\mathbf{x}}_2 \tag{1.15}$$

The first term in Eq. (1.11) is the one-electron part and the second term is the two-electron part which itself consists of Coulomb-type (Eq. (1.14)) and exchange-type integrals (Eq. (1.15)). $\hat{r}_{12} = |\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2|$ is the absolute magnitude of the distance between electrons 1 and 2 in Eqs. (1.14) and (1.15).

The electronic ground state Hartree-Fock energy is given by minimizing the Hartree-Fock energy in Eq. (1.11), subject to the constraint that all molecular orbitals are orthonormal to each other. This constraint is enforced using the method of Lagrangian multipliers. This minimization leads to the following eigenvalue problem known as the Hartree-Fock equations:

$$\hat{f}_\sigma(\hat{\mathbf{r}}_1) \psi_{i,\sigma}(\hat{\mathbf{r}}_1) = \epsilon_{i,\sigma} \psi_{i,\sigma}(\hat{\mathbf{r}}_1) \tag{1.16}$$

$$\hat{f}_\sigma(\hat{\mathbf{r}}_1) = \hat{h}(\hat{\mathbf{r}}_1) + \sum_{\sigma}^{\alpha,\beta} \sum_{j=1}^{n_\sigma} \int \frac{\psi_{j,\sigma}^*(\hat{\mathbf{r}}_2) \psi_{j,\sigma}(\hat{\mathbf{r}}_2)}{|\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2|} d\hat{\mathbf{r}}_2 - \sum_{j=1}^{n_\sigma} \int \frac{\psi_{j,\sigma}^*(\hat{\mathbf{r}}_2) \hat{P}_{12} \psi_{j,\sigma}(\hat{\mathbf{r}}_2)}{|\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2|} d\hat{\mathbf{r}}_2 \tag{1.17}$$

where $\hat{f}_\sigma(\hat{\mathbf{x}}_1)$ is the Fock operator, $\hat{P}_{12}$ is the permutation operator that swaps indices 1 and 2 and $\epsilon_{i,\sigma}$ is the orbital energy of $\psi_{i,\sigma}(\hat{\mathbf{r}}_1)$.

Substituting Eq. (1.9) into Eq. (1.16), we obtain the matrix form of the generalized eigenvalue problem:

$$\mathbf{F}_\sigma \mathbf{C}_\sigma = \mathbf{S} \mathbf{C}_\sigma \boldsymbol{\epsilon}_\sigma \tag{1.18}$$

$$F_{\mu\nu,\sigma} = \langle \phi_\mu | \, \hat{f}_\sigma \, | \phi_\nu \rangle \tag{1.19}$$

$$C_{\mu i,\sigma} = \langle \phi_\mu | \psi_{i,\sigma} \rangle \tag{1.20}$$

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle \tag{1.21}$$

$$\boldsymbol{\epsilon}_{ij,\sigma} = \epsilon_{i,\sigma} \delta_{ij} \tag{1.22}$$

where $\mathbf{F}_\sigma$ is the Fock matrix for spin σ expressed in the AO basis, $\mathbf{C}_\sigma$ is a matrix whose i^{th} column contains the coefficients of the i^{th} molecular orbital (MO) of spin σ, $\mathbf{S}$ is the AO

overlap matrix, and $\boldsymbol{\epsilon}_\sigma$ is the diagonal matrix of orbital energies with spin σ. In general, non-orthogonal atom-centered Gaussian functions are used as basis functions for molecular systems. Substituting Eq. (1.9) into Eq. (1.17), we can write the Fock matrix in AO basis in terms of 1-particle density matrix ($\mathbf{P}_\sigma$; also sometimes loosely referred to as "density matrix") as follows:

$$F_{\mu\nu,\sigma} = h_{\mu\nu} + \sum_{\lambda\xi} \langle \phi_\mu \phi_\lambda | \phi_\nu \phi_\xi \rangle \left(P_{\lambda\xi,\alpha} + P_{\lambda\xi,\beta} \right) - \sum_{\lambda\xi} \langle \phi_\mu \phi_\lambda | \phi_\xi \phi_\nu \rangle \, P_{\lambda\xi,\sigma} \tag{1.23}$$

$$P_{\mu\nu,\sigma} = \sum_{i=1}^{n_\sigma} C_{\mu i,\sigma} C^*_{\nu i,\sigma} \tag{1.24}$$

Further, we can define the Coulomb ($\mathbf{J}$) and exchange matrices ($\mathbf{K}$) from Eq. (1.23),

$$J_{\mu\nu} = \sum_{\lambda\xi} \langle \phi_\mu \phi_\lambda | \phi_\nu \phi_\xi \rangle \left(P_{\lambda\xi,\alpha} + P_{\lambda\xi,\beta} \right) \tag{1.25}$$

$$K_{\mu\nu,\sigma} = \sum_{\lambda\xi} \langle \phi_\mu \phi_\lambda | \phi_\xi \phi_\nu \rangle \, P_{\lambda\xi,\sigma} \tag{1.26}$$

The real space density ($\rho(\mathbf{r})$) can also be written in terms of the 1-particle density matrix (1-PDM) as follows:

$$\rho(\mathbf{r}) = \sum_{i=1}^{n_\sigma} \sum_{\mu\nu} P_{\mu\nu,\sigma} \phi_\mu(\mathbf{r}) \phi^*_\nu(\mathbf{r}) \tag{1.27}$$

As the Fock matrix itself is a function of the MO coefficients, Eq. (1.18) is solved iteratively using the self-consistent field (SCF) procedure. Starting with an AO basis set, charge, spin multiplicity, and positions of all nuclei in the system, the SCF algorithm can be outlined as follows:

1. Compute density independent quantities: $\mathbf{h}$ and $\mathbf{S}$
2. Guess an initial 1-PDM for the given molecular configuration and basis set
3. Construct Fock matrix using the current density using Eq. (1.23)
4. Diagonalize the Fock matrix to obtain MO coefficient matrix and update the 1-PDM using it
5. Check for convergence based on energy or density
    - If converged, compute HF energy using the converged MOs with Eq. (1.11)
    - If not converged, go back to step (3) and iterate

The SCF procedure outlined is called the Roothaan-Hall algorithm and is equivalent to fixed point iteration in numerical analysis. However, all the SCF solutions converged in this work employ some form of convergence acceleration by using gradient-like information in some form.[5–8] We refer interested readers to Refs. 9 and 10 for an in-depth discussion of the physical, mathematical, and computational aspects of Hartree-Fock theory.

The Hartree-Fock approach to solving the Schrödinger equation approximates the motion of each electron in a mean-field generated by the remaining electrons. However, in reality, electrons avoid each other and their motions are "correlated". This correlation is not captured by HF theory. The correlation energy, E_{corr}, is the difference between the exact ground state energy E_0 and the HF ground state energy for a given basis set:

$$E_{\text{corr}} = E_0 - E_{\text{HF}} \tag{1.28}$$

By definition of the electronic correlation energy in Eq. (1.28) and the variational principle (Eq. (1.10)), E_{corr} is a negative-definite quantity. While HF theory usually captures $\sim 99\%$ of the total energy, it is still vital to capture the remaining correlation energy as interesting chemical phenomena occur in this energy scale. We will discuss two starkly different categories of approaches to go beyond the mean-field approximation of HF theory. The first category is density functional theory (DFT) which offers a way to capture correlation energy at a computational cost that is comparable to HF theory. The second class of approaches is classified as *ab initio* correlated wavefunction methods, which is typically more expensive than HF theory, but offers a systematic way to improve upon it.

Density Functional Theory

Density functional theory (DFT) is the most widely used electronic structure method. Despite lack of systematic improvability inherent to its formalism, DFT has proved to be very useful for computing properties of chemical systems. DFT sits at a sweet spot in the cost-versus-accuracy trade-off curve, thus permitting application to a wide range of chemical phenomena occurring at different length and time scales. The key difference between DFT and other wavefunction-based methods is that DFT relies only on the total system density, $\rho(\mathbf{r})$, to compute the total ground state energy of the system instead of the wavefunction, $\Psi(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n)$. This leads to dramatic reduction in dimensionality, enabling mathematical tractability at multiple levels.

The premise that the ground state electronic energy of a system can be written in terms of the ground state density is based two important theorems proposed by Pierre Hohenberg and Walter Kohn.[11]

Theorem 1 *The external potential $v_{ext}(\mathbf{r})$, and hence the total energy, is a unique functional of the electron density $\rho(\mathbf{r})$.*

Theorem 2 *The ground state energy can be obtained variationally: the density that minimized the total energy is the exact ground state density.*

As theorem 1 states, there is a one-to-one mapping between electron density and external potential. This should,in principle, determine the ground state wavefunction which is now

only a functional of the ground state density:

$$E[\rho(\mathbf{r})] = \langle \Psi_0[\rho(\mathbf{r})] | \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} | \Psi_0[\rho(\mathbf{r})] \rangle \tag{1.29}$$

$$= T_e[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + \int v_{ne}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{1.30}$$

where $\hat{T}_e$ and $T[\rho(\mathbf{r})]$ are the electronic kinetic energy operator and functional respectively, $\hat{V}_{ee}$ and $V_{ee}[\rho(\mathbf{r})]$ are the electron-electron repulsion potential energy operator and functional respectively, and $\hat{V}_{ne}$ and $v_{ne}(\mathbf{r})$ are the electron-nuclear potential energy operator and functions respectively. The $V_{ee}[\rho(\mathbf{r})]$ functional in Eq. (1.30) can further be split into a classical Coulomb part ($J[\rho(\mathbf{r})]$) and a non-classical exchange part ($K[\rho(\mathbf{r})]$):

$$E[\rho(\mathbf{r})] = T_e[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + K[\rho(\mathbf{r})] + \int v_{ne}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{1.31}$$

$$J[\rho(\mathbf{r})] = \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \tag{1.32}$$

While the Hohenberg-Kohn theorem 1 gives us a way to express the ground state energy in terms of the ground state density, it does not provide an exact form for the kinetic energy and non-classical exchange energy functionals. These functionals have remained unknown till date. The magnitude of kinetic energy is usually the same as that of the total energy. Even quite good approximations to the kinetic energy functional leads to very large errors in relative energies making DFT in this original formalism ill-suited for chemical applications. Another obstacle to application of DFT is that only densities which come from integrating ground state wavefunctions are valid. This is widely referred to as the v-representability problem. These two limitations have prevented usage of DFT in its original formalism, also referred to as orbital-free DFT.[12]

The Kohn-Sham (KS) formulation of DFT reintroduces the concept of orbitals in order to circumvent the problem of the unknown kinetic energy functional. It describes a fictitious reference system comprising of n non-interacting electrons in n spin orbitals which give the same density. The kinetic energy of these spin orbitals, called KS spin orbitals, can be computed similar to that of HF theory:

$$T_s[\rho(\mathbf{r})] = \sum_{\sigma}^{\alpha,\beta}\sum_{i}^{n} \langle \phi_{i,\sigma} | -\frac{1}{2}\nabla^2 | \phi_{i,\sigma} \rangle \tag{1.33}$$

These KS orbitals give the same density $\rho(\mathbf{r})$, that is:

$$\rho(\mathbf{r}) = \sum_{\sigma}^{\alpha,\beta}\sum_{i}^{n} |\phi_{i,\sigma}(\mathbf{r})|^2 \tag{1.34}$$

The kinetic energy of this fictitious non-interacting system ($T_s[\rho(\mathbf{r})]$) is different from the actual kinetic energy of the real interacting system($T_e[\rho(\mathbf{r})]$), but the difference is much

smaller relative to the magnitude of the total ground state energy. The non-classical part of the electron-electron interaction has also remained elusive. These two unknown contributions are usually clubbed together (Eq. (1.35)) and are referred to as the exchange-correlation (XC) functional, referring to that the fact it captures the exchange interaction and the correlation contribution to the total energy (Eq. (1.36)).

$$E[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + \int v_{ne}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + E_{\text{xc}}[\rho(\mathbf{r})] \tag{1.35}$$

$$E_{\text{xc}} = \big(T_s[\rho(\mathbf{r})] - T_s[\rho(\mathbf{r})]\big) + K[\rho(\mathbf{r})] \tag{1.36}$$

While KS-DFT is exact in theory, the functional form of $E_{\text{xc}}[\rho(\mathbf{r})]$ is unknown and has remained an active area of research in the last couple of decades.

The KS energy functional can be variationally optimized by differentiating with respect to the orbitals $\psi_{i,\sigma}(\mathbf{r})$ to obtain Hartree-Fock-like KS Fock operator for computing KS orbitals:

$$\hat{f}_\sigma^{\text{KS}}\psi_{i,\sigma}(\mathbf{r}) = \epsilon_{i,\sigma}\psi_{i,\sigma}(\mathbf{r}) \tag{1.37}$$

$$\hat{f}_\sigma^{\text{KS}} = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|\hat{\mathbf{r}} - \hat{\mathbf{R}}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' + v_{\text{xc},\sigma}(\mathbf{r}) \tag{1.38}$$

where

$$v_{\text{xc},\sigma}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho(\mathbf{r})]}{\delta\rho_\sigma(\mathbf{r})} \tag{1.39}$$

Like in HF theory, the KS orbitals in Eq. (1.38) can be expressed in a basis set to give the KS generalized eigenvalue problem:

$$\mathbf{F}_\sigma^{\text{KS}}\mathbf{C}_\sigma^{\text{KS}} = \mathbf{S}\mathbf{C}_\sigma^{\text{KS}}\boldsymbol{\epsilon}_\sigma^{\text{KS}} \tag{1.40}$$

Eq. (1.40) is solved iteratively using the SCF procedure outlined in Section 1.1 and consequently has a computational cost similar to that of HF theory. For a detailed discussion of KS-DFT, we refer readers to Refs. 13, 14, 15, and 16.

Given the low cost of KS-DFT in comparison to other wavefunction methods, multiple research groups in the last couple of decades have attempted to come up with approximate functional forms for the exchange-correlation functional using empirical, semi-empirical, and non-empirical methods. Each of these approximations, termed density functional approximations (DFAs), give different accuracies for different problems.[17–19] For instance, a DFA that gives good accuracy for non-covalent interaction energies might not necessarily predict good barrier heights. The development of a DFA typically follows the below sequential pattern:

1. Hypothesizing a functional form for $E_{\text{xc}}[\rho(\mathbf{r})]$ that satisfies certain criteria that should be satisfied by the exact functional. These functional forms usually contain one or more parameters.

Table 1.1: Density Functional Approximations (DFAs) classified into different rungs of the Jacob's ladder.

Rung no.	Rung name	Depends on	Example DFAs
5	Double hybrid	$\rho_\sigma(\mathbf{r}), \nabla\rho_\sigma(\mathbf{r}), \nabla^2\rho_\sigma(\mathbf{r}), \{\psi_{i,\sigma}(\mathbf{r})\}, \{\psi_{a,\sigma}(\mathbf{r})\}$	ωB97M(2), XYG3
4	Hybrid	$\rho_\sigma(\mathbf{r}), \nabla\rho_\sigma(\mathbf{r}), \nabla^2\rho_\sigma(\mathbf{r}), \{\psi_{i,\sigma}(\mathbf{r})\}$	B3LYP, ωB97X-V
3	Meta-GGA	$\rho_\sigma(\mathbf{r}), \nabla\rho_\sigma(\mathbf{r}), \nabla^2\rho_\sigma(\mathbf{r})$	TPSS, B97M-V
2	GGA	$\rho_\sigma(\mathbf{r}), \nabla\rho_\sigma(\mathbf{r})$	PBE, BLYP
1	LDA	$\rho_\sigma(\mathbf{r})$	SPW92, SVWN5

2. Determination of these unknown parameters empirically by fitting to highly accurate reference data generated from experiments or computations

3. Testing the transferability of these fitted parameters on datasets not used for fitting in order to assess the accuracy of the DFA

The enticing promise of finding the exact functional or good approximations to it, along with a lack of rigorous recipe for systematic improvement, has caused an explosion in the number of DFAs parametrized in the recent past. The number of DFAs parametrized for chemical applications alone exceeds a couple of hundred. John Perdew has organized DFAs hierarchically based on the different terms they depend on.[20] This hierarchical classification is called the "Jacob's ladder." As one moves up the Jacob's ladder from the Hartree world with no electron exchange or correlation to the heaven of chemical accuracy, one can expect an increase in the accuracy of the DFAs. This is only an expectation and definitely does not apply to all DFAs.

We will briefly discuss the different rungs of the Jacob's ladder shown in Table (1.1). The DFAs in the first rung of the Jacob's ladder, termed Local Density Approximation (LDA), depend only on the density ($\rho_\sigma(\mathbf{r})$) of the system. The first rung functionals discussed in this work, SPW92 and SVWN5, contain Slater exchange[21] and different parameterizations for the correlation part.[22, 23] The second rung is called the generalized gradient approximation (GGA) and depends on the density ($\rho_\sigma(\mathbf{r})$) and its gradient ($\nabla\rho_\sigma(\mathbf{r})$). PBE[24] and BLYP[25, 26] are two popular DFAs from this rung. The next rung, named meta-GGA DFAs, depend on the kinetic energy density ($\nabla^2\rho_\sigma(\mathbf{r})$) in addition to the density and its gradient. In principle, the functional space of GGA is completely contained in the meta-GGA functional space and one would expect meta-GGA DFAs to perform as well or better than GGAs. However, in practice, this is always not the case as larger flexibility in space can lead to over-parameterization and consequently poorer transferability. TPSS[27] and B97M-V[28] are two examples of meta-GGA functionals. The density functionals in the first three rungs of the Jacob's ladder are commonly referred to as local or semi-local density functionals.

Semi-local functionals suffer from the problem of self-interaction where each electron interacts with itself. HF theory, which has exact treatment of exchange interactions, does not

suffer from this problem. One solution to improve the performance of DFAs is to incorporate some fraction of HF exchange (also referred to as exact exchange) into DFAs. Such DFAs are called hybrid functionals and are classified as the fourth rung and their exchange-correlation functional forms have explicit dependence on occupied MOs ($\{\psi_{i,\sigma}(\mathbf{r})\}$). Hybrid functionals come in multiple flavors: (1) Global hybrids which contain a constant fraction of HF exchange like B3LYP[29] (2) Range-separated hybrids like ωB97X-V[30] which use DFT exchange when electrons are close by and HF exchange when electrons are well-separated. Hybrid functionals are widely used because of their increased accuracy, but addition of HF exchange adds significantly to the computational cost. A potential method to partially mitigate this additional computational cost is discussed in Chapter 2. In rung five, DFAs use some fraction of correlation energy from correlated wavefunction methods like MP2[31] or RPA[32]. XC functionals of DFAs in this rung have additional dependence on virtual orbitals ($\{\psi_{a,\sigma}(\mathbf{r})\}$). These functionals, called double hybrid density functionals, have a markedly increased computational cost as they rely on correlated wavefunctions. XYG3[33] and ωB97M(2)[34] are two examples of rung five DFAs. These functionals, placed at the top of the Jacob's ladder, are typically highly accurate.

Density functional theory, being an inherently mean-field-like formalism, cannot capture certain truly non-local physical effects like dispersion. Dispersion, which is a consequence of electron correlation, has a negative-definite contribution to the total energy of systems containing more than one electron. Dispersion has been incorporated into DFAs using both empirical and pseudo-empirical methods. The most popular of these methods is Stefan Grimme's DFT-D atom-atom empirical potential.[35–39] This DFT-D method accounts for dispersion by adding a negative-definite correction which is independent of electron density. Another approach to calculate dispersion correction in an arguably less empirical way is to use a non-local correlation kernel in the density functional approximation. The vdW-DF category of methods[40, 41] and the VV09[42] and VV10[43] non-local kernels belong to this approach.

Extensive work in the field of developing empirical density functionals has resulted in highly accurate DFAs which are widely applicable to a range of chemical systems. In particular, we would like to emphasize the combinatorial optimization approach to DFA development, which has lead the formulation of B97M-V[28] (meta-GGA), ωB97X-V[30] (range-separated hybrid GGA), and ωB97M-V[44] (range-separated hybrid meta-GGA) density functional approximations. These density functionals give very good accuracy for a range of properties.[18]

Correlated Wavefunction Methods

While density functional theory and multiple DFAs provide techniques to incorporate electron correlation, they do not provide a guideline for systematic improvement in a way that guarantees a more accurate answer. Correlated wavefunction methods provide an alternative to compute the correlation energy (Eq. (1.28)) that is missing in the Hartree-Fock solution. Usually, these methods build on top of the Hartree-Fock solution and provide a systematic

recipe to improve the quality of the wavefunction. Correlated wavefunction methods can be broadly classified into three categories: (1) many-body perturbation theory, (2) coupled-cluster theory, and (3) configuration interaction theory.

Perturbation theory is most suitable for solving problems which are only slightly different from ones that can be exactly or approximately solved. By carrying out perturbation theory to higher and higher orders, one can expect to systematically improve it, making it closer to the true solution. In the context of electron correlation, the correlation energy can be considered as a perturbation to the mean-field Hartree-Fock wavefunction. The electronic Hamiltonian can be split into two parts: (1) the HF Hamiltonian ($\hat{H}_{\text{HF}}$) which can be solved exactly and (2) the remaining terms ($\hat{V}$) which can be treated as a perturbation.

$$\hat{H} = \hat{H}_{\text{HF}} + \hat{V} \tag{1.41}$$

Rayleigh-Schrödinger perturbation theory, applied to solving this problem, results in Møller-Plesset perturbation theory.[31] Møller-Plesset perturbation theory can be carried out to higher orders to achieve higher levels of accuracy in the energy and wavefunction. Møller-Plesset perturbation theory carried out to the second order is abbreviated MP2 and is perhaps the most commonly used amongst all perturbation theory based methods. The MP2 energy correction to the HF wavefunction is given by:

$$E_{\text{MP2}} = -\frac{1}{4}\sum_{i,j}^{N_{\text{occ}}}\sum_{a,b}^{N_{\text{virt}}}\frac{(\langle\chi_i\chi_j|\chi_a\chi_b\rangle - \langle\chi_i\chi_j|\chi_b\chi_a\rangle)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{1.42}$$

where the first summation in i and j runs over all occupied orbitals and the second summation in a and b runs over all virtual orbitals. ϵ_i and ϵ_j are the orbital energies of occupied orbitals χ_i and χ_j, and ϵ_a and ϵ_b are the orbital energies of virtual orbitals χ_a and χ_b.

In terms of computational cost, MP2 scales as the fifth power of the system size limiting its application to chemical systems with typically less than 100 atoms. The MP2 correlation energy also converges only polynomially with respect to the size of the one-particle basis set, thus requiring very large basis sets for an accurate computation of correlation energy. This is in contrast to HF and DFT energies which converge exponentially with respect to the size of the basis set. The requirement of large basis sets for converging the correlation energy along with the steep scaling of computational cost makes MP2 far more costly than DFT. In order to minimize this, extrapolation methods, which can extrapolate the correlation energy computed using finite basis sets to the complete basis set (CBS) limit, have been formulated.[45, 46] The polynomial extrapolation method proposed by Helgaker et al. is used in Chapters 4 and 5 to compute reference interaction energies for the datasets.

The MP2 formulation presented above does not depend on any empirical parameters. Multiple empirical variants of MP2 like the Spin Component Scaled MP2 (SCS-MP2)[47] and Spin Opposite Scaled MP2 (SOS-MP2)[48] have been proposed for improvement in performance and/or reduction in computational cost. Orbital optimization[49] in the presence of MP2 correlation energy, along with proper regularization,[50, 51] leads to higher accuracy in energetics.

Coupled-cluster theory is a more robust method to introduce electron correlation on top of the Hartree-Fock wavefunction.[52–56] Coupled-cluster theory employs an exponential ansatz to parameterize correlation corrections to the HF wavefunction.

$$\Psi_{\mathrm{CC}} = e^{\hat{\mathbf{T}}}\Phi_{\mathrm{HF}} \tag{1.43}$$

$$e^{\hat{\mathbf{T}}} = \hat{\mathbf{1}} + \hat{\mathbf{T}} + \frac{1}{2}\hat{\mathbf{T}}^2 + \ldots = \sum_{i=1}^{\infty}\frac{1}{k!}\hat{\mathbf{T}}^k \tag{1.44}$$

where Ψ_{CC} is the coupled-cluster wavefunction, Φ_{HF} is the HF wavefunction, $\hat{\mathbf{T}}$ is the excitation operator which is defined as:

$$\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2 + \ldots + \hat{\mathbf{T}}_{\mathrm{N_{occ}}} \tag{1.45}$$

where $\hat{\mathbf{T}}_1$ is the one-electron excitation operator, $\hat{\mathbf{T}}_2$ is the two-electron excitation operator and so on. The $\hat{\mathbf{T}}_i$ operator acting on the HF reference wavefunction generates all the i^{th} excited Slater determinants. The action of $\hat{\mathbf{T}}_1$ and $\hat{\mathbf{T}}_2$ are shown below:

$$\hat{\mathbf{T}}_1\Phi_{\mathrm{HF}} = \sum_{i}^{N_{\mathrm{occ}}}\sum_{a}^{N_{\mathrm{virt}}} t_i^a\Phi_i^a \tag{1.46}$$

$$\hat{\mathbf{T}}_2\Phi_{\mathrm{HF}} = \sum_{i<j}^{N_{\mathrm{occ}}}\sum_{a<b}^{N_{\mathrm{virt}}} t_{ij}^{ab}\Phi_{ij}^{ab} \tag{1.47}$$

where Φ_i^a is the Slater determinant formed by exciting the electron from i^{th} occupied orbital to a^{th} virtual orbital, and Φ_{ij}^{ab} is the Slater determinant formed by exciting electrons from i^{th} and j^{th} occupied orbitals to a^{th} and b^{th} virtual orbitals respectively. t_i^a and t_{ij}^{ab} are the coefficients of the Slater determinants Φ_i^a and Φ_{ij}^{ab} and are commonly referred to as t-amplitudes.

In practice the $\hat{\mathbf{T}}$ operator is truncated at a finite order in order to make the computations of the relevant energy and wavefunction tractable. Using $\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2$, gives the coupled-cluster with singles and doubles (CCSD) model.[57, 58] CCSD scales as $\mathcal{O}(N_{\mathrm{occ}}^2 N_{\mathrm{virt}}^4)$, overall giving a sixth order scaling method. Addition of $\hat{\mathbf{T}}_3$ to the equation leads to the coupled-cluster with singles, doubles, and triples (CCSDT) model, but makes it too computationally expensive (CCSDT scales as $\mathcal{O}(N_{\mathrm{occ}}^3 N_{\mathrm{virt}}^5)$).[59] In order to circumvent this computational expense, Raghavachari et al. proposed the coupled-cluster with singles, doubles, and perturbative triples (CCSD(T)) which approximates the $\hat{\mathbf{T}}_3$ amplitudes using perturbation theory reducing compute cost to $\mathcal{O}(N_{\mathrm{occ}}^3 N_{\mathrm{virt}}^4)$.[60] Owing to its remarkable accuracy, CCSD(T) extrapolated to the CBS limit has been the method of choice for computing energies and other properties and is often referred to as the "gold standard" of quantum chemistry. In this work, CCSD(T) extrapolated to CBS limit has been used to generate reference interaction energies in Chapters 4 and 5. Even DFA parameters are often trained using CCSD(T) reference data.

A detailed discussion and derivation of coupled-cluster theory can be found in Refs. 61, 62, and 63. For an in-depth discussion about configuration interaction theory, we refer interested readers to Refs. 9 and 64.

1.2 Computational Cost

In this section, we will present a concise discussion of the computational cost of single determinant methods (HF and DFT) to approximately solve the Schrödinger equation. The computational cost of HF theory and DFT can be discussed together as both of them make use of the SCF procedure. Apart from the one-time construction of density independent quantities, the SCF procedure consists of two important steps: (1) Fock matrix construction (2) diagonalization of Fock matrix.

The most computationally expensive components of the Fock matrix are the computation of Coulomb matrix ($\mathbf{J}$), exact exchange matrix ($\mathbf{K}$), and exchange-correlation contributions ($\mathbf{V}_{\text{xc}}$). The two-electron integrals in a Gaussian basis set are sparse and one can leverage this sparsity to eliminate computing certain integrals. The Schwarz inequality can be utilized to screen for significant two-electron four-center integrals, thus speeding up the evaluation of the Coulomb and exchange matrices:[65]

$$| \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\xi \rangle | \leq | \langle \phi_\mu \phi_\mu | \phi_\lambda \phi_\lambda \rangle |^{1/2} | | \langle \phi_\nu \phi_\nu | \phi_\xi \phi_\xi \rangle |^{1/2} \tag{1.48}$$

The number of numerically significant integrals scales quadratically with the number of basis functions. Further, it is possible to accelerate the computation of the Coulomb matrix using algorithms like the continuous fast multiple method,[66] J-engine,[67], and Fourier transform Coulomb method.[68] Methods based on resolution-of-identity have been formulated for reducing the cost of computing the exact exchange matrix,[69–71] however, this still remains the bottleneck step in Fock matrix construction in HF theory and hybrid DFT. The exchange-correlation bit of the Fock matrix in DFT is usually evaluated on a real-space atom-centered grid. The numerical integration scheme introduced by Axel Becke,[72] along with pruning techniques for reducing the size of the grid without loss in accuracy, has significantly accelerated its evaluation.[73] Overall, utilizing integral and density matrix element screening, the construction of Fock matrix asymptotically scales as $\mathcal{O}(N^2)$, where N is the number of basis functions in the system.

The other computationally demanding step of the SCF procedure is the diagonalization step. This step scales as $\mathcal{O}(N^3)$. Despite its higher asymptotic scaling in comparison to Fock matrix construction, the time taken for the latter usually dominates the cost of the SCF procedure for medium-sized systems. Through the rest of this section, we will discuss techniques that reduce the computational cost of electronic structure methods and their implications on the accuracy of the computation.

SCF for Molecular Interactions

SCF for molecular interactions (SCF-MI) is an efficient approximation to the SCF procedure that is applicable to systems which can be partitioned into two or more fragments.[74–77] Typically, SCF-MI is used to treat systems which are made up of fragments that are not covalently bound. The SCF-MI procedure relies on the use of absolutely localized molecular

orbitals (ALMOs). In the ALMO scheme, molecular orbitals are expanded only in local subsets of atomic orbitals:

$$|\psi_{Xi}\rangle = \sum_{\mu \in X} |\phi_{X\mu}\rangle \, C_{\bullet Xi}^{X\mu\bullet} \tag{1.49}$$

where the MO $|\psi_{Xi}\rangle$ on fragment X is expanded only in terms of AOs located on fragment X, giving the MO coeffient matrix a block-diagonal structure. Minimizing energy with this constraint gives us an expression like the generalized eigenvalue problem of HF (Eq. (1.18)) or DFT (Eq. (1.40)) for each fragment constituting the system, instead of one equation for the entire system:

$$[\mathbf{F}]_{\mathrm{XX}}[\mathbf{C}]_{\mathrm{XX}} = [\mathbf{S}]_{\mathrm{XX}}[\mathbf{C}]_{\mathrm{XX}}[\epsilon]_{\mathrm{XX}} \tag{1.50}$$

$$[\mathbf{F}]_{\mathrm{XX}} = (\mathbf{I} - \mathbf{SP} + \mathbf{SP}^{\mathrm{X}})\mathbf{F}(\mathbf{I} - \mathbf{PS} + \mathbf{P}^{\mathrm{X}}\mathbf{S}) \tag{1.51}$$

$$\left(\mathbf{P}^{\mathrm{X}}\right)^{Z\nu,X\mu} = (\mathbf{T})_{\bullet Zl}^{Z\nu\bullet}(\boldsymbol{\sigma}^{-1})^{Zl,Xj}(\mathbf{T}^{T})_{Xj\bullet}^{\bullet X\mu} \tag{1.52}$$

where $\mathbf{F}$ is the full system Fock matrix in the AO basis, an XX subscript indicates that it is referring to the XX$^{\mathrm{th}}$ block of the corresponding matrix. $\mathbf{P}^{\mathrm{X}}$ is the projector onto the X$^{\mathrm{th}}$ fragment given by Stoll et al.[74, 77] The SCF-MI procedure was originally proposed as a basis set superposition error free alternative to computing interaction energies.[75, 78] SCF-MI also serves as a computationally inexpensive alternative to the SCF procedure as it does not require diagonalization of the full Fock matrix in AO basis. It only requires diagonalization of a projected Fock matrix ($[\mathbf{F}]_{\mathrm{XX}}$) on each fragment, X, whose dimension is $N_X \times N_X$, where N_X is the number of AO basis functions in fragment X. Thus, it provides a more efficient alternative to SCF. Charge transfer is the only interaction SCF-MI does not capture at the single determinant level. The missing charge transfer interactions can be captured approximately using perturbation theory which provides reasonable accuracy for weakly interacting systems.[77]

Embedding theory

Embedding theory is another alternative that can be used to reduce the computational cost of quantum chemistry methods for a certain class of systems. Consider a system performing a chemical transformation that is too large to be studied using the desired method as it is too computationally expensive. Let the chemical transformation be confined to a small part of this system called the "active site" (denoted as 'A'). The remaining part of the system is not directly involved in the chemical transformation, but sufficiently modulates the active site, and thus cannot be neglected when studying this transformation. This remainder of the system is called "environment" (denoted as 'E'). Embedding theory provides a way to treat systems which can be partitioned into active site and environment by allowing the usage of highly accurate but expensive methods for treating the active site and approximate and inexpensive methods to treat the environment by capturing the coupling between them.

Enzymes with active sites confined to a few amino acids are a good example of systems that can be treated using embedding.

Different embedding methods differ in the way they treat the coupling between the active site and environment. Each of them also allows only certain methods to be used for the active site and/or environment. The hybrid quantum mechanics/molecular mechanics (QM/MM) is an embedding method that uses quantum mechanical approaches to treat the active site and molecular mechanics techniques to treat the environment.[79–83] One way in which QM/MM energy is computed is as follows:

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}}(A) + E_{\mathrm{MM}}(A+E) - E_{\mathrm{M}}(A) \tag{1.53}$$

where $E_{\mathrm{QM}}(A)$ is the energy of the active site computed using the quantum mechanical method, $E_{\mathrm{MM}}(A+E)$ is the energy of the full system computed using the molecular mechanics method, and $E_{\mathrm{MM}}(A)$ is the energy of the active site computed using the molecular mechanics method. Our own n-layered integrated molecular orbital and molecular mechanics (ONIOM) is a generalized framework that resembles the QM/MM approach but allows any two methods to be used for the active site and the environment.[84–86] The Fragment Molecular Orbital approach developed by Kitaura and co-workers is another embedding approach that explicitly allows the environment to influence the active site through Coulombic interaction.[87–89] Multiple methods have been proposed for embedding wavefunction methods in DFT.[90–95] As hybrid DFT methods are more accurate and expensive than semi-local DFT, embedding hybrid DFT in semi-local DFT is an interesting approach to treat large chemical systems otherwise inaccessible using hybrid DFT methods. The embedded mean-field theory (EMFT) approach and its variants provides a framework for this by partitioning the system and the environment at the one-particle basis set level.[96, 97] EMFT is one of the main constituents of the fragment-based approach presented in Chapter 2.

Many-Body Expansion

The many-body expansion (MBE) is a divide-and-conquer approach in computational chemistry. It provides a recipe for partitioning a large system into fragments, computing the energies of fragments, and reassembling these fragment energies to compute the energy of the full system. According to MBE, for a system consisting of F fragments, the total energy

of the full system (E_{total}) can be written as a sum of F terms:

$$E_{\text{total}} = V_1 + V_2 + V_3 + \ldots + V_F \tag{1.54}$$

$$V_1 = \sum_{i=1}^{F} E_i \tag{1.55}$$

$$V_2 = \sum_{i<j}^{F} (E_{ij} - E_i - E_j) \tag{1.56}$$

$$\begin{aligned} V_3 = \sum_{i<j<k}^{F} & [(E_{ijk} - E_i - E_j - E_k) \\ & - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k)] \end{aligned} \tag{1.57}$$

where V_1 is the one-body term, V_2 is the two-body term, V_3 is the three-body term and so on. E_i is the energy of the i^{th} fragment, E_{ij} is the energy of the dimer consisting of fragments i and j, and E_{ijk} is the energy of the trimer consisting of the fragments i, j, and k. The MBE presented in Eq. (1.54) is exact when all the F terms are included. MBE also presents an "embarrassingly parallel" approach to computing the properties of large systems as the computation of each of the terms is independent of the other terms. Such approaches are very well-suited for deployment on modern day supercomputers, which consist of thousands of processors. However, if all the terms of the MBE have to be computed, MBE does not provide any savings in computational cost in comparison to computing the energy of the full system directly. When the MBE is truncated at a finite order $k < F$, the truncated MBE can, however, provide savings in computational cost. Truncated MBEs are not exact anymore:

$$E_{\text{total}} \approx \sum_{i=1}^{k} V_i \quad k < F \tag{1.58}$$

The accuracy of the truncated MBE depends on the order of truncation (k), the chemical system and the major mechanisms governing the interactions within it, and the scheme of partitioning. Some of the missing effects of the $V_{k+1}, V_{k+2}, \ldots, V_F$ can be folded into the lower order terms ($\sum_{i=1}^{k} V_i$) by using embedding. In such embedded MBEs, each of the lower order terms is treated as an embedding problem: each k-mer as the active site and the remaining $(F-k)$ fragments as the environment. Embedded MBEs accelerate the convergence of the MBEs allowing for lower k truncations without significant loss in accuracy. The simplest of these embedded MBEs is the electrostatically-embedded MBE which captures the Coulombic effects of the $(F-k)$ fragments on the k-mer using point charges placed at the nuclei of the environment fragments.[98, 99] More sophisticated methods like the variational many-body expansion incorporate other effects like dispersion in addition to electrostatics using 3D real densities instead of just point charges.[100]

While these advances in algorithms and independent increase in the speed of computer hardware have made it possible to use these methods for medium-sized systems with hundreds of atoms, employing these methods on systems that are relevant from a molecular biology and material science perspective is still unattainable. Taking a step in that direction, we have proposed the polarized many-body expansion (PolBE) in Chapter 2, an algorithm that combines the advantages of SCF-MI, MBE, and EMFT to form an embedded many-body expansion.

1.3 Intermolecular Interactions and Energy Decomposition

Intermolecular interactions dictate a plethora of chemical phenomena. They are also ubiquitously present in systems of chemical and biological importance and play a significant role in influencing their activities. Intermolecular interaction refers to the attractive or repulsive energies that act between molecules governing their collective behavior. They also span several orders of magnitude in terms of interaction strength, ranging from weak London dispersion forces in noble gases (several kJ/mol) to strong donor-acceptor interactions in transition metal-ligand complexes (several hundred kJ/mol).

Intermolecular interactions play a crucial role in many chemical and biological processes and understanding them in detail sometimes gives us the ability to control and regulate them. For example, understanding the interactions of protein with small drug-like molecules helps us design more potent drugs. They also play a crucial role in molecular and surface catalysis and a thorough understanding of their specific roles can guide the rational design of better catalysts.[101, 102] Intermolecular interactions, specifically hydrogen bonding, are responsible for the liquid state of water at room temperature, and consequently for enabling life on Earth. They also govern surface phenomena and have been recently utilized to design porous materials capable of gas adsorption and separation.[103–106] Chapters 4 and 5 of this work concern validation of density functionals for the design of porous materials for storing H_2 gas. Control of these intermolecular interactions can also influence binding, molecular recognition, and self-assembly in materials.

Experimentally, gas phase intermolecular interactions can be measured using molecular beam experiments and by observing changes in structural and vibrational properties. Computational tools can also be used for calculating intermolecular interactions, their strengths, and trends. The advantage of computational methods is that they allow for isolating and probing specific interactions which may not be possible in experiments. Computing intermolecular interactions, being one of the most common applications of quantum chemistry methods, have prompted the development and improvement of many techniques. For example, addition of Grimme's DFT-D correction to density functional approximations increases their accuracy for non-covalent interactions.

The "supermolecular" approach is used for computing interaction energies at 0K between

molecules. For a system consisting of F fragments, the interaction energy ΔE_{INT} is computed as:

$$\Delta E_{\text{INT}} = E_{\text{TOTAL}} - \sum_{i=1}^{F} E_i \tag{1.59}$$

where E_{TOTAL} is the total energy of the entire system consisting of all fragments and E_F is the energy of the F^{th} fragment. $F + 1$ single point energy evaluations are required to compute the interaction energy using a specific model chemistry. A model chemistry is a specification of the electronic structure method (HF, DFA, MP2, etc.,) and the one-particle basis set to be used for the computation. Each of these methods gives different accuracies for intermolecular interactions. Choice of model chemistry depends on the desired accuracy, size of the system, and availability of computational resources. Over the years, computational chemists have also developed a deep understanding of the accuracies of different quantum chemistry methods for intermolecular interactions by extensive benchmarking on several representative datasets.[18] In Chapters 4 and 5, we have assembled a dataset representing H_2 interactions with binding motifs present in porous materials and benchmarked several DFA to assess their accuracy for predicting H_2 binding energies.

Physical Components of Intermolecular Interactions

Despite the huge diversity in the strength and character of intermolecular interactions, all of them are composed of a few fundamental physical components. Understanding the intricate interplay of different physical components in intermolecular interactions can help understand their origin which can further guide rational design of molecules with desired properties. Chemists usually understand concepts by attributing them to functional groups. Decomposition of intermolecular interaction energies into physical components supplement this assignment.

Different physical components are also characterized by their strength, distance dependence, and orientation profiles. For example, the electrostatic interaction is a strong long-ranged interaction. In contrast, the strength of charge transfer interaction, which is short-ranged, is heavily dependent on the chemical species and their relative orientations. Common physical components that occur in intermolecular interactions are London dispersion forces, Pauli repulsions, permanent electrostatics, polarization, and charge transfer. We will discuss some of these physical components in detail here:

Permanent electrostatics: Permanent electrostatics is the Coulombic interaction between charges or charge distributions in molecules and can be either attractive or repulsive. For example, the attractive interaction between the partial negative charge on oxygen atom in one water molecule and the partial positive charge on hydrogen atom in another water molecule in a water dimer complex can be categorized as permanent electrostatics. Permanent electrostatics show polynomial distance dependence. A charge-charge interaction decays as $1/R$ where R is the intermolecular distance. As the multipole moment increases, decay with R

also becomes faster. For example, a charge-dipole interaction decays as $1/R^2$. Permanent electrostatics are usually evaluated using Coulomb's law and the multipole expansion.

Dispersion: Dispersion, or Van der Waals' force, is a purely quantum mechanical effect that arises as a consequence of the non-classical character of electrons. Dispersion is an attractive force of interaction between instantaneous dipoles and decays as $1/R^6$ with distance R. Although weak in comparison to permanent electrostatics, dispersion interactions are always present. It is the dominant mechanism of interaction for certain cases like attraction between noble gases and π-stacked systems. HF and traditional DFT without dispersion corrections cannot capture the effect of dispersion. Dispersion-corrected DFT and correlated wavefunction methods like MP2 and CCSD(T) can capture the effects of dispersion with varying degrees of accuracy.

Polarization: Polarization is the response of the wavefunction of a fragment to the presence of other fragments. It is accompanied by reorganization of electron density within the fragment in response to electric fields of the rest of the complex in a non-overlapping regime. In the overlapping regime, polarization also includes the effects of density reorganization due to the Pauli repulsion principle. Polarization is also a truly many-body effect and gives rise to cooperativity in molecular complexes containing more than two fragments. Distance dependence of polarization energy can be approximated using classical electrostatics and multipole expansion arguments. The monopole induced-dipole interactions decay as $1/R^4$ and dipole induced-dipole interactions decay as $1/R^6$. Generally, all electronic structure methods that allow for distortion of electron density can capture polarization.

Pauli repulsion: Pauli repulsion is a truly non-classical effect that arises from Pauli's exclusion principle which states that "no two electrons can have the same spatial and spin quantum numbers." Pauli repulsion arises from the antisymmetric requirement of the electronic wavefunctions and is always repulsive in nature. This repulsion, in addition to permanent electrostatics, prevents molecules from collapsing into each other by balancing out the attractive forces. Pauli repulsion is a short-range effect and decays exponentially in that range. However, in practice, the repulsive wall is usually modelled using a polynomial with a large power in order reduce the cost of computation.

Charge transfer: Charge transfer (CT) is the energy lowering that is accompanied by transfer of electron density from one fragment to another. CT interaction is a short-ranged effect as it is usually controlled by orbital interactions. CT is also generally orientation dependent and is one of the major contributors to orientation dependence of the total interaction energy. The pairwise additive nature of CT has been demonstrated earlier[107] and again in Chapter 3. The polarized many-body expansion takes advantage of the inherent pairwise-additivity of CT to truncate the embedded many-body expansion series at the 2^{nd} order.

Energy Decomposition Analysis

Physical components of interaction energy are not quantum mechanical observables as they do not have operators associated with them; however, it is still useful to decompose the

total interaction energy into physically meaningful components. As these decomposed components have different rates of decay with distance and different orientation dependence, fitting force field parameters to the decomposed components instead of the total interaction energy has emerged as a novel technique for training force-fields.[108, 109] Energy decomposition analyses (EDA) is a class of methods that can decompose interaction energies into these physically meaningful components.

As these components do not correspond to physical observables, their definitions are not unique, especially in the overlapping regime. Multiple groups have defined different EDA schemes in the past couple of decades. The Kitaura-Morokuma (KM)-EDA scheme is one of the earliest and it decomposes the HF interaction energy into electrostatic, exchange repulsion, polarization, charge transfer, and an unassigned mixed interaction contribution.[110, 111] Ziegler and Rauk proposed a similar EDA scheme for HF that was later generalized to DFT.[112–115] The natural EDA (NEDA) approach has a self-energy and a core term that describes Pauli repulsion.[116, 117] This scheme divides the interaction energy into electrostatics, Pauli term, and orbital interaction term. Symmetry-adapted perturbation theory, which was initially proposed as an alternative method to the supermolecular approach for computing interaction energies, has also been used as an EDA method.[118, 119] For a detailed discussion of different EDA schemes and some limited comparative assessments, we refer interested readers to Refs. 120, 121, and 122.

Variational EDA is another class of EDA schemes which decomposes interaction energy based on variationally optimized wavefunctions. The block-localized wavefunction (BLW)-EDA[123–125] and absolutely localized molecular orbital (ALMO)-EDA[107, 126, 127] schemes belong to this category. The ALMO-EDA scheme decomposes interaction energy into geometry distortion (ΔE_{GD}), frozen interaction energy (ΔE_{FRZ}), polarization energy (ΔE_{POL}), and charge transfer energy (ΔE_{CT}) contributions:

$$\Delta E_{\mathrm{INT}} = \Delta E_{\mathrm{GD}} + \Delta E_{\mathrm{FRZ}} + \Delta E_{\mathrm{POL}} + \Delta E_{\mathrm{CT}} \tag{1.60}$$

The geometry distortion (GD) is the energy associated with distorting relaxed isolated fragment into their complex geometry. Frozen interactions (FRZ) is the energy associated with bringing two isolated fragments together without allowing for wavefunction relaxation. Polarization (POL) and CT have been defined in Section 1.3.

The ALMO-EDA scheme also includes a perturbative approach to further decompose CT energy and charge into pairwise additive terms.[107, 126] For example, considering a complex comprising of a charge donor and acceptor, this pairwise CT decomposition scheme can break down the CT interaction energy into a donor $\rightarrow$ acceptor component, acceptor $\rightarrow$ donor component, and a higher-order term. While successful for complexes with weak CT, this perturbation theory-based approach fails for strong CT complexes. In Chapter 3, we introduce a non-perturbative approach to pairwise CT decomposition that performs equally well for weak and strong CT complexes.

1.4 Designing Materials for Hydrogen Storage

In this section, we will discuss the application of quantum chemistry methods presented earlier for solving a specific problem in the domain of materials engineering: designing porous materials for hydrogen (H_2) storage using adsorption. We will discuss the specific parameters to consider while designing materials that are capable of storing H_2.

The use of H_2 gas as a source of energy has the potential to replace fossil fuels. H_2 is a clean and sustainable energy source and can potentially provide a way to decrease global warming-causing fossil fuel emissions. In order to be able to use H_2 as a reliable source of fuel, one must be able to store and transport sufficient quantities of H_2 in a safe and efficient manner. This is a major challenge as H_2 is a gas at room temperature and has very low volumetric and gravimetric energy density. Current technologies rely on high pressure storage of H_2 at 350 and 700 bar in tanks with thick walls. However, high-pressure storage and transportation of H_2 has serious limitations: (1) high-pressure tanks are very heavy and expensive and add significantly to the mass of the vehicle and to the cost of the fuel delivery system (2) H_2 is a combustible gas and can have disastrous consequences in case of accidents. An alternative to high pressure storage and transportation of H_2 is by adsorbing it on porous materials with high surface area. H_2 storage by adsorption alleviates the problems of high-pressure H_2 storage, significantly reducing costs and ensuring the viability of a H_2-based economy.

H_2 is a small molecule with only two electrons. It has no permanent monopole or dipole moment; the highest permanent multipole it possesses is a quadrupole. H_2 is also not a very polarizable molecule. All of these characteristics make H_2 reluctant to bind to surfaces. Porous materials have to be specifically engineered to bind H_2. Designing and characterizing such materials experimentally is a time-consuming and expensive process. Employing computational methods for designing such materials is a viable alternative. The H_2 storage community has slowly embraced computational methods for screening and designing porous materials capable of storing H_2. Recently, Farha et al. synthesized a promising simulations-motivated material capable of storing H_2 and methane at room temperature.[128] In this section, we will describe how electronic structure methods can help in the design of materials capable of storing H_2 at room temperature.

The ability of a material to store hydrogen is characterized by its usable capacity under two fixed pressures: the loading and unloading pressures. The usable capacity is defined as the amount of H_2 adsorbed at loading pressure that is released at the unloading pressure.

$$D(P_{\text{load}}, P_{\text{unload}}) = \theta(P_{\text{load}}) - \theta(P_{\text{unload}}) \tag{1.61}$$

where $D(P_{\text{load}}, P_{\text{unload}})$ is the usable capacity, $\theta(P_{\text{load}})$ is the amount of H_2 adsorbed at the loading pressure, and $\theta(P_{\text{unload}})$ is the amount of H_2 left in adsorbed form at the unloading pressure. Therefore, the problem of designing the ideal H_2 storage material can be recast as the problem of designing materials with highest usable capacity for fixed loading and unloading pressures. The U.S. Department of Energy recommends a loading and unloading

pressure of 100 and 5 bar respectively. The amount of H_2 adsorbed on a surface can be determined using the Langmuir adsorption isotherm:

$$\theta(P) = \frac{KP}{1 + KP} \tag{1.62}$$

where K is the equilibrium constant for adsorption. Substituting Eq. (1.62) into Eq. (1.61), we get

$$D(P_{\text{load}}, P_{\text{unload}}, K) = \frac{KP_{\text{load}}}{1 + KP_{\text{load}}} - \frac{KP_{\text{unload}}}{1 + KP_{\text{unload}}} \tag{1.63}$$

It is trivial to show that the optimal value of $D(P_{\text{load}}, P_{\text{unload}}, K)$ occurs when

$$K_{\text{opt}} = \frac{1}{\sqrt{P_{\text{load}} P_{\text{unload}}}} \tag{1.64}$$

Optimal value of the equilibrium constant gives an optimal value of Gibbs free energy of adsorption:

$$\Delta G_{\text{opt}} = -RT \log \frac{K_{\text{opt}}}{P_0} \tag{1.65}$$

where ΔG_{opt} is the optimal value of Gibbs free energy, R is the ideal gas constant, T is the temperature, and P_0 is the standard pressure (1 bar). Assuming a correlation between enthalpy and entropy that has been observed in multiple materials experimentally,[129–131] the optimal value of enthalpy of binding should lie in the range of -15 to -25 kJ/mol. This provides a design paradigm for engineering binding sites in porous H_2 storage materials.

DFT can be used to compute the binding enthalpy of H_2 with a binding motif. As the computational cost of DFT scales as the third power of the system size, it can be used in an *in silico* high-throughput screening setting to select potential binding sites/materials for experimental synthesis and characterization. DFT and EDA can also be used to enable rational design of binding sites with desired properties. This approach can potentially accelerate the discovery of materials with enhanced volumetric and gravimetric H_2 storage capacities. However, as discussed earlier, DFT does not refer to one method; it is a suite of approximations called density functional approximations. As each of these DFAs give different accuracies for predicting various properties, the best DFA for predicting the enthalpy of binding H_2 to different binding motifs present in porous materials is not known. We have addressed this problem in Chapters 4 and 5 by identifying DFAs which predict H_2 binding energies to a high degree of accuracy.

1.5 Outline

As DFT has been largely successful in computing various molecular properties with high accuracy for a significant portion of the chemically-relevant space, we have focused on improving different aspects of DFT in this work. In Chapter 2, we have attempted to push the

boundary of systems accessible to DFT computations by formulating a divide-and-conquer approach. In Chapter 3, we discuss a method that improves the interpretibility of DFT calculations. This method enables pairwise charge transfer analyses on a class of systems with strong charge transfer interactions, which was previously beyond of scope of this analyses. In Chapters 4 and 5, we have identified density functional approximations which work well for predicting H_2 binding energies to porous materials. A brief summary of each of the remaining chapters is presented here:

Chapter 2

The Many-Body Expansion (MBE) is a useful tool to simulate condensed phase chemical systems, often avoiding the steep computational cost of usual electronic structure methods. However, it often requires higher than 2-body terms to achieve quantitative accuracy. In this work, we propose the Polarized MBE (PolBE) method where each MBE energy contribution is treated as an embedding problem. In each energy term, a smaller fragment is embedded into a larger, polarized environment and only a small region is treated at the high-level of theory using embedded mean-field theory. The role of polarized environment was found to be crucial in providing quantitative accuracy at the 2-body level. PolBE accurately predicts non-covalent interaction energies for a number of systems, including CO_2, water, and hydrated ion clusters, with a variety of interaction mechanisms, from weak dispersion to strong electrostatics considered in this work. We further demonstrate that the PolBE interaction energy is predominantly pairwise unlike the usual vacuum MBE which requires higher-order terms to achieve similar accuracy. We numerically show that PolBE often performs better than other widely used embedded MBE methods such as the electrostatically embedded MBE. Owing to the lack of expensive diagonalization of Fock matrices and its embarrassingly parallel nature, PolBE is a promising way to access condensed phase systems with hybrid density functionals that are difficult to treat with currently available methods. This work has been published in *The Journal of Chemical Physics*.[132]

Chapter 3

Energy decomposition analysis (EDA) based on absolutely localized molecular orbitals (ALMOs) decomposes the interaction energy between molecules into physically interpretable components like geometry distortion, frozen interactions, polarization, and charge transfer (CT, also sometimes called charge delocalization) interactions. In this work, a numerically exact scheme to decompose the CT interaction energy into pairwise additive terms is introduced for the ALMO-EDA using density functional theory. Unlike perturbative pairwise charge-decomposition analysis, the new approach does not break down for strongly interacting systems, or show significant exchange-correlation functional dependence in the decomposed energy components. Both the energy lowering and the charge flow associated with CT can be decomposed. Complementary occupied-virtual orbital pairs (COVPs) that capture the dominant donor and acceptor CT orbitals are obtained for the new decomposi-

tion. It is applied to systems with different types of interactions including DNA base-pairs, borane-ammonia adducts, and transition metal hexacarbonyls. While consistent with most existing understanding of the nature of CT in these systems, the results also reveal some new insights into the origin of trends in donor-acceptor interactions. This work is under peer review in *Physical Chemistry Chemical Physics*.[133]

Chapter 4

Efficient and high capacity storage materials are indispensable for a hydrogen-based economy. In silico tools can accelerate the process of discovery of new adsorbent materials with optimal hydrogen adsorption enthalpies. Density functional theory is well-poised to become a very useful tool for enabling high-throughput screening of potential materials. In this work, we have identified density functional approximations that provide good performance for hydrogen binding applications following a two-pronged approach. First, we have compiled a dataset (H2Bind275) that comprehensively represents the hydrogen binding problem capturing the chemical and mechanistic diversity in the binding sites encountered in hydrogen storage materials. We have also computed reference interaction energies for this dataset using coupled cluster theory. Secondly, we have assessed the performance of 55 density functional approximations for predicting H_2 interaction energies and have identified two hybrid density functionals (ωB97X-V and ωB97M-V), two double hybrid density functionals (DSD-PBEPBE-D3(BJ) and PBE0-DH), and one semi-local density functional (B97M-V) as the best performing ones. We have recommended the addition of empirical dispersion corrections to systematically underbinding density functionals like revPBE, BLYP, and B3LYP for improvements in performance at negligible additional cost. We have also recommended the usage of the def2-TZVPP basis set as it represents a good compromise between accuracy and cost, limiting the finite basis set errors to less than 1kJ/mol. This work has been published in the *Journal of Chemical Theory and Computation*.[134]

Chapter 5

The availability of accurate computational tools for modeling and simulation is vital to accelerate the discovery of materials capable of storing hydrogen (H_2) under given parameters of pressure swing and temperature. Previously, we compiled the H2Bind275 dataset consisting of equilibrium geometries and assessed the performance of 55 density functionals over this dataset (Veccham, S. P.; Head-Gordon, M. *J. Chem. Theory Comput.*, **2020**, *16*, 4963–4982). As it is crucial for computational tools to accurately model the entire potential energy curve (PEC), in addition to the equilibrium geometry, we have extended this dataset with 389 new data points to include two compressed and three elongated geometries along 78 PECs for H_2 binding, forming the H2Bind78×7 dataset. Assessing the performance of 55 density functionals on this significantly larger and more comprehensive H2Bind78×7 dataset, we have identified the best performing density functionals for H_2 binding applications: PBE0-DH, ωB97X-V, ωB97M-V, and DSD-PBEPBE-D3(BJ). Addition of Hartree Fock exchange

improves the performance of density functionals, albeit not uniformly throughout the PEC. We recommend the usage of ωB97X-V and ωB97M-V density functionals as they give good performance for both geometries and energies In addition, we have also identified B97M-V and B97M-rV as the best semi-local density functionals for predicting H_2 binding energy at its equilibrium geometry.

Chapter 2

Making Many-Body Interactions Nearly Pairwise Additive: The Polarized Many-Body Expansion Approach

Reprinted with permission from Veccham, S. P.; Lee, J.; Head-Gordon, M. Making Many-Body Interactions Nearly Pairwise Additive: The Polarized Many-Body Expansion Approach. *J. Chem. Phys.* **2019**, *151*, 194101.

Introduction

Simulating condensed phase systems is considered the holy grail of quantum chemistry. However historically, the majority of quantum chemistry research has focused on molecular properties rather than bulk properties of liquids. The major hurdle towards calculating and predicting properties of bulk materials is the steep computational scaling in modeling them. For decades, a major research goal of the quantum chemistry community has been to formulate more accurate and inexpensive computational methodologies to predict properties of bulk materials.

For systems comprising of non-covalently bonded fragments, the many-body expansion method has been used as way to circumvent the steep hyper-linear scaling of quantum chemistry methods. The Many-Body Expansion (MBE) is a technique for calculating the properties of a large system by (1) partitioning the system into N fragments, (2) calculating fragment properties, and (3) reassembling them using the many-body expansion equation. This approach reduces a large supersystem calculation into a large number of much smaller fragment calculations, and each of them is completely independent of others. Thereby, MBE calculations are "embarrassingly parallel." While the MBE is in principle exact, it provides computational advantages only when it is truncated up to a certain order. Truncated

many-body expansions are not exact, and its accuracy can heavily depend on the underlying chemical interactions in the system. Some applications of the many-body expansion method are determination of lattice energies [135] and studying polymorphism in molecular crystals. [136] Many-body based cluster corrections to calculations with periodic boundary conditions have also served as powerful tools to study crystal structures using highly accurate wave-function techniques. [137] MBEs have also been used to understand and incorporate the effects of mutual polarization in classical molecular dynamics force fields.[138]

The convergence of MBE can be accelerated using embedding, in which each k-body term is calculated in the presence of the remaining $(N - k)$ fragments (hereafter referred to as the 'environment'). An active research question is then how one models this environment and its interaction with the k-body term at a given electronic structure level. Quantum embedding can be performed in the context of Green's functions, [139–142] densities or potentials,[95, 96, 143–147] or density matrices. [148, 149] In many chemical phenomena, the interesting chemical activity like binding or chemical reaction occurs only in a small part of the chemical system. Embedding techniques, taking advantage of this "nearsightedness"[150] in order to divide the system into two parts: A - the region of interest where an actual chemical phenomenon occurs and B - the environment which surrounds the system but does not directly take part in the actual chemical process. Even though the environment does not directly participate in the chemical phenomenon, it sufficiently modulates the chemical process occurring in the system that it cannot be neglected. Embedding can also be done by modelling the environment using molecular mechanics which constitute widely used QM:MM methods like ONIOM.[151] ONIOM and its extensions [152] have been extensively used to study a wide range of systems and phenomena from heterogeneous catalysis to biological macromolecules. [153] However, ONIOM does not contain any explicit system-environment coupling terms in the hamiltonian, and cannot accurately capture the effects of environment. Embedded MBEs (eMBE) make use of quantum embedding techniques to describe the k-body term using an accurate and often more expensive computational method while using a much cheaper method to represent the environment. An eMBE truncated at the 2-body level aims to reduce the relative importance of the higher order terms (3-body and higher) by folding in these many-body effects into the 1-body and the 2-body terms. Different models of the environment, which can capture distinct physical effects, are capable of encoding these many-body effects with varying degrees of accuracy and computational expense.

A simple way to go beyond vacuum MBE, where we do not include any environment, is to model the environment using atom-centered point charges, termed the Electrostatically-Embedded Many-Body Expansion (EE-MBE). [98, 154, 155] These atom-centered point charges incorporate some effect of electrostatic interaction while the finite-extent of electron density, exchange interaction, and polarization of the environment are neglected. EE-MBE is fairly simple to implement, however its accuracy is very dependent on the actual charges used for embedding. [156]

One way to go beyond a simple charge embedding model is to use the Embedded Many-Body Expansion of Manby and co-workers which uses atom-centered gaussians to represent electrostatic interactions of the environment and a simple empirical model for Pauli repulsion.

[157] A slightly more sophisticated embedding method is the Variational Many-Body Expansion (VMBE), [100] which builds upon the 1-body X-Pol wavefunction, [158–163] introduced by Jiali Gao and co-workers. In the X-Pol wavefunction, the supersystem wavefunction is written as a Hartree-product of the monomer wavefunctions. Inter-fragment coulombic interactions are simplified by using point charges in order to avoid computing two-electron integrals for the whole system. The X-Pol wavefunction can either be optimized variationally or by using a double self-consistent field procedure. [158] Dispersion, exchange repulsion, and charge transfer effects are approximated using empirical Lennard-Jones terms. A variation of the X-Pol wavefunction, X-Pol-X, can be used to incorporate the effect of exchange repulsion non-empirically. [164] While the X-Pol wavefunction treats these interactions empirically, VMBE makes an attempt to treat them quantum mechanically at the 2-body level by redefining the fragment in the X-Pol wavefunction to contain two monomers. It is important to note that the environment density is updated for each of the dimer calculations in order to include complete mutual polarization between the dimer and its environment. The Fragment Molecular Orbital (FMO) method is another idea based on the MBE, which uses densities obtained from monomer wavefunctions to represent the fragments in the environment. [165–168] It has been identified that the right choice of a zeroth-order wavefunction, a wavefunction that can be used to form an inexpensive model for the environment, can lead to significantly accelerated convergence of the many-body expansion. [95, 100] A good zeroth-order wavefunction is one that is able to incorporate the most important interactions present in the system in a computationally efficient manner.

Our choice for the zeroth order wavefunction is the Self-Consistent Field for Molecular Interaction (SCFMI) wavefunction, which was initially introduced as an efficient representation for weakly interacting systems. [74, 75, 77, 78, 169] SCFMI utilizes localized molecular orbitals and enforces fragment sparsity in the molecular orbitals. In SCFMI, molecular orbitals are expanded in local subsets of atomic orbitals and are called Absolutely Localized Molecular Orbitals (ALMOs). [74] Like other localized molecular orbitals, ALMOs can be used in chemical interpretation [123, 126, 170, 171] and formulation of reduced scaling methods using local correlation techniques.

Partitioning of AOs into subsets naturally arises for systems consisting of non-covalently bound entities like molecular clusters. ALMOs are orthogonal on fragment but non-orthogonal between fragments. ALMOs have been routinely used to perform interaction energy calculations which are basis set superposition error free,[75] employed to carry out computationally cheaper SCF alternative for weakly interacting systems, [77] and used to decompose interaction energy in energy decomposition analysis (i.e. ALMO-EDA) schemes. [123, 126, 170, 171] Unlike SCF, SCFMI avoids diagonalization of the full system Fock matrix which reduces the prefactor for linear algebra in comparision to an SCF calculation.

Moving forward using eMBE with SCFMI, another computational bottleneck will appear for Density Functional Theory (DFT) calculations. The most accurate hybrid Exchange Correlation (XC) functionals involve some amount of exact exchange, which in turn becomes a bottleneck for large-scale calculations.[18] One way to circumvent this is to further approximate the environment beyond SCFMI. We have noticed that it is sufficient to capture only

the important interactions between each of the computed terms and its surrounding in order to ensure expedited convergence of the many-body expansion. In this paper, we represent the 1-body and 2-body terms using hybrid density functionals with appropriate dispersion corrections because of their ability to accurately represent non-covalent interactions. The environment is represented using a semi-local functional due to its cheaper computational cost. The Embedded Mean-Field Theory (EMFT) developed by Manby and Miller is well-suited for embedding a hybrid density functional in a semi-local density functional. [96] EMFT is not only simple to implement, but it has also been shown to reduce the cost of exact exchange computation while maintaining accuracy. [96] EMFT treats the system and its environment as open quantum systems incorporating quantum entanglement at a mean-field level. The combination of eMBE, SCFMI, and EMFT will be referred to as PolBE.

This paper is organized as follows: the theoretical background behind the MBE will be introduced. ALMOs are defined and the procedure to obtain them self-consistently, namely SCFMI, will be described. The EMFT method and models for computing exact exchange will be presented. The performance of PolBE for several chemical systems will be assessed in comparison with other existing eMBE methods. Finally, we will discuss the computational cost of a preliminary implementation of PolBE.

2.1 Theory

Embedded Many-body Expansion (eMBE)

The Many-Body Expansion (MBE) for a system of F fragments is written as a sum of 1-body (V_1) , 2-body (V_2), ... , F-body (V_F) terms as shown in Eq. (2.1).

$$E_{\text{total}} = \sum_{i=1}^{F} V_i \tag{2.1}$$

where

$$V_1 = \sum_{i=1}^{F} E_i(B), \tag{2.2}$$

$$V_2 = \sum_{i<j}^{F} (E_{ij}(B) - E_i(B) - E_j(B)), \tag{2.3}$$

$$\begin{aligned} V_3 = \sum_{i<j<k}^{F} & [(E_{ijk}(B) - E_i(B) - E_j(B) - E_k(B)) \\ & - (E_{ij}(B) - E_i(B) - E_j(B)) - (E_{ik}(B) - E_i(B) - E_k(B)) - (E_{jk}(B) - E_j(B) - E_k(B))] \end{aligned} \tag{2.4}$$

and so on, $E_i(B)$ is the energy of the i-th monomer computed in the presence of the environment denoted as 'B', $E_{ij}(B)$ is the energy of dimers (ij) computed in the presence of the environment, and $E_{ijk}(B)$ is the energy of the trimers (ijk) computed in the presence of the environment, and so on.

Different eMBE methods differ only in their treatment of the environment. For example, in the Electrostatically-Embedded Many-Body Expansion, the 2-body term is $E_{ij}(B)$ is calculated in the presence of atom-centered point charges for all monomers $k \neq i, j$. This embedding accelerates the convergence of the MBE, in comparison to vacuum MBE in which there is no environment. PolBE represents the environment using the polarized SCFMI density. In this paper, unless otherwise stated, all the MBEs and the eMBEs have been truncated at the second order, applying the pairwise approximation.

SCFMI model for the environment

Absolutely Localized Molecular Orbitals (ALMOs) on a fragment X ($|\psi_{\mathrm{X}i}\rangle$) are linear combinations of atomic orbitals belonging to the same fragment ($|\phi_{\mathrm{X}\mu}\rangle$) as shown in Eq. (2.5).

$$|\psi_{\mathrm{X}i}\rangle = \sum_{\mu} |\phi_{\mathrm{X}\mu}\rangle C_{\bullet\mathrm{X}i}^{\mathrm{X}\mu\bullet} \tag{2.5}$$

where $\mathbf{C}$ is the MO coefficient matrix, in which all the MO coefficients $C_{\bullet\mathrm{X}i}^{\mathrm{Y}\mu\bullet}$ when $X \neq Y$ are constrained to be zero (i.e. fragment sparsity), giving the $\mathbf{C}$ matrix a block diagonal structure.

Solving for the MOs with the ALMO constraint gives the SCFMI equations for each fragment X as shown in Eq.(2.6).

$$[\mathbf{F}]_{\mathrm{XX}}[\mathbf{C}]_{\mathrm{XX}} = [\mathbf{S}]_{\mathrm{XX}}[\mathbf{C}]_{\mathrm{XX}}[\epsilon]_{\mathrm{XX}} \tag{2.6}$$

$$[\mathbf{F}]_{\mathrm{XX}} = (\mathbf{I} - \mathbf{SP} + \mathbf{SP}^{\mathrm{X}})\mathbf{F}(\mathbf{I} - \mathbf{PS} + \mathbf{P}^{\mathrm{X}}\mathbf{S}) \tag{2.7}$$

$$\left(\mathbf{P}^{\mathrm{X}}\right)^{Z\nu,X\mu} = (\mathbf{T})_{\bullet Zl}^{Z\nu\bullet}(\boldsymbol{\sigma}^{-1})^{Zl,Xj}(\mathbf{T}^{T})_{Xj\bullet}^{\bullet X\mu} \tag{2.8}$$

where $[\mathbf{F}]_{\mathrm{XX}}$ is the projected Fock matrix constructed as shown in (2.7), $[\mathbf{C}]_{\mathrm{XX}}$ is the ALMO coefficient matrix, $[\mathbf{S}]_{\mathrm{XX}}$ is the fragment AO overlap matrix, and $[\epsilon]_{\mathrm{XX}}$ contains the fragment orbital energies. The projected Fock matrix ($[\mathbf{F}]_{\mathrm{XX}}$) for each fragment is constructed using the density matrix ($\mathbf{P}$), full system AO overlap ($\mathbf{S}$) and, the Stoll projector $\mathbf{P}^{\mathrm{X}}$ as shown in Eq. (2.6). $\mathbf{T}$ is the occupied orbital coefficient matrix and $\boldsymbol{\sigma}$ is the occupied MO overlap matrix. We follow the projection operator form proposed by Stoll et al. shown in Eq. (2.8). [74] Diagonalization of the fragment Fock matrix, $[\mathbf{F}]_{\mathrm{XX}}$ of fragment AO dimension, yields MOs which are absolutely localized on that fragment.

The SCFMI wavefunction allows ALMOs on each fragment to polarize in the presence of ALMOs on all other fragments. In terms of physical interactions, the SCFMI wavefunction can capture electrostatics, exchange repulsion, dispersion (as permitted by the XC functional), and polarization between fragments as described by the mean-field theory used.

One important interaction *not captured* by the SCFMI wavefunction is the effect of charge transfer among fragments. One way to capture this charge delocalization effect is through perturbation theory, which works decently well for weakly interacting systems. [77] Another more expensive but systematic method to capture the effects of charge delocalization is the Generalized X-Pol method which treats dimer states in the spirit of non-orthogonal configuration interaction theory. [172] In terms of computational cost, SCFMI removes the diagonalization bottleneck involved in the SCF iterations by requiring only fragment Fock matrix diagonalizations. However, there is no cost saving in constructing the Fock matrix for each SCF iteration.

EMFT

The Embedded Mean-Field Theory (EMFT) is a simple embedding scheme for mean-field electronic structure theories, which partitions the system and the environment at the single-particle basis set level. [96] EMFT uses a more accurate and usually more expensive mean-field theory (referred to as 'high level') for the region of interest, while using an approximate and inexpensive mean-field theory (referred to as 'low level') for the rest of the system. The region of interest is usually one in which a chemical transformation takes place (denoted as A), and the surrounding is usually not directly involved in the chemical transformation (denoted as B). The EMFT energy is obtained by minimizing the EMFT energy function given by Eq. (2.9).

$$E_{\mathrm{EMFT}} = E_{\mathrm{low}}[\mathbf{P}] - E_{\mathrm{low}}[\mathbf{P}^{\mathrm{AA}}] + E_{\mathrm{high}}[\mathbf{P}^{\mathrm{AA}}] \tag{2.9}$$

where $E_{\mathrm{low}}[\mathbf{P}]$ is the low level energy computed using the full density matrix $\mathbf{P}$, $E_{\mathrm{low}}[\mathbf{P}^{\mathrm{AA}}]$ is the low level energy computed using the AA block of the density matrix, and $E_{\mathrm{high}}[\mathbf{P}^{\mathrm{AA}}]$ is the high level energy computed using the AA block of the density matrix. The core idea is to compute interactions in subsystem A, represented by the AA block of the density matrix $\mathbf{P}^{\mathrm{AA}}$, using the high-level mean-field theory. In the context of mean-field theories, hybrid density functionals may serve as the best choice for the high-level theory, while semi-local functionals would be suitable choices for low level. For this combination of hybrid density functional in semi-local density functional, the EMFT energy expression in Eq. (2.9) can be written as Eq. (2.10) where 'high' and 'low' denote the the hybrid density functional and semi-local density functional respectively, and α is the fraction of exact exchange present in the hybrid density functional.

$$E_{\mathrm{EMFT}} = E_{\mathrm{xc,low}}[\mathbf{P}] + \left(E_{\mathrm{xc,high}}[\mathbf{P}^{\mathrm{AA}}] - E_{\mathrm{xc,low}}[\mathbf{P}^{\mathrm{AA}}]\right) + \alpha E_{\mathrm{K}}[\mathbf{P}] \tag{2.10}$$

$$\mathbf{F}_{\mathrm{EMFT}} = \frac{\partial E_{\mathrm{EMFT}}}{\partial \mathbf{P}} \tag{2.11}$$

$$= \mathbf{F}_{\mathrm{xc,low}}[\mathbf{P}] + \left(\mathbf{F}_{\mathrm{xc,high}}[\mathbf{P}^{\mathrm{AA}}] - \mathbf{F}_{\mathrm{xc,low}}[\mathbf{P}^{\mathrm{AA}}]\right) + \alpha \mathbf{K}^{\mathrm{AA}}[\mathbf{P}] \tag{2.12}$$

The EMFT Fock matrix expression follows directly from the energy expression (shown in Eq. (2.12)). The EMFT paper reports two models for computing exact exchange, EX0 and

EX1 (shown in Eq. 2.13 and Eq. 2.14 respectively).

$$E_{\mathrm{K,EX0}}[\mathbf{P}^{\mathrm{AA}}] = -\frac{1}{4} \sum_{\mu,\nu,\sigma,\lambda \in A} \mathbf{P}^{\mathrm{AA}}_{\lambda,\mu} (\mu\nu|\sigma\lambda) \mathbf{P}^{\mathrm{AA}}_{\nu\sigma} \tag{2.13}$$

$$E_{\mathrm{K,EX1}}[\mathbf{P}^{\mathrm{AA}}] = -\frac{1}{4} \sum_{\substack{\nu,\sigma \\ \mu,\lambda \in A}} \mathbf{P}^{\mathrm{AA}}_{\lambda,\mu} (\mu\nu|\sigma\lambda) \mathbf{P}_{\nu\sigma} \tag{2.14}$$

The EX0 requires the construction of only the AA block of the exact exchange matrix ($\mathbf{K}$) from the AA block of the density matrix ($\mathbf{P}^{\mathrm{AA}}$), providing substantial savings in cost. However, this model of computing the exact exchange suffers from severe numerical instability due to unnatural density shift to the part of the system containing exact exchange (A), as shown in Fig. S1. This is caused, in part, due to the lack of balancing counterpart terms in the semi-local functionals to the negative-definite exact exchange contribution of $\alpha E_{\mathrm{EX0}}[\mathbf{P}^{\mathrm{AA}}]$ of hybrid DFT to the EMFT energy expression. In the original EMFT paper, [96] this is overcome by using the more expensive EX1 model for exact exchange which computes the exact exchange matrix using the full density matrix, instead of the just the AA block. Another technique to alleviate this problem, while still using the the EX0 model, comprises of doing the computations in AO basis orthogonal between A and B. [97] This approach in turn breaks the fragment sparsity in $\mathbf{C}$ and is therefore not suitable for our purposes. In this work, we employ the EX0 model for computing exact exchange but constrain the particle number on each fragment by solving for the SCFMI equations instead of the full SCF equations. This fixes the numerical instability problem in EMFT. While the problem of unphysical collapse in the wavefunction is solved, particle number fluctuation (i.e. charge delocalization) is allowed at a two-body level.

PolBE theory

As mentioned before, we have identified that the SCFMI wavefunction is well-suited to achieve the delicate balance between computational feasibility and incorporating many-body effects into lower order terms. We use EMFT to treat the system using an accurate, expensive mean-field theory while treating the environment using a more approximate and relatively inexpensive mean-field theory. Further computational advantages can be achieved by using a smaller basis set for the environment. For the rest of the paper, we will use XC1 to denote the high level functional and XC2 to denote the low level functional. Similarly, BS1 will refer to the larger basis set and BS2 to the smaller basis set.

1. The first step in PolBE is to calculate the polarized density of the full system at XC2/BS2. The MO coefficient matrix, illustrated in Fig. 2.1(a), is completely represented in BS2. This approximate polarized density is used for embedding all k-body terms treated using XC1/BS1.

2. For a system consisting of F fragments, each 1-body term is calculated at XC1/BS1 in the presence of the approximate polarized density of the remaining $(F-1)$ fragments.

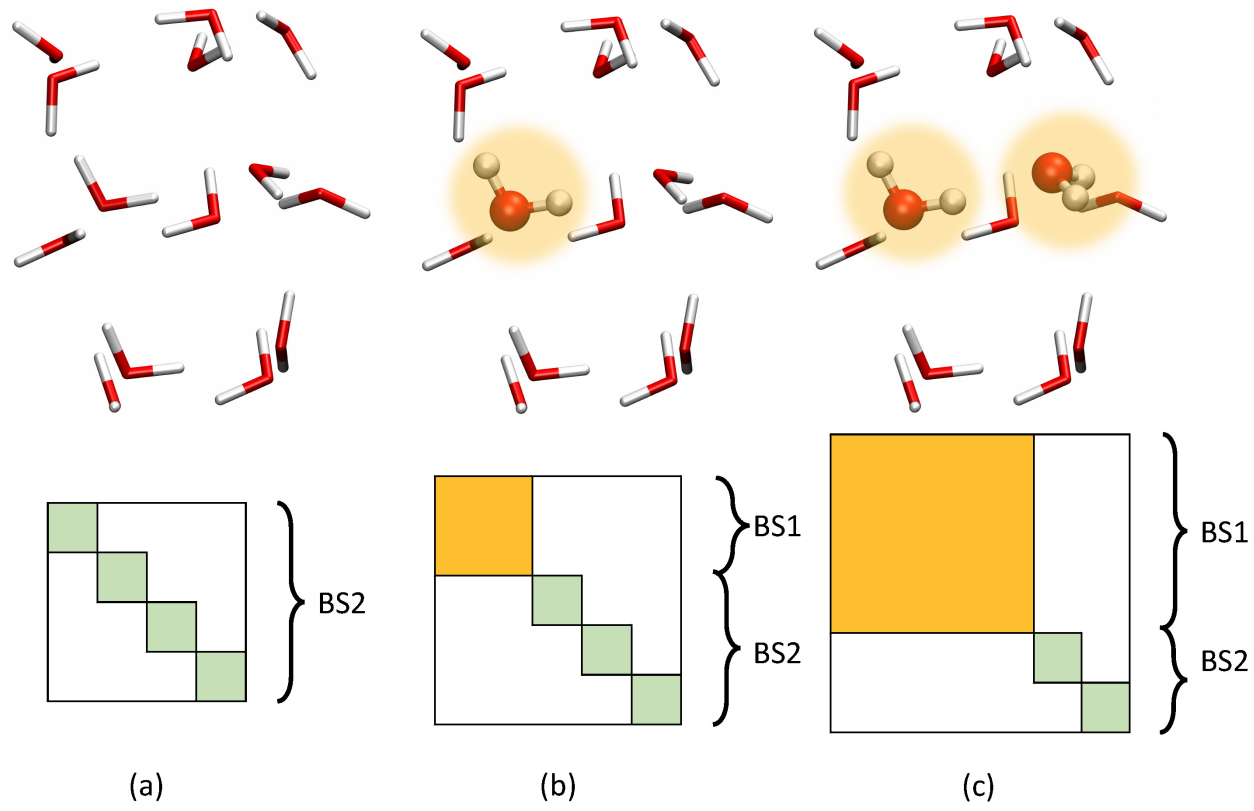


Figure 2.1: A schematic of the PolBE algorithm. The top panel schematically shows the density functional and basis set used for each of the monomers. Highlighted monomers are treated at XC1/BS1, and the others are treated at XC2/BS2. The bottom panel shows the structure of the MO coefficient matrix and the size of each of the blocks (a) V_0: All fragments are treated at XC2/BS2 (b) V_1: One of the fragments is treated at XC1/BS1 while the others are treated at XC2/BS2 (c) V_2: Two of the fragments are treated at XC1/BS1 while the others are treated at XC2/BS2

Fig. 2.1(b) shows one of the F 1-body terms in which the monomer is represented in BS1 while the rest of the environment is represented in BS2. In each monomer calculation, EMFT with EX0 is used to build Fock matrix treating the monomer at XC1 while the rest of environment is treated at XC2. Instead of diagonalizing the full Fock matrix, SCFMI is used to construct a projected Fock matrix only for the monomer fragment which is diagonalized, and its density is updated until convergence is reached. It should be emphasized that the density of the environment is not updated.

3. At a 2-body level, dimers are formed by combining all possible monomers into a larger fragment and treated at XC1/BS1 as described earlier. One of the ${}^{N}C_{2}$ dimers is shown as a large fragment consisting of two monomers and is represented in BS1 as shown in Fig. 2.1(c).

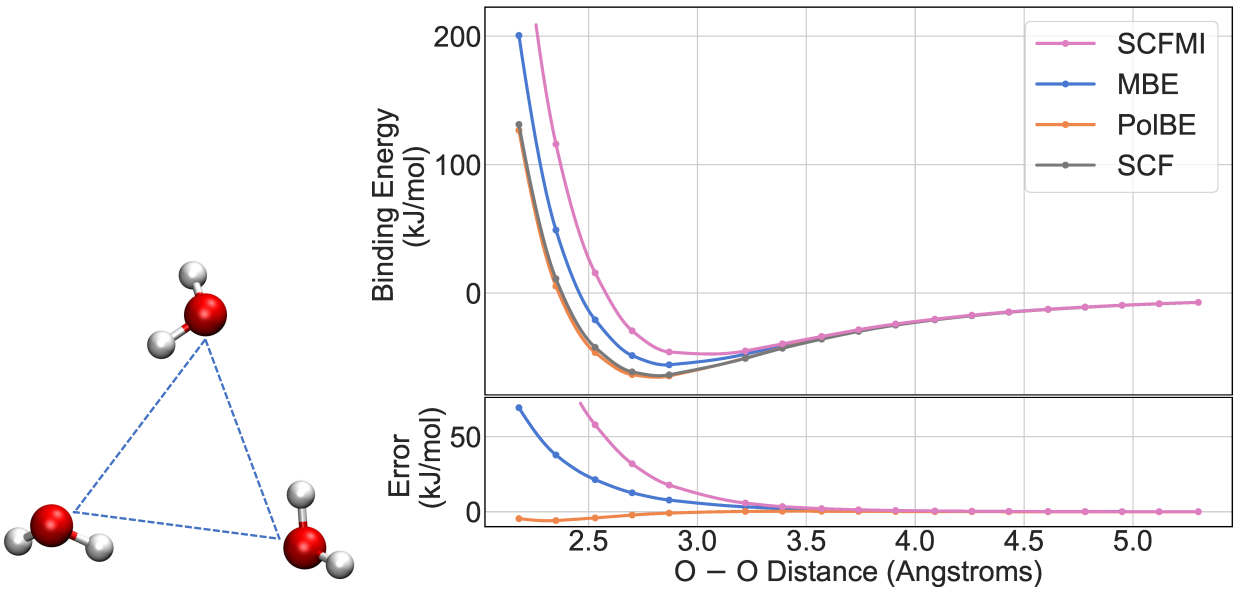
2.2 Computational Details

ωB97M-V, a combinatorially optimized range-separated hybrid meta-GGA functional with VV10 non-local correlation, [44] has been known to provide very good performance for a range of properties, including binding energies for non-covalent dimers and clusters. [18] We used this functional to represent the system denoted earlier as XC1. We used PBE, a more inexpensive and less accurate GGA functional, to represent the environment (XC2). [24] Def2-TZVPPD, a triple-zeta basis set with polarization and diffuse functions was used for BS1.[173] Def2-SV(P), a double-zeta split-valence basis set with polarization functions only on heavy atoms was used for BS2. [174] A quadrature grid consisting of 99 Euler-MacLaurin radial points and 590 Lebedev angular points was used for integrating both the XC functionals. SG-1 was used for the non-local VV10 part.[175] Binding energies reported below are vertical binding energies, computed as the difference between the energy of the full system and the sum of fragment energies at the super system geometry. The reference SCF binding energy was computed using XC1/BS1 (ωB97M-V/def2-TZVPPD); the goal of PolBE is to accurately reproduce the full SCF energy of the supersystem. PolBE was implemented in the new SCF manager of Q-Chem[176] which was developed over Refs. 51, 177, 178, and 179.

2.3 Results and Discussion

Water Trimer PES

A desirable property of an embedded Many-Body expansion is that it should provide a smooth Potential Energy Surface (PES) and consequently a continuous nuclear gradient so that it can used for performing geometry optimizations, transition state searches, and molecular dynamics simulations. [180, 181] In order to evaluate the performance of PolBE as a function of nuclear position, we have investigated the PES of water trimer. PolBE relies on



(a) (b)

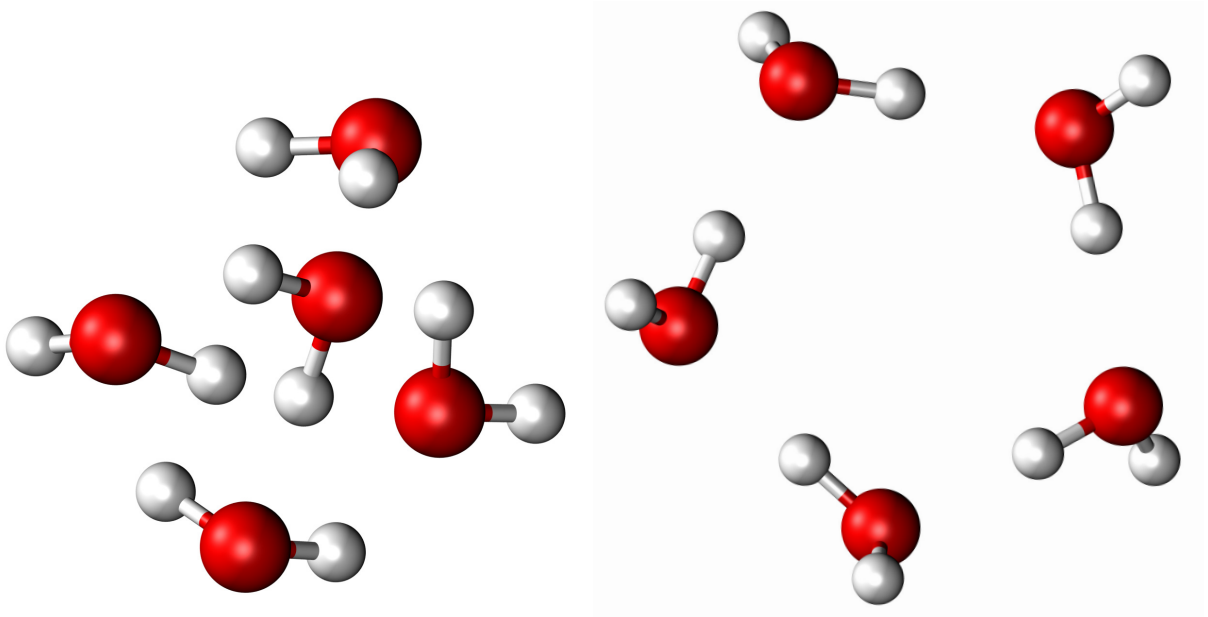
Figure 2.2: (a) Water trimer formed by stretching molecules away from each other (b) Water PES as computed by full SCF, vacuum MBE, and PolBE. Associated errors are shown in the bottom panel.

natural fragmentation arising from the lack of covalent bonding between different molecules in the cluster. There are no cut-off parameters which depend on real-space distances, and thus PolBE provides smooth potentials as shown in Fig. 2.2. PolBE is also trivially size-consistent as it recovers the non-interacting limit at large distances. In the interacting regime, PolBE binding energies are in good agreement with the SCF reference binding energies. PolBE also successfully recovers the repulsive part of the PES, while the EX0 model of EMFT cannot as shown in Fig. S1.

This numerical instability does not appear in PolBE despite using the inexpensive EX0 model for exact exchange as the ALMO constraint prevents any charge transfer between fragments. However, it should be emphasized that the lack of charge transfer between fragments simply yields inadequate binding energies (as shown in the SCFMI curve in Fig. 2.2(b)). Therefore, allowing for pairwise charge transfer as is done in PolBE is very effective in recovering the missing charge transfer interaction energy. Non-additive dispersion interactions are also not captured by PolBE. These higher-order interactions are typically very small. In the case of water trimer at equilibrium geometry, higher order dispersion interaction would contribute an additional 0.06 kJ/mol to the PolBE binding energy.

Table 2.1: Binding energy errors for the cage and cyclic conformers. Relative binding energy of the cage conformer with respect to the cyclic conformer as predicted by vacuum MBE truncated at 2-body (MBE-2), vacuum MBE truncated at 3-body (MBE-3), and PolBE truncated at 2-body (PolBE-2). The SCF binding energies for the cage and cyclic conformers are −153.7 and −160.7 kJ/mol. All values are in kJ/mol.

	E_{bind} error		
	cage	cyclic	E_{bind}(cage) − E_{bind}(cyclic)
MBE-2	28.4	45.4	−9.9
MBE-3	1.1	5.9	2.3
PolBE-2	−7.2	−4.5	4.3



(a) (b)

Figure 2.3: (a) Cage conformer of water pentamer (b) Cyclic conformer of water pentamer.

Water pentamer conformers

When a many-body expansion is truncated at the 2-body level,higher order binding cooperativity is not captured. The change in the binding affinity of molecule upon being bound to another water molecule is termed cooperativity. The phenomenon of cooperativity is found ubiquitously in chemical systems including water, [182] methanol,[183] and hydrogen fluoride. [184] The importance of understanding cooperativity cannot be overemphasized as it is crucial for understanding the bulk phase properties of these molecules. For instance, it enhances the strength of hydrogen bonding in water at a molecular level leading to anomalous adsorption properties in bulk water. [185] Cooperativity or non-pairwise additivity can be quantified as the amount of interaction energy that is not captured by the many-body expansion truncated at the 2-body level. In this work, we show how the higher order effect of cooperativity may be captured at a 2-body level using the PolBE.

There has been considerable interest in predicting the relative energies of various molecular clusters, both qualitatively and quantitatively. [186–188] More often than not, hydrogen bonded molecules in clusters interact in a non-pairwise fashion,[183, 184] making the total binding energy not a simple sum of pairwise interaction energies. It has been shown that homodromic hydrogen bonding ring-like networks are associated with the highest cooperativity. [182] This defining characteristic of these structures make them the most stable conformer in small water clusters, from trimer through pentamer, even though they contain only one hydrogen bond per water molecule while other conformers have more than one hydrogen bond per water molecule.

In this section, we investigate the case of two conformers of water pentamer: cyclic and cage using geometries taken from Ref. 189. The cyclic conformer, with the homodromic ring-like structure, exhibits more cooperativity in binding than the cage conformer. Vacuum MBE trucated at the 2-body level makes errors of 28.4 and 45.4 $kJmol^{-1}$ for the cage and cyclic conformer respectively showing that that the cyclic conformer has more cooperativity than the cage variant. Vacuum MBE not only predicts quantitatively wrong binding energies, but also incorrectly predicts that the cage conformer is the more stable than the cyclic one. Vacuum MBE needs the 3-body term in order to capture this many-body effect, as shown in Table. 2.1. A full DFT calculation (ωB97M-V/def2-TZVPPD) on these conformers shows that the cyclic conformer is more stable than the cage by 7.0 kJ/mol. PolBE which correctly embeds the 1-body and 2-body terms in an approximate representation of the environment makes much smaller binding energy errors of 7.2 and 4.5 kJ/mol. PolBE is able to correctly pre-polarize the monomers, thus decomposing the interactions in a pairwise additive fashion. This demonstrates the ability of our 0^{th} order wavefunction (SCFMI wavefunction) to precisely capture the mutual polarizations between the different water molecules which is clearly a many-body effect. PolBE also predicts the relative energetics correctly, predicting that the cyclic conformer is more stable with a relative energy error of 2.7 kJ/mol, smaller than chemical accuracy of 1 kcal/mol. In the context of molecular simulation, it is more important to get the relative binding energy ordering and separations correctly rather than absolute binding energies. While vacuum MBE needs to be carried out till the third order

to even get the relative ordering of conformers correct qualitatively, PolBE can achieve both the qualitative and quantitative relative energy ordering correct at the two-body level.

The water simulation community has recognized the importance of these higher-body effects and has tried to incorporate them either implicitly through inclusion of classical polarization [190, 191] or explicitly by including 3-body and other higher order terms. [192–196] This parametrization is further complicated by the fact that different phases of water (clusters, liquid water, and ice) show different magnitudes of non-additivities. [197, 198] PolBE does not contain any parametrization, and can approximate the 'high level' density functional with reasonable accuracy. PolBE also suggests an alternative way to account for these many-body effects, through the use of an inexpensive, albeit crude, zeroth order wavefunction.

CO_2 clusters

There has been widespread interest in studying CO_2 molecular clusters because of their presence in earth's atmosphere and that of other planets. [199] There is a significant interest in the green chemistry community as supercritical CO_2 can be used as a benign solvent on an industrial scale. [200] Theoretical studies on CO_2 clusters have served as a bridge to understand its condensed phase properties.

CO_2 are classic examples of weakly-bound clusters. A study on the intermolecular potential of CO_2 has shown that dispersion is one of the leading terms governing the interaction. [201] We performed an ALMO-based energy decomposition analysis [126, 127, 202] of the binding energy of a single CO_2 molecule completely surrounded by other CO_2 molecules for all the clusters considered in this study. This decomposition, as summarized in Fig. 2.4(a), shows that polarization, which is the energy associated with the relaxation of molecular orbitals on each fragment in response to molecular orbitals on other fragments, contributes 38% of the interaction energy with charge transfer causing an additional stabilization of 27%. Nevertheless, as shown in Fig. 2.4(a), dispersion is the main mechanism which binds these CO_2 clusters, and the ability to describe it accurately is critical to understanding the molecular processes governing the chemistry of carbon dioxide. In fact, the semi-local functional PBE used for representing the environment, which does not contain any dispersion corrections, predicts that the CO_2 clusters are unbound. The VV10 non-local correlation functional is a sophisticated method to accurately capture the effect of dispersion. The high level functional used in this study, ωB97M-V, contains the VV10 non-local correlation functional and correctly predicts the interaction energy of CO_2 trimer to be -14.98 kJ/mol (CCSD(T)[203] interaction energy is -15.96 kJ/mol). This shows that the correct treatment of dispersion is very much essential in predicting binding energies of such systems accurately.

We have investigated the ability of PolBE to predict interaction binding energies for a series of CO_2 clusters with geometries taken from Ref. 204. In Ref. 204, the geometries for the CO_2 clusters were optimized at M06/aug-cc-pVDZ. The binding energy per monomer increases with system size from -5.1 kJ/mol for a trimer to -14.1 kJ/mol for the 16-mer. The binding energy per monomer shows signs of saturation starting at the 13-mer, as it

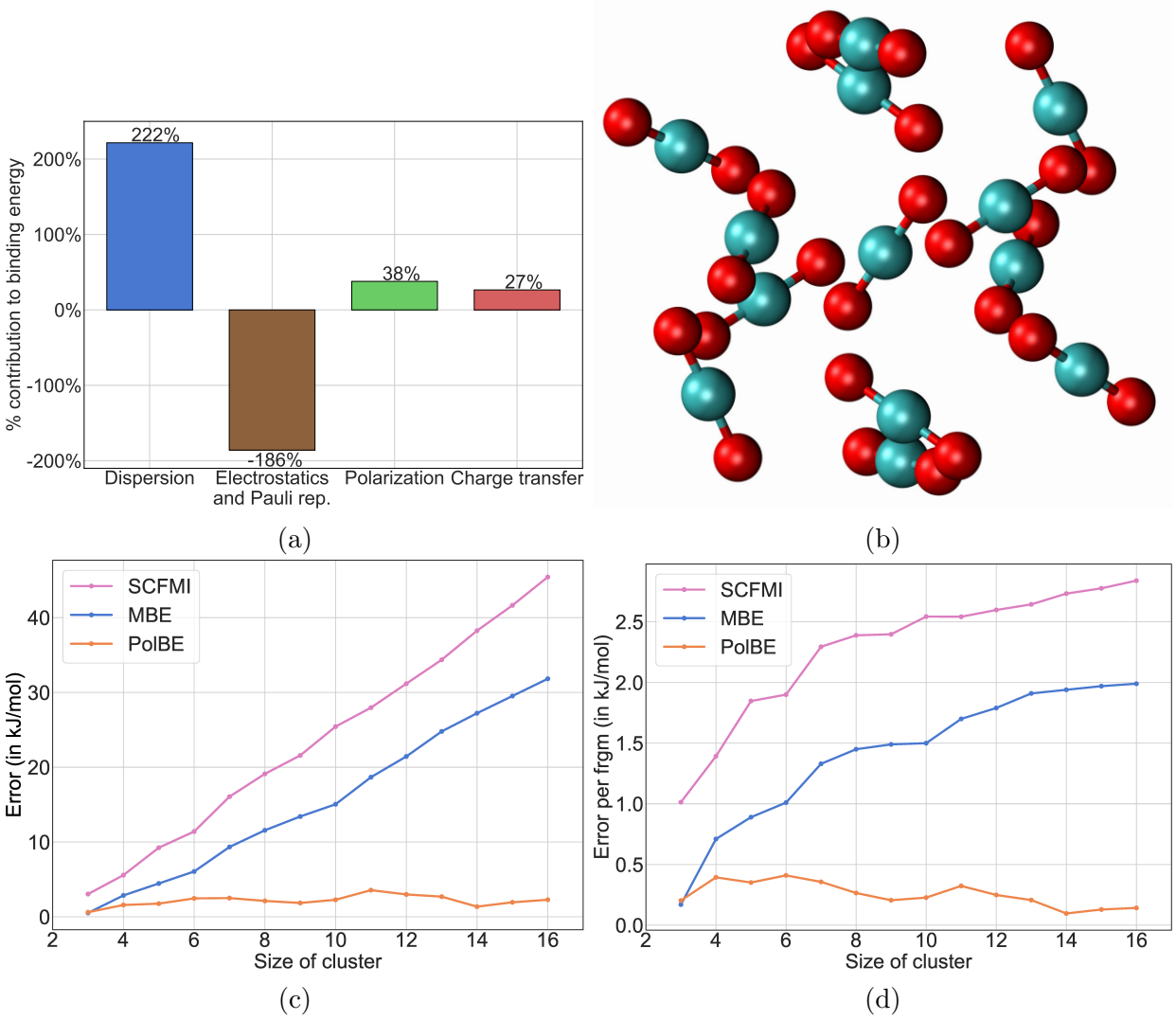


Figure 2.4: (a) Energy decomposition analysis for all CO_2 clusters considered in this study. Dispersion is the interaction between induced dipoles of different fragments. Pauli repulsion is the steric energetic penalty that occurs when two non-bonded atoms are forced to occupy the same space. Electrostatics is the energy associated with the coulombic interactions between fragments. Polarization is the energy associated with relaxation of molecular orbitals. Charge transfer is the energy associated with the flow of electrons between the fragments. (b) A CO_2 cluster containing 13 molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of CO_2 cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of CO_2 cluster size. For comparison, the reference binding energy for 16-mer is 225.3 kJ/mol.

only increases by 0.5 kJ/mol from 13-mer to 16-mer. The saturation of binding energy per monomer at such a small size of cluster is a consequence of two effects: 1. the loose packing of CO_2 molecules in the cluster and 2. Dispersion interactions which dominate the binding energy of CO_2 are short-range effects. The errors in the predicted binding energies for SCFMI (XC1/BS1), vacuum MBE and PolBE have been plotted in Figure 2.4(c) as a function of cluster size. The binding energy error predicted by SCFMI and vacuum MBE truncated at the 2-body level increases continuously with system size. At all fragment sizes, SCFMI, vacuum MBE, and PolBE underestimate the binding energy but the magnitude of underestimation varies significantly. The binding energy error prediction of the SCFMI wavefunction is the largest, followed by that of vacuum MBE and then PolBE. For the 16-mer, the largest CO_2 cluster considered in this study, the vacuum MBE error is 31.83 kJ/mol while the PolBE error is only 2.28 kJ/mol, smaller by more than an order of magnitude.

The error per monomer saturates for all the three methods as shown in Figure 2.4(d). For the largest clusters investigated in this study, the binding energy error per monomer saturates at around 0.15 kJ/mol for PolBE whereas it saturates at 2 kJ/mol for vacuum MBE and 2.8 kJ/mol for SCFMI. This saturation of error per monomer is very encouraging and its magnitude is expected to be transferable to large condensed phase systems of CO_2.

Water clusters

Water clusters are of fundamental importance in biological systems and processes and in atmospheric chemistry.[205] Unlike bulk water, water clusters contain molecules at the surface with unsaturated hydrogen-bonds serving as models for liquid-vapor interfaces. Almost every possible experimental tool has been used to characterize the structure of water clusters in an attempt to understand its anomalous properties.[206] It is also possible to use highly accurate ab initio quantum chemical methods to study small water clusters, enabling direct comparison between experiment and theory.

Water molecules interact with each other through hydrogen bonding, participating both as donor and acceptor. Water molecules in both water clusters and bulk water interact with each other in a highly non-pairwise additive fashion. [182] Dispersion is an important mechanism contributing 35% to the total interaction energy of water clusters as shown in Fig. 2.5(a). The gas phase dipole moment of 1.855D of water is enhanced in the condensed phase by about 40%. [207–210] This enhancement in dipole moment coupled with the increasing polarizability of water clusters with size of the cluster [211] leads to stronger binding through magnified polarization contributing about 29% to the interaction energy. While different energy decomposition schemes differ on the actual magnitude of charge transfer in a hydrogen bond in water, all methods agree on that fact that charge transfer plays a significant role in hydrogen bonding placing its relative importance somewhere between 20% to 30%. [126, 212] Performing an ALMO-based energy decomposition of the interaction energy for each term of the many-body expansion has shown that polarization and charge transfer play an important role in the 2-body terms, while polarization dominates the 3-body

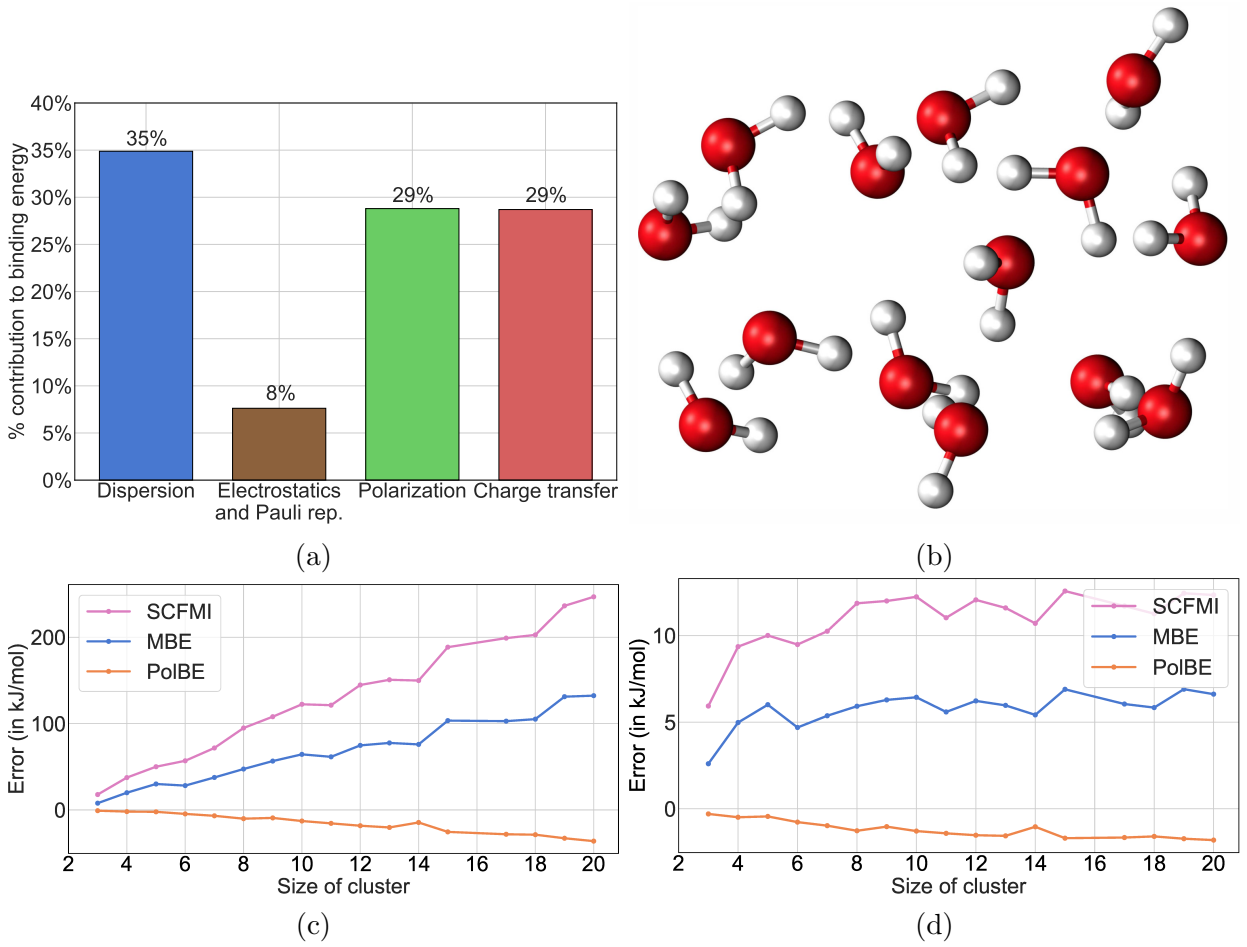


Figure 2.5: (a) Energy decomposition analysis for all water clusters considered in this study (b) A water cluster containing 14 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of water cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of water cluster size. For comparison, the total reference binding energy for the water 20-mer is 845.2 kJ/mol.

terms. [213] Charge transfer is mostly a pairwise additive effect and higher order terms are negligible.

In this section, we investigate the binding interactions of water clusters using geometries taken directly from a paper by Gadre and co-workers.[214] Binding in water clusters is much stronger than that in CO_2 clusters as can be seen from the larger interaction energies at every cluster size. The binding energy of the water trimer considered in this study has an interaction energy of 63.6 kJ/mol while the CO_2 trimer has a much smaller interaction energy of 15.4 kJ/mol. The presence of long-range interactions like dipole-dipole interactions slows down the saturation of binding energy per monomer as well. Even for the 20 water molecule cluster, the binding energy per water molecule is 42.3 kJ/mol and does not seem to be saturated. As a consequence, SCFMI, vacuum MBE, and PolBE errors per fragment also show slower signs of saturation. As in the case of CO_2 clusters, SCFMI and vacuum MBE systematically underbinds water clusters. The magnitude of the vacuum MBE errors, i.e. the errors arising from truncating the many-body expansion at the 2-body level, is an indicator of the importance of 3-body and higher body effects in the binding of water clusters. For a water 20-mer, vacuum MBE underestimates the binding energy by 132 kJ/mol which is about 16 % of the total interaction energy. PolBE errors are much smaller, increasing from less than 1 kJ/mol for water trimer to -36.2 kJ/mol for the water 20-mer. It is interesting to note that PolBE consistently overbinds water clusters while it underbinds CO_2 clusters.

The error per fragment of vacuum MBE increases, though not monotonically, from the trimer value of 2.6 kJ/mol to 6.6 kJ/mol for the 20-mer. SCFMI errors per fragment are much larger, consistently upwards of 11 kJ/mol for large clusters. PolBE makes much smaller errors per fragment starting at 0.3 kJ/mol for trimer and increasing to 1.8 kJ/mol for the 20-mer. The underlying many-body effect in water clusters, the polarization term, is treated in a complete many-body sense in PolBE. Furthermore, charge transfer is predominantly pairwise additive in nature and is treated upto 2-body in PolBE. The success of PolBE can be attributed to treating these underlying physical interactions accurately and consistently. Both vacuum MBE and PolBE error curves show kinks as a function of cluster size (see Fig. 2.5(d)) due to the lack of a systematic structure in water clusters with increasing size of the cluster. Certain geometries of water clusters are less compact than others, leading to fewer hydrogen bonds and consequently a smaller interaction energy in comparison to its neighboring clusters. The kinks closely follow the discontinuities in the binding energy per fragment plot as a function of water cluster size. Unlike the case of CO_2 where both the binding energy per fragment and the errors per fragment are saturated at the largest cluster size considered, these quantities continue to increase even at the 20-mer for water. Long-range electrostatics are, in part, responsible for this behavior and pose a considerable challenge to simulation of large systems of water. In order to investigate the saturation of SCF reference binding energy per fragment and PolBE binding energy error per fragment, we have investigated much larger water clusters $n = 32$ and $n = 64$. The SCF binding energy for $n = 32$ and $n = 64$ is -26.36 and -30.15 kJ/mol, and are not saturated. The PolBE binding energy error per fragment is -3.4 and -3.9 kJ/mol for $n = 32$ and $n = 64$ respectively. We see that the the PolBE binding energy error per fragment is saturating with

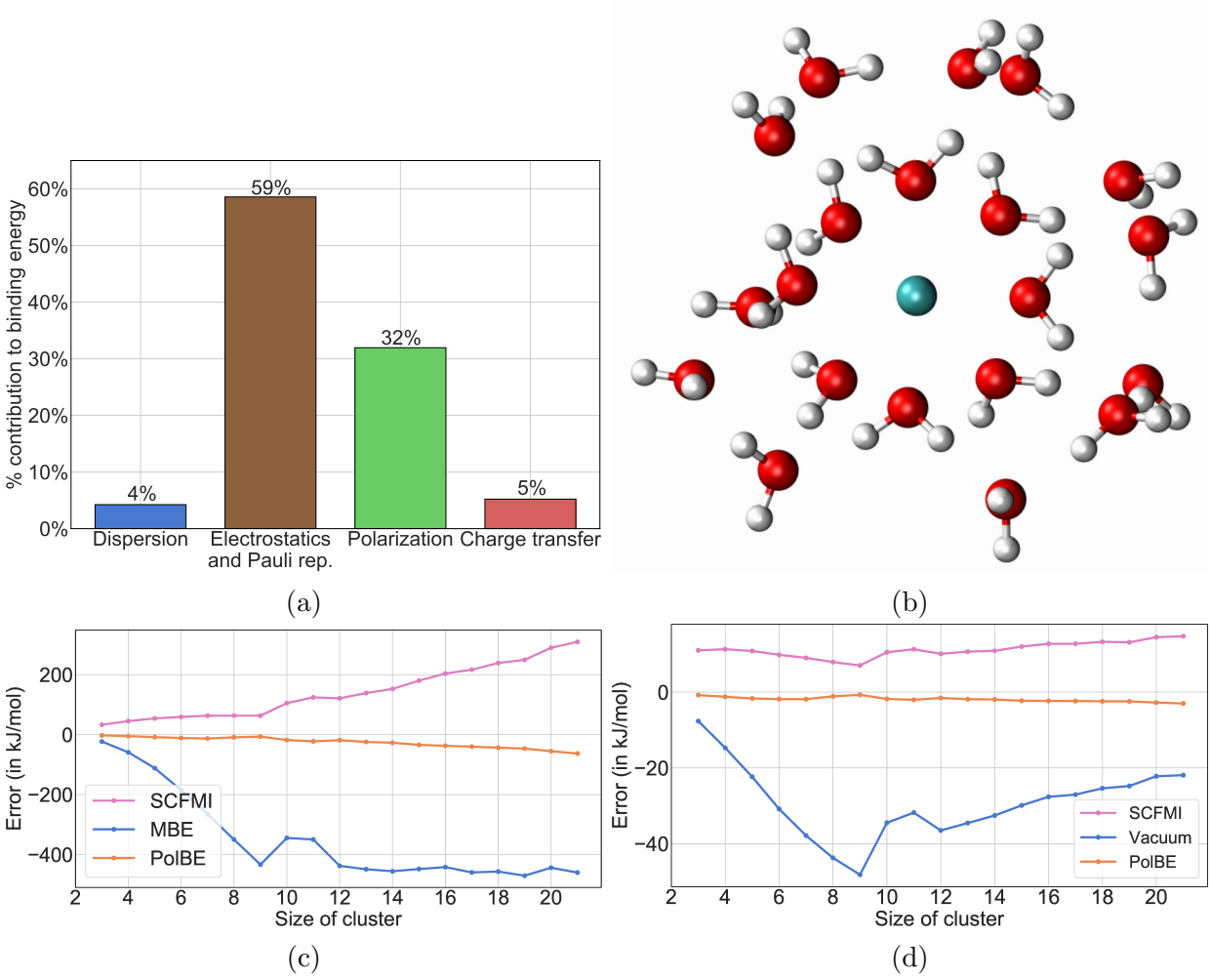


Figure 2.6: (a) Energy decomposition analysis for hydrated Ca^{2+} water clusters (b) Hydrated Ca^{2+} containing 20 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of cluster size. For comparison, the binding energy of $Ca^{2+}(H_2O)_{20}$ is 1935.6 kJ/mol.

cluster size, and hence showing that PolBE is size-extensive.

Hydrated ion clusters

Aqueous ions play a crucial role in many biological processes, and help sustain life on earth. While studying bulk aqueous ions has its own set of challenges, studying ion(s) in water clusters serve as a surrogate to understanding their properties. The small size of these clusters also enable the usage of accurate and expensive ab initio computational tools and direct comparison with experiment. These clusters also serve as models for studying hydration of

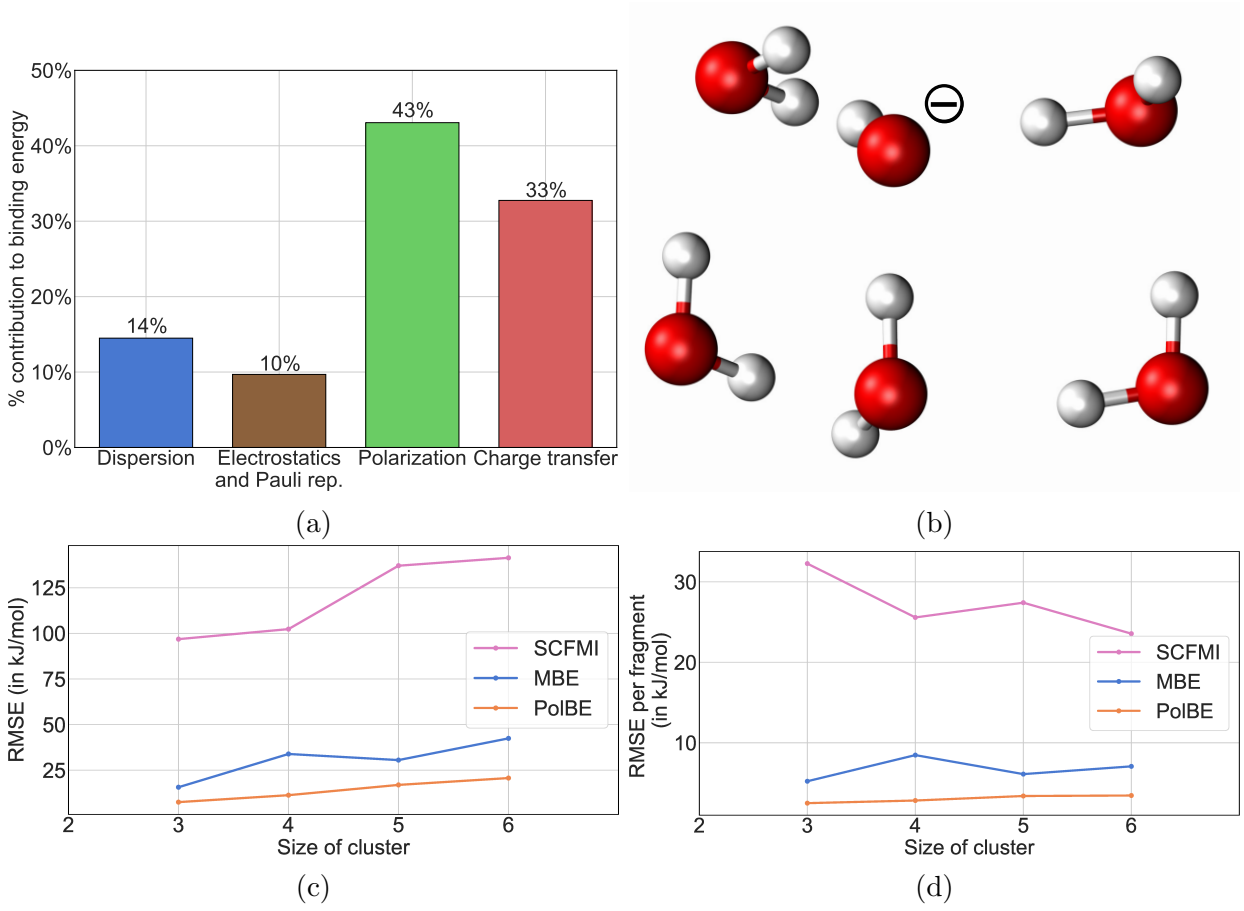


Figure 2.7: (a) Energy decomposition analysis for hydrated OH^- water clusters (b) Hydrated OH^- containing 5 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of cluster size. As multiple conformers of each cluster size were investigated, the error presented here are the root mean square errors of all conformers of a given size. For comparison, the reference RMS binding energy for $OH^-(H_2O)_5$ is 435.6 kJ/mol.

ions and their coordination shell structure. In this section, we present the performance of PolBE and vacuum MBE for cationic water clusters ($Ca^{2+}(H_2O)_n$, $n = 2 - 20$) and anionic water clusters ($OH^-(H_2O)_n$, $n = 2 - 6$). $Ca^{2+}(H_2O)_n$ cluster geometries were taken from Ref. 215 and $OH^-(H_2O)_n$ geometries are taken from Ref. 216.

The interaction of ions with water is very strong as can be seen from the large binding energies of the hydrated ion clusters. The binding energy of $Ca^{2+}(H_2O)_{20}$ per fragment is about 92 kJ/mol and that of $OH^-(H_2O)_5$ is 72 kJ/mol. Decomposition the interaction energy between the Ca^{2+} ion and its surrounding water molecules shows that the interaction is mainly dominated by electrostatics and polarization. The bare ion carries a large charge density giving rise to large frozen and polarization terms. At a frozen density level, this monopole interacts with the surrounding water dipoles which makes permanent electrostatics strong. These effects are also more long-ranged than in pure water due to the slower decay of monopole-dipole interaction. The length scale of these effects makes this class of systems poorly suited for fragmentation methods, and hence an extreme challenge for PolBE. The large charge density also creates a strong field polarizing the surrounding water molecules very significantly, thus explaining the large polarization term.

In the case of hydrated Ca^{2+} ions, the SCFMI wavefunction underestimates the interaction energy as it is unable to capture all the effects of charge transfer. The magnitude of underestimation increases from 33 kJ/mol for $Ca^{2+}(H_2O)_2$ to 310 kJ/mol for $Ca^{2+}(H_2O)_{20}$. On the other hand, vacuum MBE error increases from -22.9 kJ/mol to -460.0 kJ/mol for $Ca^{2+}(H_2O)_{20}$. For the $Ca^{2+}(H_2O)_{20}$ cluster, this error is about 23% of the total interaction energy. Vacuum MBE truncated at the two-body level misses certain important higher order effects. Vacuum MBE also overestimates the amount of charge transfer at the 2-body level when the ion comprises one of the two fragments. Overestimating the interaction energies at a 2-body level causes net overbinding, while neglecting higher order terms in the MBE causes underestimation of interaction energies. These two effects compete with each other to cause a temporary saturation in the total vacuum MBE error at around 450 kJ/mol in clusters of sizes 14 to 21. The saturation of total error with increasing cluster size causes the error per fragment to decrease after the first coordination shell of Ca^{2+} has been saturated. PolBE performs significantly better by making much smaller total errors, starting from 2.3 kJ/mol for $Ca^{2+}(H_2O)_2$ and increasing steadily to 63.0 kJ/mol for $Ca^{2+}(H_2O)_{20}$. PolBE error, even in these strongly interacting systems, ranges from 0.5 % to around 3 % of the total interaction energies, roughly an order of magnitude smaller than vacuum MBE errors. The PolBE error per fragment is only 3.0 kJ/mol for $Ca^{2+}(H_2O)_{20}$ compared to 1.8 kJ/mol for the water 20-mer.

The first important difference between hydrated Ca^{2+} clusters and hydrated OH^- clusters is the charge density. The single negative charge of the OH^- anion has a much smaller charge density causing smaller total interaction energies. Comparing the clusters of the same size, hydrated hydroxide ion interaction energies are roughly half of the hydrated calcium ion interaction energies due to the reduced strength of the monopole. The second difference is the contribution of charge transfer component to the interaction energies. In hydrated calcium clusters, charge transfer contributes only 5% to the total interaction energy, while in

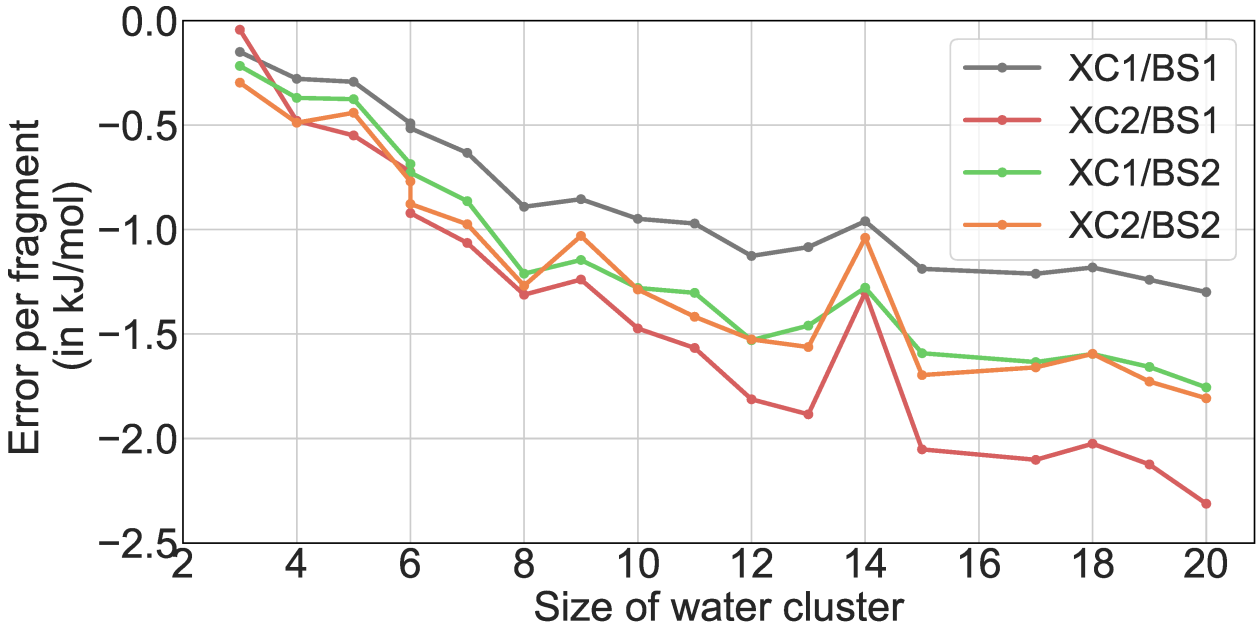


Figure 2.8: PolBE error per monomer for water clusters as function of size of the cluster with different representations of the environment. In this figure, XC1 is ωB97M-V, XC2 is PBE, BS1 is def2-TZVPPD, and BS2 is def2-SV(P). The system is always treated at XC1/BS1.

hydrated hydroxide anion clusters it contributes 33%. This is expected because Ca^{2+} is not a good Lewis acid, but OH^- is a strong Lewis base and donates electrons to water readily.

As the contribution of charge transfer to interaction energy is significant, the SCFMI wavefunction severely underestimates the interaction energy. Both vacuum MBE and PolBE overestimate the binding energies for hydrated hydroxide anions, but the errors are shown as positive values as error metric used is root mean square error. Vacuum MBE binding energy error increases from 15.7 kJ/mol to 42.4 kJ/mol for $OH^-(H_2O)_5$. PolBE errors are only about half the magnitude, overestimating the binding energy of $OH^-(H_2O)_5$ by 20.7 kJ/mol. Similar trends are seen in the binding energy errors per fragment.

Error analysis of different environment models

The PolBE method uses EMFT to build the Fock matrix and solves the SCFMI equations to obtain ALMOs. EMFT relies on the fact that an approximate description of the environment (B) is sufficient to accurately capture the essential chemistry occurring in the system (A). PolBE relies on this assumption, providing an approximate embedding density for computing 1-body and 2-body terms accurately. In this section, we discuss the validity of this assumption. As a mean-field theory, PolBE attempts to represent the environment approximately by using a semi-local functional (XC2) in a much smaller basis set (BS2). The PBE functional, while much cheaper to compute than ωB97M-V (XC1), is also grossly

inadequate for computing the binding energies of CO_2, as it does not contain any dispersion and predicts that the CO_2 clusters are unbound. The def2-SV(P) basis set used for the environment, denoted as BS2, is much smaller than the def2-TZVPPD used for the system. A single water molecule has 74 basis functions in BS1 and has only 18 basis functions in BS2. This makes the number of contributions of an environment water molecule to the Fock matrix smaller by a factor of 4 in comparison to the number for a system water molecule.

In this section, we analyze the sources of error coming from the functional and basis set approximations used to represent the environment. In Fig. 2.8, different approximations of the environments are compared in order to isolate the source of error that comes from each of them. The best representation of the environment uses XC1/BS1 for both the system and the embedding environment. This treatment of the environment density on the same footing as the system does not use EMFT to build the Fock matrix, and is equivalent to SCFMI for each of the 1-body and 2-body terms. Therefore, it is the baseline error to compare against (shown as the gray curve in Fig. 2.8). When the basis set of the environment is changed to BS2 and the functional is fixed at XC1, an additional error of 0.5 kJ/mol per monomer is added at the water 20-mer. Similarly, changing the environment functional to XC2 while keeping the basis set fixed at BS1 causes the largest error, adding an error of 1 kJ/mol per water molecule to the baseline error. The method proposed in this paper, which represents the environment using XC2 in a small basis set BS2, adds saturating error of only 0.5 kJ/mol to the baseline error. This shows that an approximate representation of the environment is sufficient to accurately and efficiently capture the interaction at a 1-body and 2-body level, decomposing the binding energy in an almost pairwise additive fashion. A systematic analysis of error for different representations of the environment (different local/semi-local functionals and basis sets) is beyond the scope of this paper.

Comparison with existing eMBE methods

Developing different eMBEs has been an active research topic of many research groups for the last couple of decades. These embedded MBE methods differ only in their description of the environment, ranging from using fixed atom-centered point charges, [154] using variationally-optimized atom-centered point charges, [100] to using actual monomer densities.[166]

Of these methods, the Electrostatically Embedded Many-Body Expansion (EE-MBE) [154] developed by Truhlar and co-workers, is the possibly the simplest and most widely used. In this method, each k-body term is computed in the presence of atom-centered point charges for the rest of system. While an approximate representation of the environment can be sufficient to attain useful accuracy, EE-MBE oversimplifies it, which causes important interactions like exchange repulsion, coulombic interaction of electron clouds, and polarization to be omitted, consequently leading to poorer performance. EE-MBE is also not parameter-free as the binding energies depend on the charges as shown by Herbert and co-workers previously. [156] Including higher-order terms in EE-MBE can lead to significant precision problems which would require tightening of SCF convergence and integral screening thresholds making them more expensive. [217] Fig. 2.9 shows the errors of EE-MBE performed

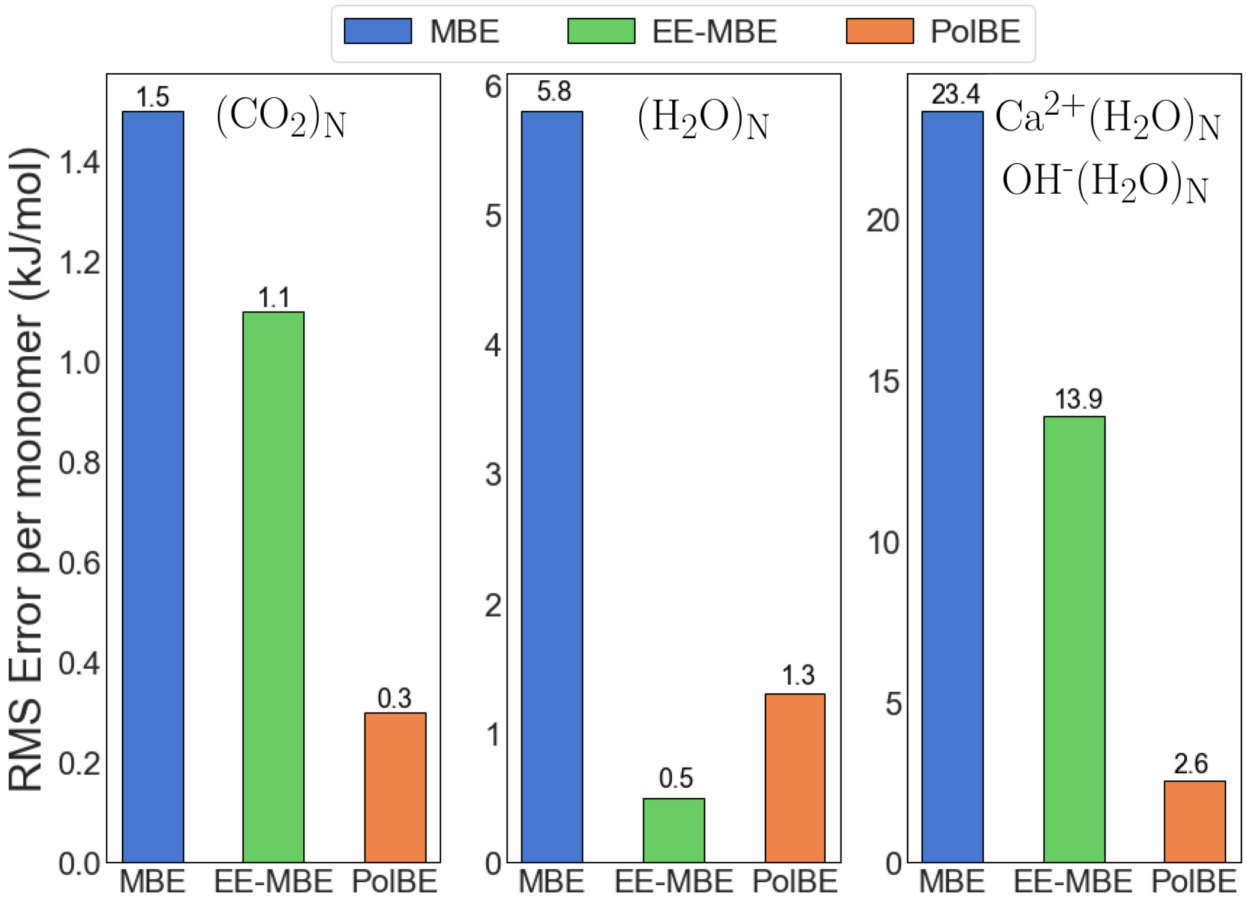


Figure 2.9: Root Mean Square Error per fragment for all carbon dioxide, water, and hydrated ion clusters. The errors for vacuum MBE, EE-MBE, and PolBE are shown. For $(CO_2)_N$, N ranges from 3 to 16. For $(H_2O)_N$, N ranges from 3 to 20. For $Ca^{2+}(H_2O)_N$, N ranges from 2 to 20 and for $OH^-(H_2O)_N$, N ranges from 2 to 5.

Table 2.2: A table showing the computational cost scaling with system size for construction of Fock matrix and diagonalization steps in SCF and PolBE

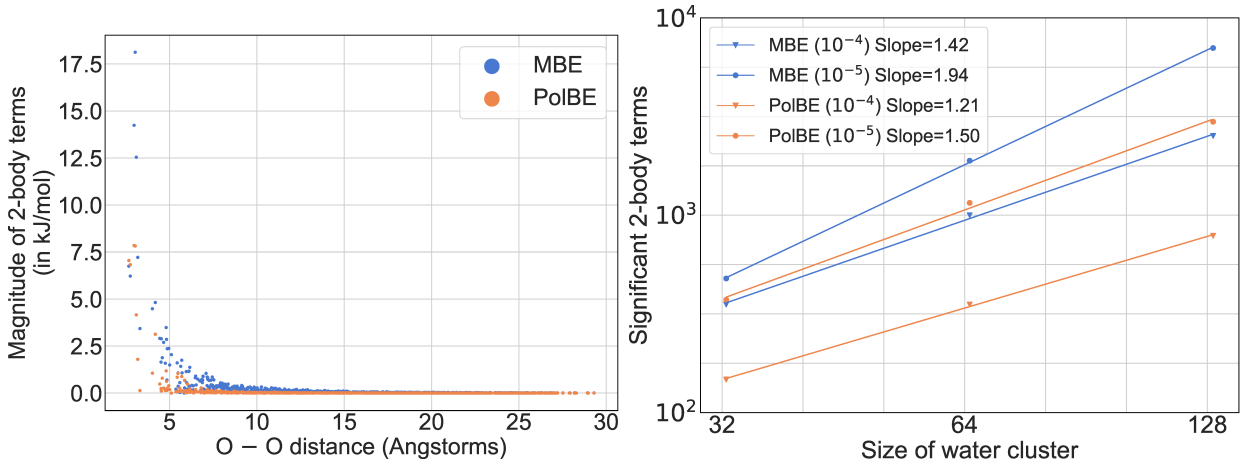
	Fock build		Diagonalization	
	SCF	PolBE	SCF	PolBE
2-body screening	$\mathcal{O}(F^2)$	$\mathcal{O}(F^2)+\mathcal{O}(F^3)+\mathcal{O}(F^3)$	$\mathcal{O}(F^3)$	$\mathcal{O}(F)+\mathcal{O}(F)+\mathcal{O}(F)$
No 2-body screening	$\mathcal{O}(F^2)$	$\mathcal{O}(F^2)+\mathcal{O}(F^3)+\mathcal{O}(F^4)$	$\mathcal{O}(F^3)$	$\mathcal{O}(F)+\mathcal{O}(F)+\mathcal{O}(F^2)$

using CHELPG charges, which are atom-centered charges fitted to reproduce the molecular electrostatic potential at a number of points around the molecule.[218] The CHELPG charges were determined for a water monomer at its equilibrium geometry, and the same set of charges were used for representing all environment water molecules. In comparison to EE-MBE, PolBE errors are much smaller in the case of CO_2 clusters and hydrated ion clusters. For all the conformers of hydrated ion clusters studied in this work, the RMS error per fragment of PolBE is only 2.6 kJ/mol whereas that of EE-MBE is 13.9 kJ/mol as seen in Fig. 2.9. For the water clusters investigated, EE-MBE errors are smaller than PolBE errors for the water clusters investigated. However, the performance of EE-MBE is heavily dependent on the embedding charges used as shown in Table S8. For instance, EE-MBE performed using Mulliken charges increases the error per fragment for water molecules to 2.8 kJ/mol. The errors of EE-MBE also depend on the functional and basis set used to run these calculations. The performance of PolBE is robust across the range of systems tested in this work and is also parameter-free. PolBE consistently matches the SCF binding energies computed using XC1/BS1 with modest errors irrespective of the choice of XC1 and BS1. Different choices for representation of the environment, viz. different XC2 and BS2, does not affect the predicted binding energies too significantly.

Computational timing analysis

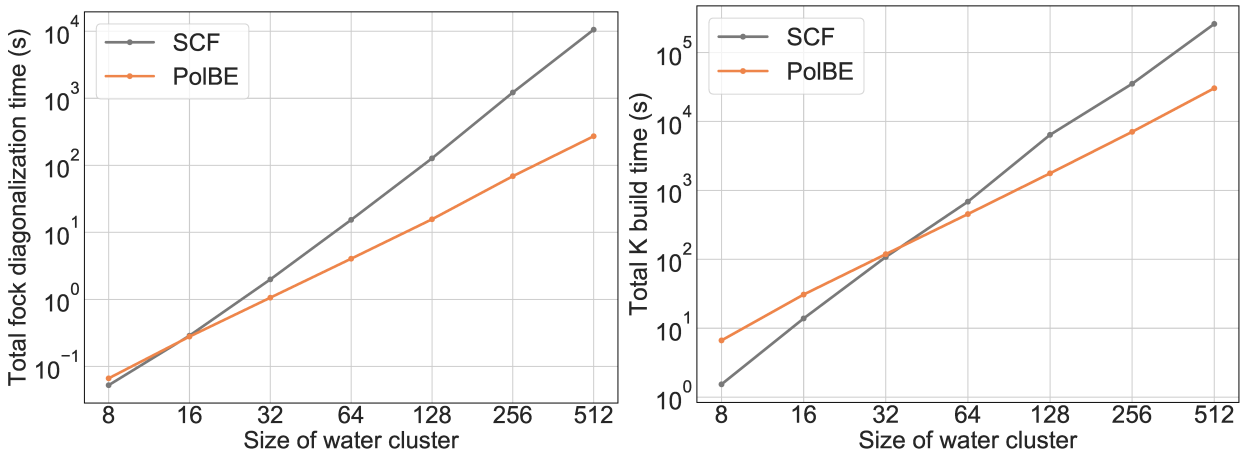
The computational expense in the SCF and PolBE procedures can be classified into two categories: diagonalization and Fock build. In this section, we discuss the computational cost of SCF and PolBE in terms of these two categories. All timings reported in this section were run on a 2.3 GHz 16-core AMD Opteron™ processor. Timings reported are the wall clock timings for one iteration of SCF and the wall clock total time for one iteration for all terms in PolBE. B3LYP[26, 219] and PBE[24] were used for XC1 and XC2, and def2-SVPD [173] and STO-3G[220] were used for BS1 and BS2 in PolBE. For the SCF calculation, the reported timings are for running B3LYP/def2-SVPD. Different density functional and smaller basis sets were used for this part of the timings analysis in order to access larger system sizes. Spherical water clusters used in this study were obtained from Ref. 221.

The number of terms at each order increases combinatorially with increasing order of the MBE. While the number of terms in theory increases combinatorially in PolBE as well, not all of them are numerically significant. This effect can be attributed to embedding, which



(a) (b)

Figure 2.10: (a) A plot of the magnitude of 2-body terms in MBE and PolBE as a function of the distance between the fragments in a water cluster containing 512 monomers (b) A log-log plot of the number of significant 2-body terms for MBE and PolBE as a function of size of the water cluster. The number of significant terms are indicated determined as the number of terms exceeding the threshold as indicated in parenthesis



(a) (b)

Figure 2.11: (a) Wall clock time taken for Fock matrix diagonalization(s) in SCF and PolBE as a function of cluster size (b)Wall clock time taken for construction of exact exchange matrix(ces) for SCF and PolBE as a function of cluster size. F is the number of monomers in the system which indicates system size. Timings reported are the wall clock timings for one iteration of SCF procedure and the wall clock total time for one iteration for all terms in PolBE.

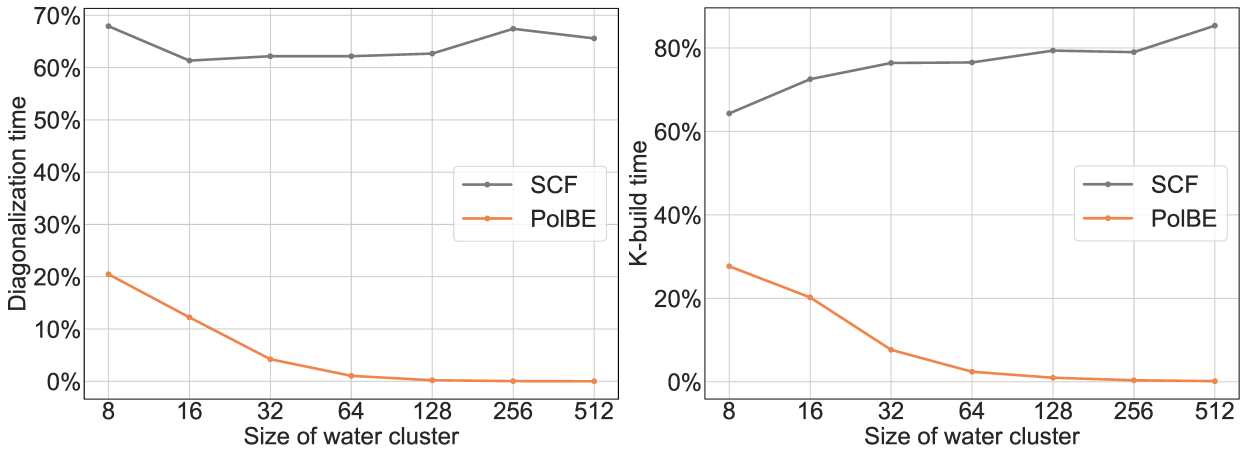


Figure 2.12: (a) Wall clock time taken for diagonalization of Fock matrix(ces) as a percentage of total time taken for all linear algebra manipulations for SCF and PolBE as a function of cluster size (b) Wall clock time taken for construction of the exact exchange bit as a percentage of total time taken for construction of the Fock matrix(ces) for SCF and PolBE as a function of cluster size. Timings reported are the wall clock timings for one iteration of SCF procedure and the wall clock total time for one iteration for all terms in PolBE.

incorporates physical effects like electrostatics and many-body polarization in the 0-body term as captured at XC2/BS2. These effects are further corrected using a the high-level functional XC1 in a large basis set BS1 at a 1-body level. After incorporating 1-body effects, the only physical interactions that are not yet captured are the corrections to the many-body polarization at XC1/BS1 and the effect of charge transfer. The 2-body terms capture only the effect of 2-body charge transfer and 2-body correction to the polarization, both of which decay rapidly with increasing distance between the pairs. Fig. 2.10(a) shows rapid decrease in the magnitude of 2-body terms with distance between the pair. It also shows that the magnitude of 2-body terms in PolBE are much smaller and decay much faster than 2-body vacuum MBE terms. This results in the linear increase in the number of 2-body terms in contrast to the quadratic number of all 2-body terms. Employing a conservative distance-based cutoff will limit the number of 2-body computations to scale linearly with system size. Considering cluster sizes of 32, 64, and 128 water molecules with a threshold of 10^{-5} Hartrees, Fig. 2.10(b) shows that the number of significant 2-body terms scales sub-quadratically with system size. Using looser thresholds and/or larger system sizes will show a scaling that is much closer to unity.

Consider a system with F fragments. Let N_1 and N_2 represent the total number of basis functions in BS1 and BS2 respectively. Let n_1 and n_2 represent the numbers of basis functions per monomer in BS1 and BS2 respectively. The cost of Fock matrix construction for SCF scales as $\mathcal{O}(N_1^2)$ leading to$\mathcal{O}(F^2)$ scaling with system size.[222] PolBE involves

construction of three types of Fock matrices: (a) One Fock matrix at XC2/BS2 for V_0 (b) A linear number of Fock matrices for the V_1 term (c) A linearor quadratic number of Fock matrices for the V_2 term depending on whether screening is employed or not employed. For each 1-body term, the size of the XC2 Fock matrix is roughly N_2. V_1 Fock matrices would also involve the computation of two additional Fock contributions from XC2 and XC1, but their size is limited to n_1 and does not scale with system size. The total cost of computation of the V_1 terms scales as $F \times \mathcal{O}(N_2^2)$. Each Fock matrix for the V_2 term also exhibits similar computational costs. Summing up, the cost of constructing all the Fock matrices scales as $\mathcal{O}(F^2) + \mathcal{O}(F^3) + \mathcal{O}(F^3)$ after employing 2-body screening.

The density update step, which involves diagonalization of the effective hamiltonian matrix, scales as $O(M^3)$, where M is the size of the system. For large system sizes, diagonalization of the Fock matrix becomes the most expensive step computationally. Diagonalization accounts for the majority of the computational cost of all linear algebra manipulations, constituting more than 60% of the cost across all fragment sizes as shown in Fig. 2.12(a). PolBE, which is based on the SCFMI wavefunction, removes this diagonalization bottleneck for large systems as shown in Fig. 2.12(a). As the cost of diagonalization scales as the third power of the dimension of the matrix, density update step in SCF scales as $\mathcal{O}(N_1^3)$. In PolBE at the V_0 level, the total cost of diagonalization scales as $F \times \mathcal{O}(n_2^2)$. As n_2 is constant, the cost of diagonalization for V_0 term scales as $\mathcal{O}(F)$. For each term in V_1 and V_2, only the MOs of the monomer and dimers are updated while the MOs of the environment are fixed. This requires diagonalization of projected Fock matrices with dimension n_1 and $2n_1$ for the V_1 and V_2 terms respectively. In conclusion, the diagonalization step scales as $\mathcal{O}(F) + \mathcal{O}(F) + \mathcal{O}(F)$ with system size. The total cost of diagonalization in SCF quickly exceeds that of PolBE for cluster sizes of 16 water molecules (624 basis functions) as shown in Fig. 2.11.

Fock matrix construction, comprisescoulomb, XC and exact exchange contributions, and scales as the 2^{nd} power of its size. Exact exchange dominates the computational cost as shown in Fig. 2.12(b). Hybrid density functionals, known to alleviate the self-interaction energy problem, provide a more reliable and accurate prediction of system properties, including non-covalent interaction energies. [18] However, addition of exact exchange relative to pure density functionals makes them more expensive to compute. [70, 71, 223] Different methods, ranging from screening based on locality to density-fitting have been proposed for reducing the computational cost of forming the Fock matrix.[223] However, computation of exact exchange has remained the bottleneck step. As shown in Fig. 2.12(b), construction of the exact exchange part of the Fock matrix takes up upwards of 60% of the total Fock construction time. PolBE, which embeds hybrid DFT in semi-local DFT, requires construction of only the AA block of the exact exchange matrix which is much smaller than the entire Fock matrix. The size of the AA block is only n_1 for each 1-body term and $2n_1$ for each 2-body term. The total SCF K-build time exceed the total K-build time for all the ${}^{F}C_1 + {}^{F}C_2$ terms in PolBE at 32 water molecules (1248 basis functions) as shown in Fig. 2.11(b). PolBE thus accelerates the computational bottleneck step in both linear algebra manipulations and Fock construction. In terms of the total computational cost, PolBE costs more than a full SCF calculation owing to the large number of Fock constructions at a 2-body level. However, all

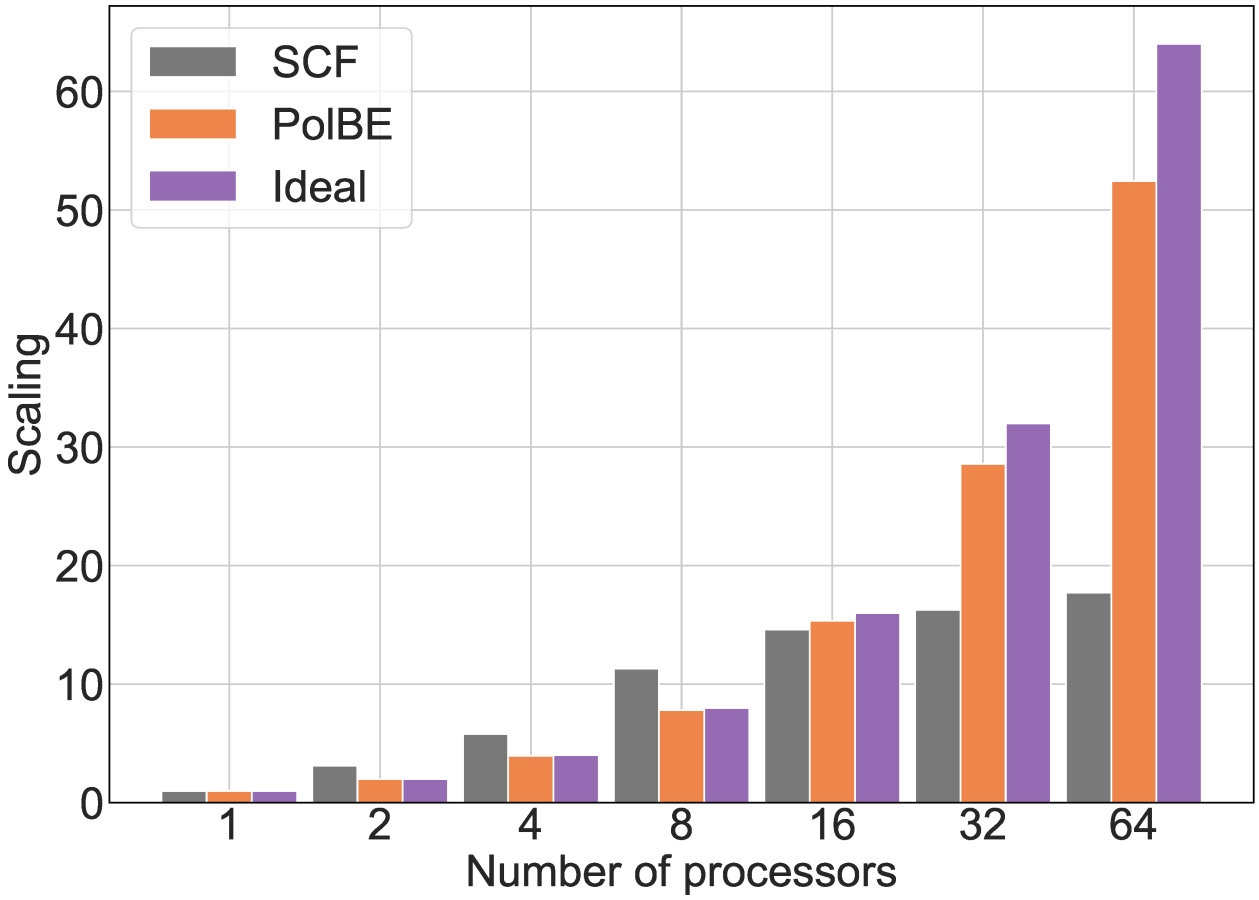


Figure 2.13: Scaling of SCF and PolBE with increasing number of processors for water 20-mer. Computations were performed using ωB97M-V/def2-TZVPPD in PBE/def2-SV(P) for PolBE and ωB97M-V/def2-TZVPPD for SCF.

of each of these computations are trivially parallel as described below.

A practical challenge in most DFT software packages is to be able to take full advantage of the multiple cores available in modern day high performance computing architectures. [224] PolBE provides a very simple way to take advantage of this. Once the embedding density has been formed, PolBE involves computation of ${}^{F}C_1 + {}^{F}C_2$ terms, all of them completely independent of each other. This property of PolBE, inherited from the many-body expansion, makes it trivial to distribute these jobs among any number of given processors making PolBE "embarrassingly parallel." As shown in Fig. 2.13, PolBE for water clusters exhibits nearly ideal parallel scaling. In principle, one can achieve even better scaling with PolBE by conceiving smarter master-slave process which are aware of the approximate time it would take to run each of the jobs *a priori*.

2.4 Conclusions and Outlook

In this work, we investigated the importance of accurately representing the environment in order accelerate the convergence of the many-body expansion. In particular, we found that capturing the effects of the environment using EMFT while partitioning the fragments using SCFMI accurately captures the many-body cooperativity effects commonly found in condensed phase systems. We have shown that the PolBE method proposed in this paper reproduces binding energies of a wide range of molecular clusters of various sizes and consisting of different interaction mechanisms. PolBE consistently performs better than other popular embedded many-body expansion methods EE-MBE. The computational cost of PolBE and its ability to remove the computational bottleneck steps was analyzed and demonstrated.

Comparison of the performance of PolBE to other embedded many-body expansion methods[157] would be of significant interest. The implementation of PolBE reported here can potentially be improved using various strategies. Larger systems can be made computationally more tractable by using a real-space cutoff for computing two-body terms. These cut-offs must depend on the dominant long-range interactions present in the system and the accuracy desired. Such cut-offs have not been investigated in this paper and are an interesting topic for future study. As PolBE is an MBE based method, its accuracy can be systematically improved by including higher order terms in the expansion. While the number of higher order terms increases with each order, some screening of important terms, either based on real-space distances or on the magnitude of lower terms, can be employed. The description of the environment can be made coarser to achieve additional speedups. PolBE, as suggested in this work, uses different basis sets and functionals for the system and environment. The extension of PolBE to use different DFT quadrature grids for the system and the environment is straightforward. Similar ideas can be extended to use different density fitting basis sets for the coulombic interaction term. The ALMO-based representation of the system used in PolBE yields fragment-specific occupied and virtual orbitals. These fragment-specific virtuals, which are both fragment-sparse and smaller in number in comparison to the virtual space of the whole system, provides a compact representation of the fragment virtual space which can be exploited in the extension of PolBE to double hybrids density functionals[34] and to correlated wavefunction methods like MP2, orbital-optimized MP2,[49–51, 177–179] and coupled cluster theories.[60] Extension of these ideas to correlated wavefunction methods is currently being investigated.

Chapter 3

A Non-Perturbative Pairwise-Additive Analysis of Charge Transfer Contributions to Interaction Energies

3.1 Introduction

Electronic structure calculations, such as modern density functional theory (DFT),[18] are capable of yielding accurate results for intermolecular interactions. However no physical or chemical insight beyond the calculated observable is obtained. Many different energy decomposition analyses (EDAs) [110, 115, 118, 123, 126, 225–227] have attempted to address this need. At their core, EDA attempts to partition computed interaction energies into physically and chemically motivated terms like electrostatics, dispersion, polarization, and charge transfer. EDA thereby aims to provide insights into the fundamental nature of molecular interactions which can, in principle, help guide the design of molecules with properties of interest. The development of force-fields based on these partitions of interaction energies is another utility of EDA schemes.[108, 109]

Interaction of molecules by electron delocalization or charge transfer (CT) is one of the most fundamental drivers of complex formation.[228] CT from a filled donor orbital of one molecule to an empty acceptor orbital of another lowers the total energy of the complex, and favors binding of the molecules. CT via forward and back donation in the Dewar-Chatt-Duncanson model[229, 230] plays a very important role in transition metal-ligand complex formation, and is instrumental in many transition metal catalytic transformations. Even hydrogen bonding, which is a much weaker interaction that is ubiquitous for (bio)molecules in aqueous solution, has a significant contribution from CT.[228, 231] Another interesting aspect of charge transfer stabilization is its diversity in the strength of interaction: it can vary in strength from a few kJ/mol in hydrogen-bonding complexes to hundreds of kJ/mol

in transition metal-ligand interactions.

The energy decomposition analysis scheme based on Absolutely Localized Molecular Orbitals (ALMOs) has been developed for mean-field theories (Hartree Fock and Density Functional Theory) [123, 126, 170, 232, 233] and Møller-Plesset perturbation theory.[171, 234, 235] The mean-field ALMO-EDA scheme partitions the interaction energy into geometric distortion, frozen, polarization, and CT components using variationally optimized intermediate wavefunctions as illustrated in Eq. (3.1):

$$\Delta E_{\mathrm{INT}} = \Delta E_{\mathrm{GD}} + \Delta E_{\mathrm{FRZ}} + \Delta E_{\mathrm{POL}} + \Delta E_{\mathrm{CT}} \tag{3.1}$$

The energy associated with distorting the relaxed isolated fragment geometries to their complex geometry is represented by the geometric distortion term (ΔE_{GD}). The frozen term (ΔE_{FRZ}) is the energy associated with bringing infinitely separated fragments to their complex geometry while retaining their infinitely-separated wavefunctions. This term includes permanent electrostatics, Pauli repulsion, and dispersion.[202] The polarization term (ΔE_{POL}) represents the energy lowering associated with relaxing the wavefunctions of each of the fragments in the presence of other fragments while not allowing CT between fragments.[126, 127] The energy lowering associated with CT between fragments is captured by ΔE_{CT}. CT can be further broken down into pairwise additive components for both energy lowering and its associated charge flow based on perturbation theory.[107] This decomposition scheme can also be used to extract chemically relevant Complementary Occupied-Virtual orbital Pairs (COVPs) which are the most important orbitals associated with CT. [107] However, this charge decomposition scheme breaks down in cases of strongly interacting species as it is based on perturbation theory.[107, 126] In this paper, we provide an alternative non-perturbative CT decomposition scheme which works robustly for all regimes of interaction energy, while nevertheless providing an effective pairwise description of CT.

There are many other proposed measures of CT, which we can only briefly summarize here. The charge decomposition analysis method (CDA) by Dapprich and Frenking defined forward and backward donation components in intermolecular complexes.[236] While this method is conceptually simple, some of the terms in CDA do not have a well-defined physical meaning.[107] Other EDA variants define charge transfer energy as the amount of energy associated with mixing occupied orbitals of the donor with virtual orbitals of the acceptor. Such ideas are used in the reduced variational space (RVS)[237] and constrained space orbital variation (CSOV)[238] schemes. However, these methods only account for CT in one direction at a time and cannot provide a pairwise additive decomposition of CT such that the contributions sum up to the total CT energy. Methods such as EDA-NOCV (natural orbital for chemical valence) [115, 239] and localized molecular orbital (LMO)-EDA [226] report only the sum of polarization and CT, which corresponds to the "orbital interaction" and "polarization" terms in these methods, respectively. The same choice is commonly followed in symmetry-adapted perturbation theory (SAPT) calculations of intermolecular interactions,[118, 119] where polarization and CT are combined in the "induction" term, although several approaches have been proposed to quantify CT in SAPT. [240, 241] The popular

Natural Bond Orbital (NBO) method can also be used to define charge decomposition between pairs of fragments using orthogonal molecular orbitals.[227, 242] However NBO and its associated Natural Energy Decomposition Analysis[116, 117] (NEDA) is known to greatly exaggerate the magnitude of CT energies.[212, 243] Constrained DFT (CDFT)[244–246] has also been used to estimate the magnitude of CT energy.[247, 248] As it constrains charge populations, rather than preventing charge-delocalization, CDFT was shown to substantially underestimate CT in most scenarios.[228]

In the ALMO-EDA, total CT is based on the energy difference between two variationally optimized wavefunctions: the fully relaxed final (FULL) wavefunction, and the constrained polarized (POL) wavefunction (which is evaluated by the self-consistent field for molecular interactions (SCF-MI) method[74, 75, 77, 170]):

$$\Delta E_{\mathrm{CT}} \equiv E(\Phi_{\mathrm{FULL}}) - E(\Phi_{\mathrm{POL}}) \tag{3.2}$$

This will be a faithful description of CT if Φ_{POL} describes a "CT-free" state of the complex: in other words it is constrained to prohibit CT whilst allowing on-fragment polarization in response to the rest of the complex. To achieve this target, the SCF-MI wavefunction optimizes a wavefunction in which the MO coefficient matrix is block-diagonal in the molecules making up the complex. Specifically, the ALMOs of each molecule are described by the mixing of its optimal occupied orbitals in isolation (the frozen occupied orbitals) with its own set of virtual orbitals. The simplest choice is to use *all* the virtual orbitals of the given isolated molecule for this purpose. However, this choice becomes ill-defined as the span of the AO basis of one molecule overlaps that of other molecules in the complex more and more. In the limit of a linearly dependent AO basis, the use of all fragment virtuals in a block-diagonal MO coefficient matrix can no longer guarantee a CT-free state. [127, 248, 249] To overcome this formal limitation (which can also be avoided in practice by using AO basis sets without excessively diffuse functions), one can instead use "fragment electric response functions" (FERFs) [127] as a limited set of virtuals for each molecule in the complex that describe exactly the response of the molecules to applied dipole (D) and quadrupole (Q) fields. The FERF-DQ model provides a well-defined complete basis set (CBS) limit for the resulting SCF-MI energy, and thus for ΔE_{CT}.

Since the CT energy is associated with orbital mixing between occupied levels on one fragment (molecule or ligand) of a complex and virtual orbitals of another, there are observable manifestations such as red-shifting of vibrational frequencies, as well as sometimes changes in structure. To quantify such effects, the adiabatic EDA[250] employs geometry optimization on each constrained surface to define a sequence of energy lowerings. The adiabatic EDA has proven useful to understand the role of charge transfer on observables[228, 250, 251]. It has also been extended to variationally quantify the role of forward and back donation individually on observable properties, via the recently introduced variational forward-backward (VFB) analysis of the relaxations due to CT.[252] This approach uses a generalized SCF-MI scheme that permits only uni-directional CT coupled to polarization in the entire system.

At a given geometry of a complex, an existing perturbative CT analysis (CTA) within the ALMO-EDA (reviewed in detail later) can decompose the CT energy into pairwise ad-

ditive components, plus a residual "higher-order" (HO) CT correction.[126]. This scheme is general enough to also decompose the charge flow, ΔQ_{CT} associated with CT into pairwise additive components.[107] The perturbative CTA relies on the HO correction being small, but unfortunately, it can vary substantially in both sign and magnitude based on the chemical system investigated (and even the choice of density functional). Furthermore, this perturbative decomposition can break down when CT becomes strong, as seen for example in the interaction between a transition metal and ligand in a transition metal complex. The recently proposed VFB approach [253] alleviates the issues associated with the perturbative CTA to some extent when it is employed to quantify forward and backward CT energies, but it does not incorporate the HO term fully. In addition, the current VFB formulation lacks a well-defined basis set limit.

To go beyond the perturbative CTA, in this work, we present an alternative which exactly (at least up to machine precision) decomposes the CT energy into pairwise additive terms irrespective of the strength of interaction. This scheme is applicable to a wide range of intermolecular interactions: from weakly bound hydrogen bonded complexes to strongly bound transition metal complexes. Improving upon its perturbative predecessor, the new CTA also provides a complete pair-wise additive decomposition of the quantity of charge transferred.

This paper is organized as follows. After a brief summary of the ALMO-based EDA and perturbative CTA, the non-perturbative CTA is introduced. We specify how the total CT can be exactly decomposed into pairwise additive interactions. Next, we illustrate the limitations of the perturbative CTA and demonstrate how the non-perturbative CTA overcomes these limitations. We show that the non-perturbative CTA has a well-defined basis set limit. Subsequently, we apply the new CTA to complexes ranging from hydrogen bonding in DNA base-pairs to strongly interacting transition metal hexacarbonyls. We illustrate how this new scheme debunks certain traditional ideas about borane adducts by revealing new mechanisms of interaction. Finally, we discuss the computational expense of the non-perturbative CTA relative to the perturbative CTA.

3.2 Theory

Brief summary of the ALMO-EDA

The ALMO-EDA scheme[126, 127] decomposes the interaction energy into chemically relevant quantities given in Eq. (3.1) by lifting imposed constraints one by one. The geometric distortion energy (ΔE_{GD}) associated with distorting relaxed (free) monomer geometries into their corresponding supersystem geometries is the difference between the SCF energy of each monomers, X, at the complex geometry ($\Phi_X^{\mathrm{complex}}$) and at their optimal free geometry (Φ_X^{free}):

$$\Delta E_{\mathrm{GD}} \equiv \sum_X \left(E_{\mathrm{SCF}}(\Phi_X^{\mathrm{complex}}) - E_{\mathrm{SCF}}(\Phi_X^{\mathrm{free}}) \right) \tag{3.3}$$

The frozen wavefunction of the complex, Φ_{FRZ}, is a Slater determinant composed of the occupied MOs from all isolated fragments (which are non-orthogonal). The corresponding energy, $E(\Phi_{\mathrm{FRZ}})$, accounts for Pauli repulsion by forming a valid density from the non-orthogonal frozen MOs. The frozen interaction (ΔE_{FRZ}) is defined as the difference between $E(\Phi_{\mathrm{FRZ}})$ and the isolated energies of all monomers at their complex geometry (Eq. (3.4)).

$$\Delta E_{\mathrm{FRZ}} \equiv E(\Phi_{\mathrm{FRZ}}) - \sum_X E_{\mathrm{SCF}}(\Phi_X^{\mathrm{complex}}) \tag{3.4}$$

Physically, ΔE_{FRZ} contains contributions from permanent electrostatic interactions, dispersion, and Pauli repulsions.[202] The polarized wavefunction (Φ_{POL}) is computed by relaxing the occupied MOs on each fragment in presence of the MOs of all other fragments while forbidding CT from one fragment to another by enforcing the ALMO constraint (that the MO coefficient matrix remains fragment block diagonal). This optimization of non-orthogonal orbitals with the ALMO constraint is achieved by SCF-MI,[74, 75, 77, 170] and the polarization energy is defined as the energy difference between the polarized and frozen wavefunctions.

$$\Delta E_{\mathrm{POL}} \equiv E(\Phi_{\mathrm{POL}}) - E(\Phi_{\mathrm{FRZ}}) \tag{3.5}$$

The use of FERF virtual orbitals to give the polarization term a well-defined basis set limit[127] has already been discussed above. So too has the definition of the energy lowering due to CT, which was shown in Eq. (3.2). By construction, we can see that the EDA terms sum up to the total interaction energy, defined in Eq. (3.1)

Perturbative Charge-Transfer Analysis

To set the stage for the new non-perturbative approach to pairwise decomposing the CT energy, we first discuss the existing perturbative CTA.[107] The usual second order perturbation correction to an energy may be written as $\Delta E^{(2)} = \mathrm{Tr}[\mathbf{F}^{(1)}\mathbf{X}^{(1)}]$, where $\mathbf{F}^{(1)}$ is the first order perturbed Hamiltonian and $\mathbf{X}^{(1)}$ is the first order perturbed wavefunction. The polarized wavefunction, Φ_{POL}, only zeros the mixing between occupied (O) and virtual (V) orbitals on fragments, but not between fragments. Therefore, given the polarized Fock matrix, $\mathbf{F}^{\mathrm{POL}} = \mathbf{F}(\mathbf{P}^{\mathrm{POL}})$, the perturbation is the residual occupied-virtual mixing, $\mathbf{F}_{OV}^{\mathrm{POL}}$, which is exclusively *between* fragments. This perturbation, or Roothaan Step (RS), in turn yields a perturbative mixing, $\mathbf{X}_{VO}^{\mathrm{RS}}$, of virtuals on a given fragment into occupieds of another.

The perturbative approximation to orbital mixing is obtained by a single diagonalization (Roothaan step) of the polarized Fock matrix, which is equivalent[254, 255] to solving the following quadratic equations for $\mathbf{X}_{VO}^{\mathrm{RS}}$:

$$\mathbf{F}_{\mathrm{VO}}^{\mathrm{POL}} + \mathbf{F}_{\mathrm{VV}}^{\mathrm{POL}}\mathbf{X}_{\mathrm{VO}}^{\mathrm{RS}} - \mathbf{X}_{\mathrm{VO}}^{\mathrm{RS}}\mathbf{F}_{\mathrm{OO}}^{\mathrm{POL}} - \mathbf{X}_{\mathrm{VO}}^{\mathrm{RS}}\mathbf{F}_{\mathrm{OV}}^{\mathrm{POL}}\mathbf{X}_{\mathrm{VO}}^{\mathrm{RS}} = \mathbf{0}_{\mathrm{VO}} \tag{3.6}$$

Equation (3.6) is written in the orthogonalized MO basis which is discussed in the Chapter B. The resulting $\mathbf{X}^{\mathrm{RS}}$ is still a perturbative solution, because it gives the energy lowering of a single diagonalization rather than iterating to self-consistency.

$$\Delta E_{\mathrm{CT}}^{\mathrm{RS}} = 2\,\mathrm{Tr}[\mathbf{F}_{\mathrm{VO}}\mathbf{X}_{\mathrm{OV}}^{\mathrm{RS}}] \tag{3.7}$$

Note that solving Eqs. (3.6) and (3.7) for $\Delta E_{\mathrm{CT}}^{\mathrm{RS}}$ is equivalent to infinite-order single excitation perturbation theory with a fixed Fock matrix,[254, 255] and thus is preferable to $\Delta E_{\mathrm{CT}}^{(2)}$. However the form still couples occupied MOs on one fragment with virtuals on another through $\mathbf{X}_{VO}^{\mathrm{RS}}$, and therefore $\Delta E_{\mathrm{CT}}^{\mathrm{RS}}$ is pairwise decomposable, just like $\Delta E_{\mathrm{CT}}^{(2)}$. The correction to $\Delta E_{\mathrm{CT}}^{\mathrm{RS}}$ for its lack of self-consistency is a non-pairwise decomposable, higher order (HO) term, $\Delta E_{\mathrm{CT}}^{\mathrm{HO}} = \Delta E_{\mathrm{CT}} - \Delta E_{\mathrm{CT}}^{\mathrm{RS}}$, which means the CT energy is represented as the sum of the RS contribution, and the residual HO term:

$$\Delta E_{\mathrm{CT}} = \Delta E_{\mathrm{CT}}^{\mathrm{RS}} + \Delta E_{\mathrm{CT}}^{\mathrm{HO}} \tag{3.8}$$

Equation (3.7) can be rewritten in terms of projectors onto the occupied space ($\mathbf{P}^{\mathrm{POL}}$) and virtual space ($\mathbf{Q}^{\mathrm{POL}}$) of the polarized wavefunction as follows.

$$\Delta E_{\mathrm{CT}}^{\mathrm{RS}} = \mathrm{Tr}\left[\mathbf{F}^{\mathrm{POL}}\mathbf{P}^{\mathrm{POL}}\mathbf{X}^{\mathrm{RS}}\mathbf{Q}^{\mathrm{POL}}\right] \tag{3.9}$$

$\mathbf{P}^{\mathrm{POL}}$ and $\mathbf{Q}^{\mathrm{POL}}$ are sums of projectors onto the polarized occupied orbitals and virtual orbitals on all fragments respectively.

$$\begin{aligned} \mathbf{P}^{\mathrm{POL}} &= \sum_{X,i} \mathbf{P}_{Xi}^{\mathrm{POL}} \\ \mathbf{Q}^{\mathrm{POL}} &= \sum_{X,a} \mathbf{Q}_{Xa}^{\mathrm{POL}} \end{aligned} \tag{3.10}$$

Eqs. (3.10) can be inserted into Eq. (3.9) to obtain the corresponding energy lowering due to transfer of charge from occupied orbital i on fragment X to virtual orbital a on fragment Y as follows.

$$\begin{aligned} \Delta E_{\mathrm{CT}}^{\mathrm{RS}} &= \sum_{X,Y} \sum_{i,a} \Delta E_{Xi\to Ya} \\ \Delta E_{Xi\to Ya} &= \mathrm{Tr}\left\{\mathbf{F}^{\mathrm{POL}}\mathbf{P}_{Xi}^{\mathrm{POL}}\mathbf{X}^{\mathrm{RS}}\mathbf{Q}_{Ya}^{\mathrm{POL}}\right\} \end{aligned} \tag{3.11}$$

Similarly, it has been shown that the total charge transferred can also be decomposed into pairwise additive components:[107]

$$\begin{aligned} \Delta Q_{\mathrm{CT}}^{\mathrm{RS}} &= \sum_{X,Y} \sum_{i,a} \Delta Q_{Xi\to Ya} \\ \Delta Q_{Xi\to Ya} &= \mathrm{Tr}\left\{\mathbf{P}^{\mathrm{RS}}\mathbf{P}_{Xi}^{\mathrm{POL}}\mathbf{X}^{\mathrm{RS}}\mathbf{Q}_{Ya}^{\mathrm{POL}}\right\} \end{aligned} \tag{3.12}$$

Here, $\mathbf{P}^{\mathrm{RS}}$ is the density matrix of the RS occupied orbitals. Using $\mathbf{P}^{\mathrm{RS}}$ to replace the Fock operator in Eq. (3.11) yields the charge transfer decomposition of Eq. (3.12).

While this approach is appealing and useful, its perturbative nature (i.e. its lack of self-consistency) is a disadvantage, because $\Delta E_{\mathrm{CT}}^{\mathrm{RS}}$ can never be exact. In particular, $\mathbf{X}^{\mathrm{RS}}$ generates the energy lowering, $\Delta E_{\mathrm{CT}}^{\mathrm{RS}}$, of a single diagonalization rather than the proper

self-consistent energy lowering, $\Delta E_{\rm CT}$, that results from lifting the SCF-MI constraint of no charge delocalization between fragments. In complexes containing large charge transfer, this can lead to the higher order term becoming large, not because CT cannot be partitioned in a pairwise additive fashion, but because the approximation of a single diagonalization becomes inadequate. We therefore take up the challenge of lifting this approximation.

Exactly Pairwise-Additive Charge-Transfer Analysis

In this section, we will show that it is possible to generate an exactly pairwise-additive charge transfer analysis (at least to as much precision as we wish). The working form of the result for the charge-transfer energy is:

$$\Delta E_{\rm CT} = E(\Phi_{\rm FULL}) - E(\Phi_{\rm POL}) = 2\,{\rm Tr}\left[\mathbf{F}_{VO}^{\rm CT}\mathbf{X}_{OV}^{\rm CT}\right] \tag{3.13}$$

This has the same form as the approximate expression, Eq. (3.7) discussed above, but with new definitions of the Fock matrix and the occupied-virtual mixings.

To begin, since we know the polarized state (i.e. $E(\Phi_{\rm POL})$) and the final, fully relaxed state (i.e. $E(\Phi_{\rm FULL})$), we also, at least implicitly, know the occupied-virtual mixings, $\mathbf{X}_{OV}^{\rm CT}$, necessary to connect them. We shall discuss how we explictly obtain them after establishing the exactly pairwise addition decomposition. Let us define $\lambda = 0$ as the state $\Phi_{\rm POL}$, and $\lambda = 1$ as the state $\Phi_{\rm FULL}$ which are connected along the straight-line path specified by $\lambda\mathbf{X}^{\rm CT}$. Applying the fundamental theorem of line integrals along this path, it is true by definition that:

$$\Delta E_{\rm CT} = E(1) - E(0) = \sum_{i,a}\int_0^1 \frac{\partial E(\lambda)}{\partial\lambda X_{ia}^{\rm CT}}\cdot X_{ia}^{\rm CT} d\lambda \tag{3.14}$$

Since:

$$F_{ai}(\mathbf{X}^{\rm CT}) \equiv \frac{1}{2}\frac{\partial E(\mathbf{X}^{\rm CT})}{\partial X_{ia}^{\rm CT}} \tag{3.15}$$

We are led directly to the desired result, Eq. (3.13), where evidently the appropriate Fock matrix arises from integrating along the line:

$$\mathbf{F}_{VO}^{\rm CT} = \int_0^1 \mathbf{F}_{VO}[\lambda\mathbf{X}^{\rm CT}]d\lambda \tag{3.16}$$

Let us turn next to the charge reorganization that occurs as the electron density is rearranged. To connect directly to the energy changes discussed above, we choose to write it in an isomorphic form (other more obvious possibilities[115] also exist, of course, and are independently useful). The charge that is rearranged, $\Delta Q_{\rm CT}$, as the system evolves from the POL state to the final FULL state from $\lambda = 0$ to $\lambda = 1$ along the straight-line path given by $\lambda\mathbf{X}$ can likewise be represented by a line integral:

$$\Delta Q_{\rm CT} = Q(1) - Q(0) = \int_{\lambda=0}^{\lambda=1}\sum_{i,a}\frac{\partial Q[\lambda\mathbf{X}^{\rm CT}]}{\partial\lambda X_{ia}^{\rm CT}} X_{ia}^{\rm CT} d\lambda \tag{3.17}$$

Here $Q[\lambda \mathbf{X}^{\mathrm{CT}}]$ is the charge that is promoted from occupied levels in the polarized state to orbitals that are virtual in the polarized state:

$$Q[\lambda \mathbf{X}^{\mathrm{CT}}] = \mathrm{Tr}\big\{\mathbf{Q}_0 \mathbf{P}[\lambda \mathbf{X}^{\mathrm{CT}}]\mathbf{Q}_0\big\} = \mathrm{Tr}\big\{\mathbf{P}[\lambda \mathbf{X}^{\mathrm{CT}}]\mathbf{Q}_0\big\} \tag{3.18}$$

The variation of $Q[\lambda \mathbf{X}^{\mathrm{CT}}]$ with respect to elements, X_{ia}^{CT} of X^{CT} is:

$$\begin{aligned}\frac{\partial \,\mathrm{Tr}\big\{\mathbf{P}[\lambda \mathbf{X}^{\mathrm{CT}}]\mathbf{Q}^{\mathrm{POL}}\big\}}{\partial \lambda X_{ia}^{\mathrm{CT}}} &= -\lambda(\mathbf{P}^{\mathrm{POL}}\mathbf{X}\mathbf{Q}^{\mathrm{POL}})_{ai} - \lambda(\mathbf{Q}^{\mathrm{POL}}\mathbf{X}\mathbf{P}^{\mathrm{POL}})_{ai} \\ &= -\lambda(\mathbf{X}_{VO}^{\mathrm{CT}})_{ai} \\ &= P_{ai}(\lambda \mathbf{X}^{\mathrm{CT}})\end{aligned} \tag{3.19}$$

Substituting Eq. (3.19) into Eq. (3.17), we get

$$\Delta Q_{\mathrm{CT}} = \int_{\lambda=0}^{\lambda=1} \sum_{i,a} P_{ia}(\lambda \mathbf{X}^{\mathrm{CT}}) X_{ai}^{\mathrm{CT}} d\lambda - \int_{\lambda=0}^{\lambda=1} \sum_{i,a} P_{ai}(\lambda \mathbf{X}^{\mathrm{CT}}) X_{ia}^{\mathrm{CT}} d\lambda \tag{3.20}$$

$$= 2\,\mathrm{Tr}\big\{\mathbf{P}_{VO}^{\mathrm{CT}}\mathbf{X}_{OV}^{\mathrm{CT}}\big\} \tag{3.21}$$

where

$$\mathbf{P}_{VO}^{\mathrm{CT}} = \int_{\lambda=0}^{\lambda=1} \mathbf{P}_{VO}[\lambda \mathbf{X}^{\mathrm{CT}}] d\lambda \tag{3.22}$$

Two issues must be addressed to employ this approach in practice: (i) we must find the orbital mixings associated with CT, $\mathbf{X}_{OV}^{\mathrm{CT}}$, and (ii) we must develop a suitable quadrature to efficiently and accurately numerically evaluate Eq. (3.16) and Eq. (3.22). We address these points in turn below.

Density matrices are independent of the redundant occupied-occupied and virtual-virtual mixings that are needed to fully specify molecular orbitals. Working in an orthonormal basis, we are seeking the unitary transformation, $\mathbf{U}^{\mathrm{CT}}$ connecting the polarized density matrix, $\mathbf{P}^{\mathrm{POL}}$ and the fully relaxed density matrix, $\mathbf{P}^{\mathrm{FULL}}$:

$$\mathbf{P}^{\mathrm{FULL}} = \left(\mathbf{U}^{\mathrm{CT}}\right)^{\dagger} \mathbf{P}^{\mathrm{POL}} \mathbf{U}^{\mathrm{CT}} \tag{3.23}$$

$\mathbf{U}^{\mathrm{CT}}$ can be written just in terms of occupied-virtual mixings, $\mathbf{X}_{OV}^{\mathrm{CT}}$, as the matrix exponential:

$$\mathbf{U}^{\mathrm{CT}} = \begin{bmatrix} \mathbf{U}_{OO}^{\mathrm{CT}} & \mathbf{U}_{OV}^{\mathrm{CT}} \\ \mathbf{U}_{VO}^{\mathrm{CT}} & \mathbf{U}_{VV}^{\mathrm{CT}} \end{bmatrix} = \exp\left\{ \begin{bmatrix} \mathbf{0} & \mathbf{X}_{OV}^{\mathrm{CT}} \\ -(\mathbf{X}_{OV}^{\mathrm{CT}})^{\dagger} & \mathbf{0} \end{bmatrix} \right\} \tag{3.24}$$

To solve for $\mathbf{X}_{OV}^{\mathrm{CT}}$ we define a cost function that vanishes when Eq. (3.23) is satisfied:

$$C = ||\mathbf{P}^{\mathrm{FULL}} - \mathbf{U}^{\mathrm{CT}}\mathbf{P}^{\mathrm{POL}}\left(\mathbf{U}^{\mathrm{CT}}\right)^{\dagger}||_F^2 \tag{3.25}$$

As shown in the Chapter B, the analytical gradient of the cost function, C, with respect to the ia^{th} element of $\mathbf{X}_{OV}^{\text{CT}}$ is given by:

$$\frac{\partial C}{\partial \left(\mathbf{X}_{OV}^{\text{CT}}\right)_{ia}}\Big|_{\mathbf{X}_{OV}^{\text{CT}}=\mathbf{0}} = 4\left[\left(\mathbf{U}_{\text{curr}}^{\text{CT}}\right)^{\dagger}\mathbf{P}^{\text{FULL}}\mathbf{U}_{\text{curr}}^{\text{CT}}\right]_{ia} \tag{3.26}$$

Here $\mathbf{U}_{\text{curr}}^{\text{CT}}$ is the current approximation to the unitary transformation. We choose a working orthonormal basis that symmetrically orthogonalizes the ALMO occupieds, and the ALMO virtuals (after projecting out their occupied components), and canonically orthogonalizes the projected virtual void orbitals (which were excluded from the evaluation of $\mathbf{P}^{\text{POL}}$). Minimizing C using the analytical gradient is performed via standard iterative techniques for unconstrained non-linear equations like quasi-Newton methods and DIIS in order to obtain $\mathbf{X}_{OV}^{\text{CT}}$. In principle, one can solve for $\mathbf{X}^{\text{CT}}$ analytically by taking the logarithm of $\mathbf{U}^{\text{CT}}$. However, we do not use this method as this leads to non-zero OO and VV blocks of $\mathbf{X}^{\text{CT}}$ and hence cannot be used for pairwise decomposition directly.

Now, we will address how to develop a suitable numerical quadrature scheme to evaluate $\mathbf{F}_{VO}^{\text{CT}}$ and $\mathbf{P}_{VO}^{\text{CT}}$. We can numerically evaluate the integral in Eq. (3.16) and Eq. (3.22) using Gauss quadrature rules in the interval $[0, \mathbf{X}^{\text{CT}}]$. Specifically, we use the 5-point Gauss-Lobatto quadrature rules,[256] as they include the end points of the interval which have already been evaluated for the purpose of EDA. The expressions for $\mathbf{F}_{VO}^{\text{CT}}$ and $\mathbf{P}_{VO}^{\text{CT}}$ are shown in Eq. (3.27) and Eq. (3.28).

$$\begin{aligned}\mathbf{F}_{VO}^{\text{CT}} = &\frac{1}{20}\mathbf{F}_{VO}[0] + \frac{49}{180}\mathbf{F}_{VO}\left[\frac{1}{2}\left(1-\sqrt{\frac{3}{7}}\right)\mathbf{X}^{\text{CT}}\right] \\ &+ \frac{16}{45}\mathbf{F}_{VO}\left[\frac{1}{2}\mathbf{X}^{\text{CT}}\right] + \frac{49}{180}\mathbf{F}_{VO}\left[\frac{1}{2}\left(1+\sqrt{\frac{3}{7}}\right)\mathbf{X}^{\text{CT}}\right] + \frac{1}{20}\mathbf{F}_{VO}[\mathbf{X}^{\text{CT}}]\end{aligned} \tag{3.27}$$

$$\begin{aligned}\mathbf{P}_{VO}^{\text{CT}} = &\frac{1}{20}\mathbf{P}_{VO}[0] + \frac{49}{180}\mathbf{P}_{VO}\left[\frac{1}{2}\left(1-\sqrt{\frac{3}{7}}\right)\mathbf{X}^{\text{CT}}\right] \\ &+ \frac{16}{45}\mathbf{P}_{VO}\left[\frac{1}{2}\mathbf{X}^{\text{CT}}\right] + \frac{49}{180}\mathbf{P}_{VO}\left[\frac{1}{2}\left(1+\sqrt{\frac{3}{7}}\right)\mathbf{X}^{\text{CT}}\right] + \frac{1}{20}\mathbf{P}_{VO}[\mathbf{X}^{\text{CT}}]\end{aligned} \tag{3.28}$$

As $\mathbf{F}_{VO}[0]$ and $\mathbf{P}_{VO}[0]$ are the Fock and density matrices of the polarized wavefunction (Φ^{POL}) and $\mathbf{F}_{VO}[\mathbf{X}^{\text{CT}}]$ and $\mathbf{P}_{VO}[\mathbf{X}^{\text{CT}}]$ are the Fock and density matrices of the fully relaxed wavefunction (Φ^{FULL}), we would have to evaluate only three new Fock and density matrices at points specified by Eqs. (3.27) and (3.28). Empirically, for the systems studied in this work, using the 5-point quadrature formula is sufficient to recover the variational charge transfer energy with sub-Joule per mole accuracy (See Table B.6).

To partition CT we need an appropriate set of projectors onto the occupied and unoccupied subspaces of each of the interacting fragments. In the second-generation ALMO-EDA,[202] the Hilbert space of each polarized fragment X is spanned by its occupied frozen

orbitals and the virtual Fragment Electric Response Functions (FERFs)[127] that are basis-independent and provide a well-defined basis set limit for ΔE_{POL}. After polarization, let us denote these spaces as P_X and V_X. The unoccupied AO space on each fragment not spanned by FERF virtuals (hereinafter virtuals) is termed the "void space" (denoted by R_X) and the entire Hilbert space on fragment X (denoted by H_X) can be written as shown in Eq. (3.29)

$$H_X = P_X \oplus V_X \oplus R_X \tag{3.29}$$
$$= P_X \oplus Q_X \tag{3.30}$$

where $Q_X = V_X \oplus R_X$ denotes the total unoccupied space formed by combining virtuals and voids.

Following Ref. 107, the projector onto the i^{th} occupied orbital on fragment X, denoted by $\hat{P}_{Xi}$, can be formed as shown below:

$$\hat{P}_{Xi} = |\phi_{Xi}\rangle \left\langle \phi^{Xi} \right| \tag{3.31}$$
$$= \sum_{Y,j} |\phi_{Xi}\rangle \, (\boldsymbol{\sigma}_{OO}^{-1})^{Xi,Yj} \, \langle \phi_{Yj}| \tag{3.32}$$
$$(\mathbf{P}_{Xi})^{\mu,\nu} = \sum_{Y,j} (\mathbf{C}_{\mathrm{POL}})_{\bullet Xi}^{X\mu\bullet} (\boldsymbol{\sigma}_{OO}^{-1})^{Xi,Yj} ((\mathbf{C}_{\mathrm{POL}})^T)_{Yj\bullet}^{\bullet Y\nu} \tag{3.33}$$

In the equations above, the biorthogonal covariant and contravariant notation was used for dealing with non-orthogonal molecular orbitals.[257] One further consideration regarding the virtual space projector is the treatment of near and exact linear dependence in large AO basis sets, which is part of ensuring a decomposition with well-defined basis set limits. Linear dependence between basis functions within the same fragment can be resolved by simply discarding one of the linearly dependent molecular orbital, for example by canonical orthogonalization. However, this is not possible with linear dependence between fragments without losing the association of the retained functions to fragments. Therefore inter-fragment linear dependence in the virtual space is treated by using the Moore-Penrose generalized inverse of the overlap matrix ($\boldsymbol{\sigma}_{VV}^{+}$) by discarding the null space.[257] For CT decomposition, we do not make any distinction between FERF virtuals and the void space.

$$\hat{Q}_{Xa} = |\phi_{Xa}\rangle \left\langle \phi^{Xa} \right| \tag{3.34}$$
$$= \sum_{X,a} |\phi_{Xa}\rangle \, (\boldsymbol{\sigma}_{VV}^{+})^{Xa,Yb} \, \langle \phi_{Yb}| \tag{3.35}$$
$$(\mathbf{Q}_{Xa})^{\mu,\nu} = \sum_{Y,b} (\mathbf{C}_{\mathrm{POL}})_{\bullet Xa}^{X\mu\bullet} (\boldsymbol{\sigma}_{VV}^{+})^{Xa,Yb} ((\mathbf{C}_{\mathrm{POL}})^T)_{Yb\bullet}^{\bullet Y\nu} \tag{3.36}$$

With the aid of these projectors, we can now exactly (up to arbitrary precision) decompose the CT energy, Eq. (3.13) into contributions from each pair of orbitals, on each pair of

interacting fragments in the complex:

$$\begin{aligned}\Delta E_{\mathrm{CT}} &= \sum_{X,Y}\sum_{i,a}\Delta E_{Xi\to Ya}\\ \Delta E_{Xi\to Ya} &= \mathrm{Tr}\left\{\mathbf{F}^{\mathrm{CT}}\mathbf{P}^{\mathrm{POL}}_{Xi}\mathbf{X}^{\mathrm{CT}}\mathbf{Q}^{\mathrm{POL}}_{Ya}\right\}\end{aligned} \tag{3.37}$$

Similarly, the total charge transferred can be exactly decomposed into pairwise additive orbital components:

$$\begin{aligned}\Delta Q_{\mathrm{CT}} &= \sum_{X,Y}\sum_{i,a}\Delta Q_{Xi\to Ya}\\ \Delta Q_{Xi\to Ya} &= \mathrm{Tr}\left\{\mathbf{P}^{\mathrm{CT}}\mathbf{P}^{\mathrm{POL}}_{Xi}\mathbf{X}^{\mathrm{CT}}\mathbf{Q}^{\mathrm{POL}}_{Ya}\right\}\end{aligned} \tag{3.38}$$

In practice, the results are compacted into the CT energy and flow between each pair of fragments. For the occupied orbitals on one fragment, X, coupling to virtual orbitals on another fragment, Y, it is convenient to singular value decompose their couplings, $\left(\mathbf{X}^{\mathrm{CT}}_{OV}\right)^{XY}$.

$$\left(\mathbf{X}^{\mathrm{CT}}_{OV}\right)^{XY} = \mathbf{L}^{X}_{O}\mathbf{s}\left(\mathbf{R}^{Y}_{V}\right)^{\dagger} \tag{3.39}$$

Often there is only 1 significant singular value, in which case there is a single complementary donor-acceptor orbital pair (COVP) controlling CT, where the donor orbital is represented by the first column of $\mathbf{L}^{X}_{O}$ and the acceptor is given by the first column of $\mathbf{R}^{Y}_{V}$.

Finally we note that $\mathbf{X}^{\mathrm{CT}}_{OV}$ describes CT couplings from all occupied orbitals to all unoccupied orbitals irrespective of the fragment that the virtuals belong to. This applies not just to the case where the fragments X and Y are different, but also to the case where they are the same. As a consequence of these non-zero couplings on the same fragment, we will have non-zero $\Delta E_{X\to X}$ terms, because the on-fragment blocks of $\mathbf{F}^{\mathrm{CT}}_{VO}$ are also non-zero. This is in contrast to the perturbative mixing, $\mathbf{X}^{\mathrm{RS}}_{OV}$, that has zero on-fragment blocks of $\mathbf{F}^{\mathrm{POL}}_{VO}$ (a result of the SCF-MI iterations) leading to zero $E_{X\to X}$ terms. The interpretation of the non-zero pairwise on-fragment terms for the exactly pairwise additive scheme is that this a re-polarization of the fragments in response to charge transfer.

3.3 Computational Details

The first consideration in decomposing charge transfer is computing the densities of the polarized ($\mathbf{P}^{\mathrm{POL}}$) and FULL ($\mathbf{P}^{\mathrm{FULL}}$) wavefunctions. After computing these quantities, we can solve for the non-perturbative mixing matrix ($\mathbf{X}^{\mathrm{CT}}$) by solving Eq. (3.23).

The perturbative and non-perturbative pairwise charge decomposition schemes were implemented in a developmental version of Q-Chem 5.0.[176] In particular, the new CTA method was implemented within the `libgscf` and `libloco` libraries which are, respectively, new SCF and SCFMI modules in Q-Chem.[51] ωB97X-D[258] with def2-TZVPD[173, 174] basis set was used for geometry optimization and energy decomposition analyses unless

stated otherwise. ωB97X-D is a range-separated hybrid density functional and has shown to give superior performance on a large number of systems for a wide range of chemical properties including non-covalent interaction energies.[18] All geometries were confirmed to be a minimum on the potential energy surface by confirming that the Hessian has no negative eigenvalues. All interaction energies computed adiabatically by including the relaxation energies of the individual fragments when they are infinitely separated from each other. All Complementary Occupied-Virtual Pairs (COVPs) are plotted with an isosurface value of ±0.07 au. All plots were created using `Matplotlib`[259] and molecule figures were generated using VMD.[260]

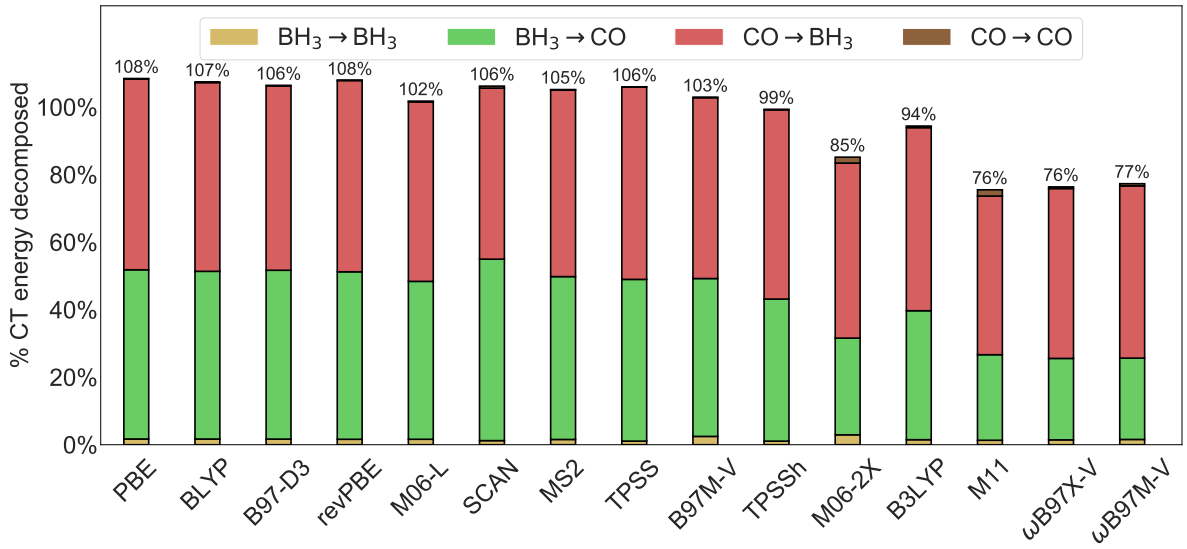
3.4 Results and Discussion

Perturbative CTA vs non-perturbative CTA

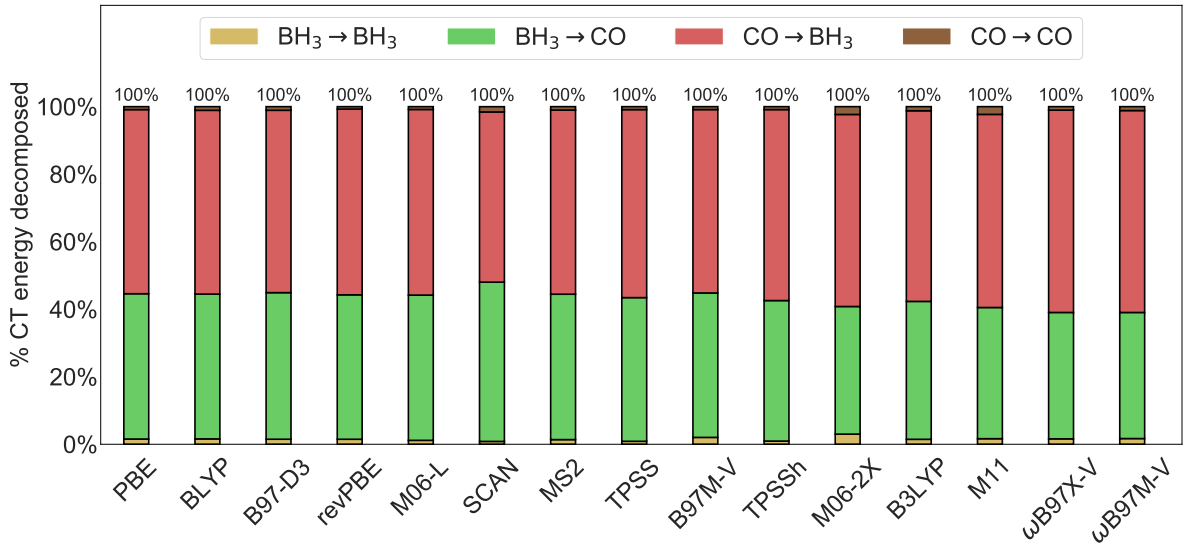
The perturbative CTA has some well-known limitations, especially its inability to pairwise decompose the entire CT energy. A manifestation of this problem is that the fraction of CT energy decomposed into pairwise additive terms depends on the density functional used, as illustrated in Fig. (3.1) for the case of the borane-carbonyl complex (BH_3−CO). BH_3−CO is bound by −123.6 kJ/mol at the ωB97X-D/def2-TZVPD level of theory as a result of strong CT character, with active forward (−171.0 kJ/mol) and backward donation (−113.8 kJ/mol). As shown in Fig. (3.1), the perturbative Roothaan step treatment overestimates CT for semi-local functionals by amounts ranging from 8% for PBE to about 2% for M06-L. Concurrently, the Roothaan step underestimates CT for hybrid and range-separated hybrid functionals by amounts ranging from 1% for TPSSh to rather severe values of 24% for M11 and ωB97X-V. Among the pairwise components, the perturbative BH_3→CO back-donation sees the most variation, fluctuating by more than 90 kJ/mol across the density functionals tested. This functional-dependent performance of the Roothaan step in the perturbative CTA is reminiscent of density functional minimal adaptive basis calculations based on the same perturbative corrections.[261]

On the other hand, the non-perturbative pairwise CTA consistently decomposes 100% of the charge transfer energy irrespective of the density functional employed. It is particularly encouraging to see from Fig. (3.1) that the functional-dependence of pairwise contributions in the non-perturbative approach is also significantly lower than their corresponding perturbative counterparts.

Consistent performance of EDA and CTA across the potential energy surface (PES) is necessary for training EDA-based force-fields which can benefit from these decompositions for parametrizing force-field terms. The limitation of perturbative CTA manifests in another form when considering different points on a PES. Consider a simple one-dimensional rigid PES formed by stretching the BH_3–CO complex along the B−C bond with the geometry of BH_3 and CO fixed to be the same as their equilibrium geometries. Fig. (3.2b) shows the total percentage of CT decomposed into pairwise additive terms for both perturbative and non-



(a)



(b)

Figure 3.1: Dependence of pairwise decomposition of charge transfer energy on different density functionals in the perturbative (upper panel) and non-perturbative (lower panel) treatment of charge transfer energy decomposition for the BH_3−CO complex in the def2-TZVPD basis set. The on-fragment CT terms (BH_3→BH_3 and CO→CO) are very small in most cases and cannot be seen in this figure for multiple density functionals. The different components of CT are shown as bars stacked on each other and appear in this order from bottom to top: BH_3→BH_3, BH_3→CO, CO→BH_3, and CO→CO.

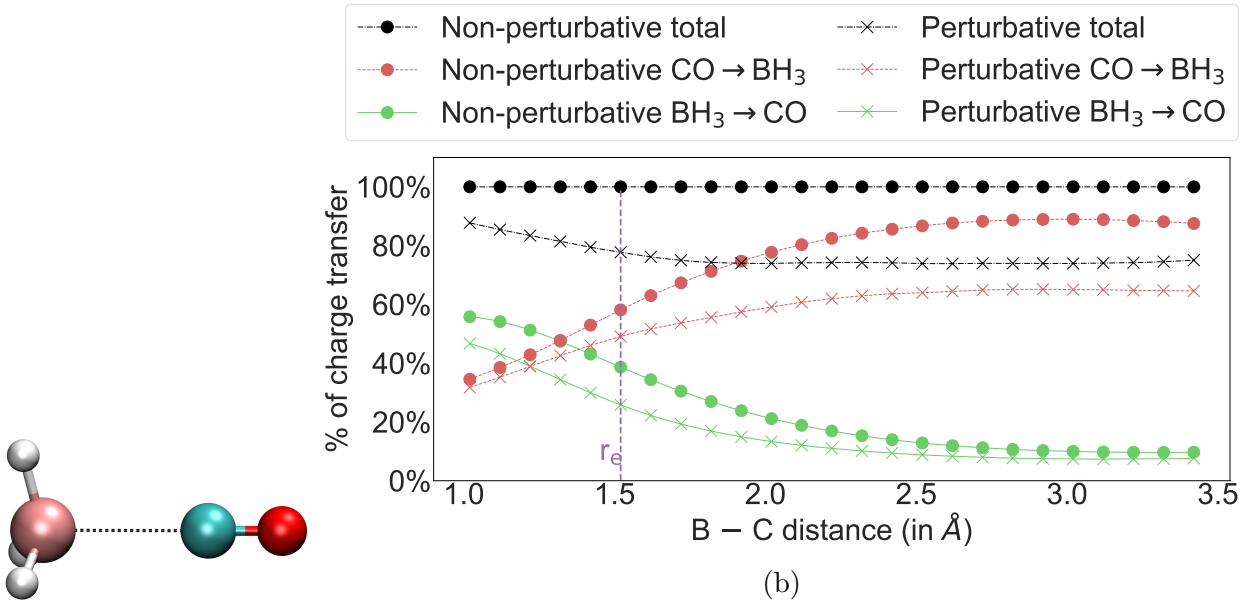


Figure 3.2: (a) BH_3–CO complex stretched along the dashed line (b) BH_3→CO, CO→BH_3, and total charge decomposed is shown as a percentage of the total charge transfer energy for the perturbative and non-perturbative CTAs for different B – C bond distances. The equilibrium bond distance (r_e) is shown in purple.

perturbative schemes. The total percentage of CT decomposed by the perturbative scheme varies along the potential energy surface. When the B−C bond is compressed to 1 Å, 88% of total CT is decomposed, while at equilibrium, this number falls to 77% and saturates at 74% as the bond is stretched. By contrast, and by virtue of its design, the non-perturbative CTA does not have this dependence at all. The charge transfer energy in Fig. (3.2b) varies over two orders of magnitude from −736.8 kJ/mol at 1 Å separation to −1.5 kJ/mol at 3.4 Å. From Fig. (3.2b), it is also interesting to note that the perturbative CTA yields components that deviate from the exact non-perturbative values by significantly different fractions across the PES: the perturbative CO→BH_3 component deviates significantly from the exact fraction at long distances, while it agrees much better at the shortest distances. On the other hand, the perturbative BH_3→CO fraction agrees well with the exact value at long-range, and deviates significantly at short-range. The advantage of the new non-perturbative CTA for pairwise decomposition is quite clear.

Basis set dependence

As modern density functionals show their best performance when approaching the basis set limit, it is advisable to run all calculations with as large a basis set as computationally

feasible [18]. Hence, one of the essential properties an energy decomposition scheme should have is a reasonably stable and physically meaningful basis set limit for all the contributing energy terms. The Fragment Electric Response Functions (FERFs),[127] which construct a polarization subspace for each fragment based on its response to electric fields, are used in the ALMO-EDA to give the polarization energy and CT energy well-defined basis set limits. In this section, we assess the basis set convergence properties of each of the pairwise charge and energy transfer components by again considering the example of BH_3–CO adduct at its equilibrium distance. Fig. (3.3) shows the partition of the CT energy into its 4 pairwise contributing terms as a function of increasing size of the AO basis. Similarly, Fig. (3.4) shows the corresponding partition of the pairwise decomposed charge flow terms. The size of the basis set can be increased in two complementary ways: (1) Increasing the highest angular momentum (cardinal number, X) of the one-particle basis set (denoted by the DZ, TZ, QZ, and 5Z sequence) (2) Increasing the level of augmentation of diffuse basis functions (denoted by the augmented (aug-)and doubly-augmented (d-aug-) prefixes). The latter effect is important for treating ions, excited states and strongly polarized systems with comparable accuracy to relatively non-polar ground states, but leads to near linear-dependence of the basis, which presents a challenge for methods such as CTA that use Hilbert space partitioning.

Figures (3.3) and (3.4) show that the CT partition into four components is well-behaved with respect to increasing X and augmentation. Considering the two major components of the CT energy, we can see that the $BH_3 \rightarrow CO$ component converges to 37% and the $CO \rightarrow BH_3$ converges to 56%. Similarly, the $BH_3 \rightarrow CO$ charge flow component converges to 67% and the $CO \rightarrow BH_3$ component converges to 35%. The double-ζ basis sets, contaminated with basis set superposition error (BSSE) as well as incompleteness error, are too small to reliably capture either the CT energy or charge flow. If the DZ basis sets are excluded, the two major components are almost converged at the triple-ζ basis set level, while the repolarization (on-fragment) components contribute little to CT regardless of basis set.

An interesting aspect of CT analysis for BH_3-CO is the fact that the pairwise energy and charge components are not correlated. The $CO \rightarrow BH_3$ energy component is larger than the $BH_3 \rightarrow CO$ CT energy component, while the opposite is true for the pairwise charge flow components. This result is consistent with perturbative analysis presented earlier.[107] The underlying reason has its origin in the different quantities being decomposed: the CT energy versus the charge flow associated with CT. Let us denote V_{DA} as the matrix element coupling donor (D) and acceptor (A) orbitals, with orbital energies ϵ_{D} and ϵ_{A}. From dimensionality (or perturbation theory) arguments, the CT energy is $\sim V_{\mathrm{DA}}^2/(\epsilon_{\mathrm{D}} - \epsilon_{\mathrm{A}})$. On the other hand, the charge flow behaves as $\sim V_{\mathrm{DA}}/(\epsilon_{\mathrm{D}} - \epsilon_{\mathrm{A}})$. Figures showing the magnitude of these pairwise decomposed components in kJ/mol and me^- are included in the Chapter B.

DNA base-pairs

Hydrogen bonding between DNA bases is one of the most important non-covalent interactions as it modulates a myriad of biological phenomena, such as the melting temperature of

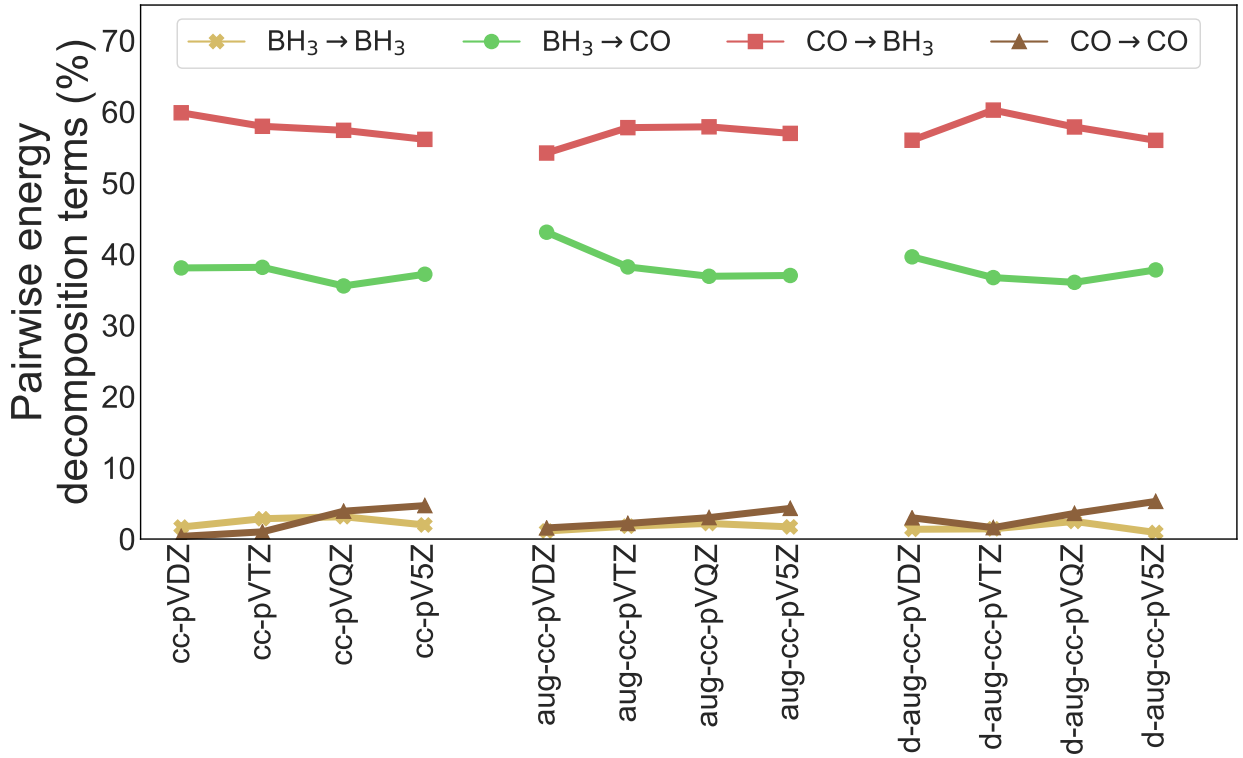


Figure 3.3: Convergence properties of the non-perturbative pairwise decomposition components of the total CT energy with respect to increasing the highest angular momentum of the basis set for the Dunning basis set sequence: cc-pVXZ, aug-cc-pVXZ, and d-aug-cc-pVXZ (X=D, T, Q, and 5) for the BH_3–CO system at its equilibrium geometry using ωB97X-D.

oligonucleotide sequences which is a critical parameter in molecular biology experiments.[262] The Watson-Crick base pairs, adenine-thymine (A-T) and guanine-cytosine (G-C), shown in Fig. (3.5), interact by characteristic hydrogen bonding. The A-T base pair is bound by two hydrogen bonds: one from N3 of thymine to N1 of adenine, and the other from N6 of adenine to O4 of thymine. The G-C binding energy is −120.8 kJ/mol which is much larger than the A-T binding energy of −69.4 kJ/mol as G-C has three hydrogen bonds (two H-bond donors on G and one donor on C) while A-T has only two.

The pairwise CTA for these base pairs is shown in Table (3.1). From a CTA perspective, the G-C pair really has two N lone pair donors on cytosine, coupled to two charge accepting H−N groups on guanine, rather than the opposite proton donor perspective. This is reflected in the larger C→G term versus the smaller G→C contribution. The main Complementary Occupied-Virtual Pairs (COVPs) shown in Fig. (3.6) for the A-T pair are a very convenient way to understand these hydrogen bonds. The most significant COVP of the A-T base pair contributes a majority (about 67%) of the CT interactions. The COVP donor is located on

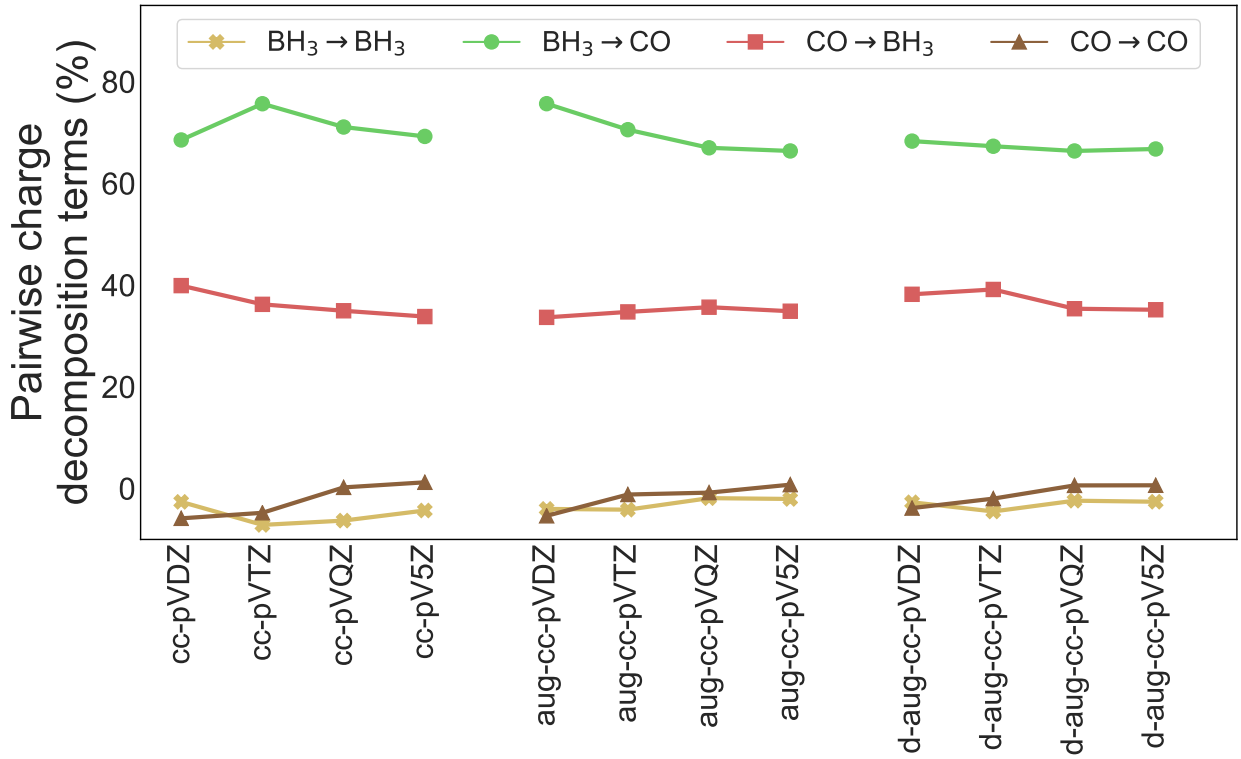


Figure 3.4: Convergence properties of the non-perturbative pairwise decomposition components of total charge transfer with respect to increasing the highest angular momentum of the basis set for the Dunning basis set sequence: cc-pVXZ, aug-cc-pVXZ, and d-aug-cc-pVXZ (X=D, T, Q, and 5) for the BH_3–CO system at equilibrium geometry using ωB97X-D.

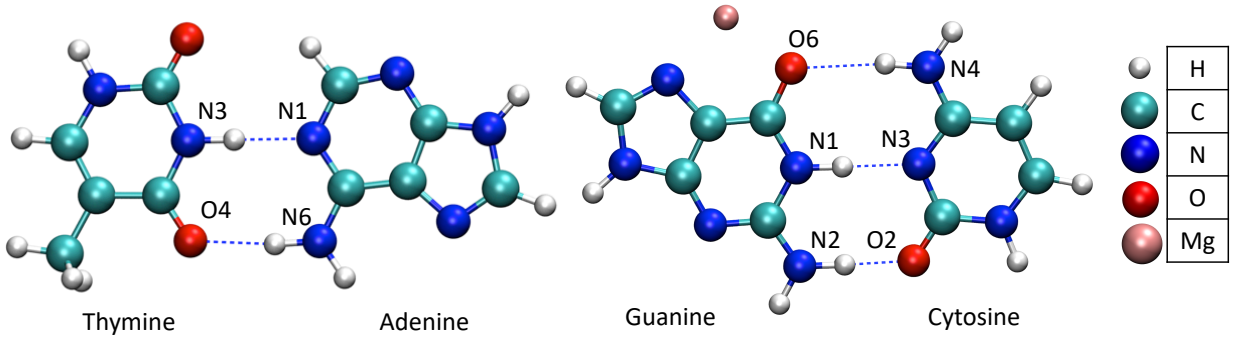
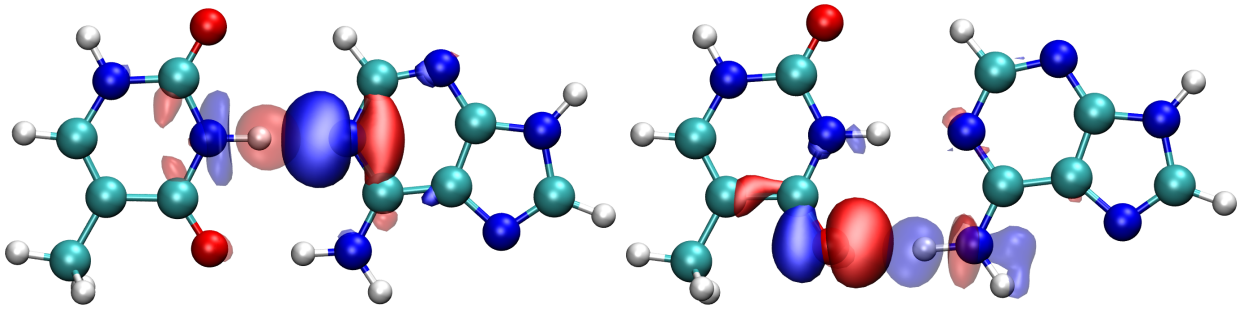


Figure 3.5: DNA base pairs adenine-thymine (A-T; left panel) and guanine-cytosine (G-C) with Mg^{2+} (right panel)

Table 3.1: Energy decomposition analysis and non-perturbative charge decomposition analysis (in kJ/mol) for the DNA base pairs thymine(T)-adenine(A), guanine(G)-cytosine(C), and their corresponding metallated versions.

	Energy decomposition analysis					Non-perturbative decomposition of CT energy			
	ΔE_{INT}	ΔE_{GD}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}	1→1	1→2	2→1	2→2
T(1):A(2)	-63.8	5.6	-7.4	-26.2	-35.8	0.0	-10.7	-25.4	0.3
G(1):C(2)	-120.8	13.5	-35.6	-52.2	-46.5	0.0	-19.0	-27.6	0.2
Na^+ G(1):C(2)	-138.2	11.0	-45.4	-55.8	-48.1	-0.2	-10.2	-38.0	0.4
Mg^{2+} G(1):C(2)	-195.7	18.5	-56.3	-90.9	-67.1	-0.3	-5.3	-62.2	0.7
Ca^{2+} G(1):C(2)	-178.9	16.3	-54.6	-80.9	-59.7	-0.3	-5.4	-54.6	0.6



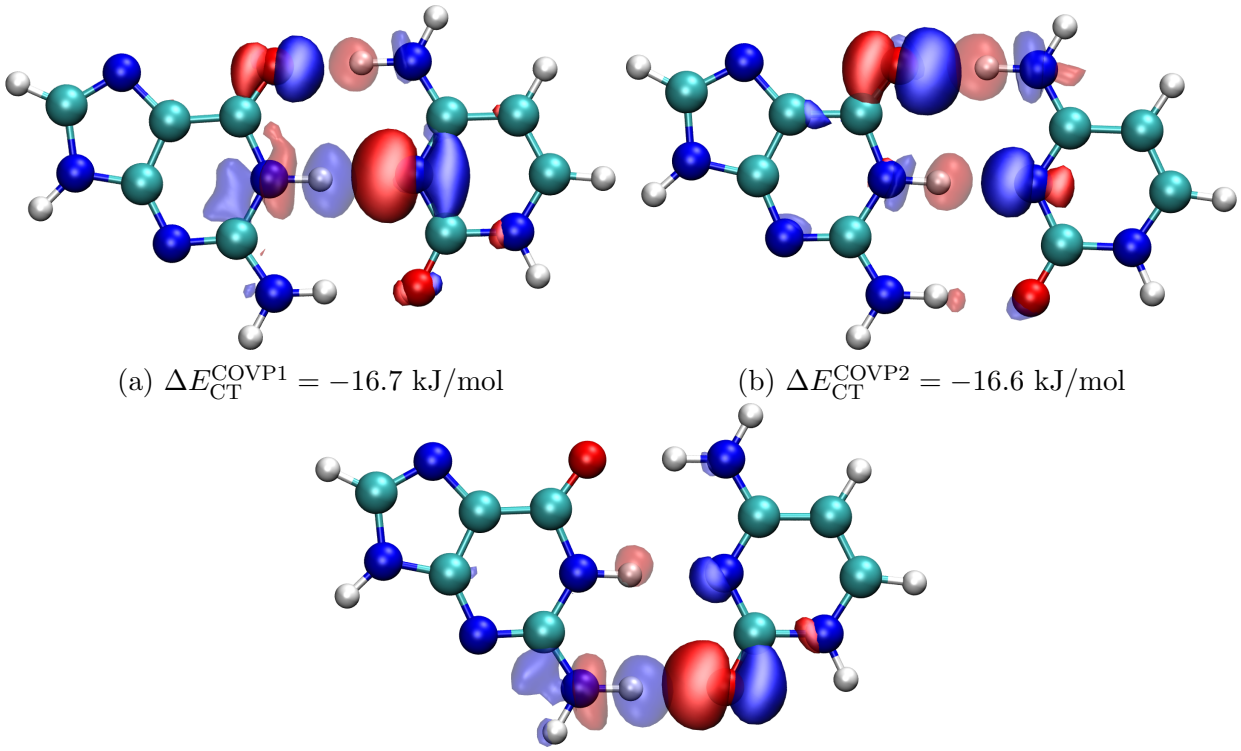
(a) $\Delta E_{\text{CT}}^{\text{COVP1}} = -23.9$ kJ/mol (b) $\Delta E_{\text{CT}}^{\text{COVP2}} = -8.7$ kJ/mol

Figure 3.6: (a) The most significant Complementary Occupied-Virtual Pair (COVP) for the A-T DNA base pair (b) Second most significant COVP for the A-T DNA base pair. Atom color codes are shown in Fig. (3.5).

N1 of adenine and the COVP acceptor is located on N3−H of thymine. The second hydrogen bond has much weaker CT, which accounts for the remaining CT stabilization.

By contrast, as shown in Fig. (3.7), the G-C base pair with its three hydrogen bonds contains two equally significant COVPs, each of which contributes −16.6 kJ/mol to the CT energy. Unlike the A-T COVPs, these G-C COVPs do not completely localize on any one particular hydrogen bond although the first pair (shown in Fig. (3.7a)) is mainly on the O6-N4 hydrogen bond, while the second pair (shown in Fig. (3.7b)) dominates the N3−N1 interaction. For two almost degenerate COVPs such as this, it is possible to localize the two occupied and two virtual orbitals if a more localized picture is desired.

The presence of metal cations has a significant effect on biophysical processes such as stabilization of DNA triple and quadruple helices. At low concentrations, metal cations produce a stabilizing effect by neutralizing the negatively charged phosphate backbone. However, at high concentrations they affect the structural integrity of DNA by disrupting the hydrogen bonding interactions.[263, 264] Metal cations interact with the N7 position of guanine as shown in the right panel of Fig. (3.5).[265–267] An understanding of the effect of metal cation coordination on the binding energy of DNA base pairs will help illuminate whether



(a) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP1}} = -16.7$ kJ/mol

(b) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP2}} = -16.6$ kJ/mol

(c) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP3}} = -9.3$ kJ/mol

Figure 3.7: (a) The most significant Complementary Occupied-Virtual Pair (COVP) for the G-C DNA base pair (b) Second most significant COVP for the G-C DNA base pair (c) Third most significant COVP for the G-C DNA base pair. Atom color codes are shown in Fig. (3.5).

this 3-body interaction affects the stability of the DNA duplex. The effect of metal cation coordination of the strength of hydrogen bonds in DNA base pairs has been studied using Natural Bond Orbital Analysis.[268]

As a primitive model for this effect, we consider guanine of a G-C dimer binding in a bidentate fashion via its N7 and O6 sites to three metal cations (Na^+, Mg^{2+}, and Ca^{2+}), as shown for Mg^{2+} in the right panel of Fig. (3.5). The EDA and non-perturbative CTA of the three metallated G-C complexes are shown in Table (3.1). All 3 metallations of guanine increase the binding energy of the complex. This increment is small (17.4 kJ/mol) when the metal cation is Na^+, but quite large when the metal cation is Mg^{2+} or Ca^{2+}. The EDA reveals that this increment comes from increases in all major mechanisms of interaction: frozen interactions, polarization, and charge transfer. The adiabatic EDA[250] (see Table B.1) confirms that the increase in frozen and polarization interactions is larger than the increase in CT. The predominance of permanent and induced electrostatics agrees with chemical understanding of this toy model of metallated DNA base pairs, which is

Table 3.2: EDA and non-perturbative CT energy decomposition analysis (in kJ/mol) for the series of adducts BX_3–NH_3 where X = F, Cl, or Br.

	Energy Decomposition Analysis					Non-perturbative decomposition of CT energy			
	ΔE_{INT}	ΔE_{GD}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}	$BX_3 \rightarrow BX_3$	$BX_3 \rightarrow NH_3$	$NH_3 \rightarrow BX_3$	$NH_3 \rightarrow NH_3$
BF_3–NH_3	-90.8	99.7	122.8	-157.4	-156.0	0.0	-5.1	-151.2	0.2
BCl_3–NH_3	-107.7	101.2	259.7	-282.2	-186.4	0.2	-9.8	-179.3	2.6
BBr_3–NH_3	-119.2	94.3	295.5	-322.6	-186.5	0.2	-11.6	-177.8	2.7

controlled by interactions with the unscreened charge.

Considering the case of Mg^{2+}-G-C versus G-C, frozen interactions, polarization, and CT contribute an additional 20.6 kJ/mol, 38.7 kJ/mol, and 20.6 kJ/mol respectively to the interaction energy. Applying the non-perturbative CTA to the charge transfer contribution reveals an interesting pattern. Metallation of guanine increases the C→G CT energy while it decreases the G→C CT. This serves to decrease the strength of the O6(G)$\cdots$H-N4(C) hydrogen bond and increase the strength of the N3(C)$\cdots$H-N1(G) and O2(C)$\cdots$H-N2(G) hydrogen bonds. The increase in the strength of the latter CT interactions is larger than the decrease of the former, thereby leading to an overall increase in CT upon metallation. COVPs for the metallated G-C base pairs are included in the Chapter B (See Figs. B.3, B.4, and B.5). The significant COVPs are localized on the N3-N1 and O2-N2 hydrogen bonds, while the COVP localized on the O6-N4 hydrogen bond contributes little to the total CT energy.

These aspects of hydrogen bonding in nucleobases have been previously studied using delocalization indices from Atom-In-Molecule (AIM) theory.[269] While it is difficult to compare our method with AIM theory, both analysis methods agree on the selective strengthening and weakening of hydrogen bonds upon metallation in the G-C base pair. While AIM theory measures this change in terms of change in delocalization index, our analysis can directly compute an energy value associated with these interactions, thus enabling a richer and more direct comparison to the interaction energy.

Borane-amine adducts

Borane-amine adducts are textbook examples of Lewis acid-base pairs. The ammonia-borane complex has been studied particularly in detail as it has been considered a promising hydrogen storage material that contains 19.6 wt% of hydrogen.[270, 271] This adduct consists of an electron-deficient group 13 center and an electron-rich group 15 center. Traditionally, bonding in this adduct has been understood as a dative bond as a result of donation of an electron pair from the nitrogen to the boron center. This is supported by the fact that borane, which is a planar molecule, pyramidalizes upon complexation with ammonia.

Understanding the nature and strength of the dative bond in these adducts is key to tuning the strength of such dative interactions, and can potentially be used to design ligands

Table 3.3: Energy decomposition analysis and non-perturbative charge decomposition analysis (in kJ/mol) for the adduct BH_3–NMe_pH_q ($p+q=3$)

	Energy Decomposition Analysis					Non-perturbative Decomposition of CT energy			
	ΔE_{INT}	ΔE_{GD}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}	$BH_3 \rightarrow BH_3$	$BH_3 \rightarrow NMe_pH_q$	$NMe_pH_q \rightarrow BH_3$	$NMe_pH_q \rightarrow NMe_pH_q$
BH_3–NH_3	-133.1	56.3	116.5	-148.1	-157.7	-2.7	-16.4	-138.1	-0.5
BH_3–$NMeH_2$	-153.2	59.7	114.0	-172.9	-154.0	-3.4	-17.4	-131.4	-1.9
BH_3–NMe_2H	-163.1	62.8	110.7	-185.9	-150.6	-3.7	-17.8	-127.7	-1.4
BH_3–NMe_3	-163.8	65.8	107.7	-178.4	-159.0	-3.6	-18.1	-136.5	-0.8

on a catalyst or engineer protein-drug interactions. One common example is the series of halogenated boranes binding ammonia: BX_3–NH_3 where X = F, Cl, or Br. The order of stability of these adducts is BBr_3–NH_3 > BCl_3–NH_3 > BF_3–NH_3 as shown in Table (3.2). This ordering is rather counter-intuitive as one would naively expect the reverse ordering consistent with ordering of electronegativity of the halogens (F > Cl > Br). One explanation proposed for this ordering is that electron donation from the halogen to the empty p-orbital of boron reduces the Lewis acidity of boron.[272] As shape and sizes of B and F match better than B and Cl, it was suggested that donation from F to B was stronger. While this explanation suggests that BF_3 has a stronger π-bond character than BCl_3, and consequently should have a higher energy of pyramidalization, the opposite has been found to be true (See Ref. 272 and ΔE_{GD} in Table (3.2)). It was also found that the Natural Bond Orbital π-overlap between F and B in BF_3 and Cl and B in BCl_3 are more or less identical.[273] Another explanation correlating the Lewis acidity of halogenated boranes with its LUMO level was proposed[274, 275] and then contradicted.[273]

In this section, we attempt to explain the trend in the strength of interaction of trihaloborane-ammonia adducts. The interaction energy increases upon going from BF_3–NH_3 to BBr_3–NH_3 from −90.8 to −119.2 kJ/mol. Upon going down the periodic table from BF_3–NH_3 to BCl_3–NH_3, the polarization and charge transfer components of the interaction energy increase (Table (3.2)). While the increase in polarization is cancelled out by the increase in the repulsive frozen interactions, charge transfer causes a true increase in the strength of interaction. Upon decomposition of CT into pairwise additive terms, it is clear that the increase in the $NH_3 \rightarrow BX_3$ component is the most significant and leads to enhanced binding energy. On the other hand, upon going from BCl_3–NH_3 to BBr_3–NH_3, total CT and the $NH_3 \rightarrow BX_3$ component remain unchanged. Most of the increase in binding energy arises from a decrease in the geometry distortion term. While most of the increase in the polarization energy is offset by the increase in repulsive frozen interactions, the increase in the polarization term still contributes a little to the overall increase in the binding energy. Energy decomposition in an adiabatic picture (see Table B.3) further emphasizes the importance of CT relative to polarization and frozen interactions. The adiabatic energy decomposition also shows an increase in CT upon going from BF_3–NH_3 to BBr_3–NH_3.

Another interesting observation is that placing electron donating groups on B decreases the binding energy of the adduct, while placing the same group on N increases the binding

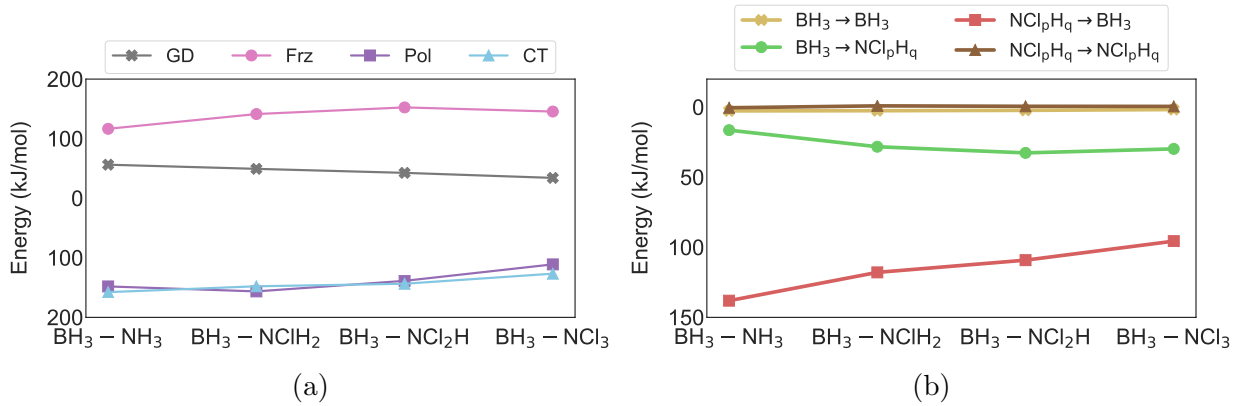


Figure 3.8: (a) Energy decomposition of BH_3–NCl_pH_q binding energy into geometry distortion (GD), frozen (Frz), polarization (Pol), and charge transfer (CT) terms. (b) Non-perturbative decomposition of charge transfer into pairwise additive terms. A table containing the energetics shown in this figure is included in the Chapter B (Table B.2).

strength of the adduct. This phenomenon can be classically understood as the enhancement of Lewis basicity when electron donating groups are placed on nitrogen. In order to gain a further understanding of this phenomenon, we consider a series of boranes bound to ammonia substituted with increasing numbers of methyl (−Me) groups. Methyl is an electron donating group, and with addition of each methyl group on the nitrogen center the interaction energy increases as shown in Table (3.3). Experimentally, gas phase dissociation enthalpies also increase.[276] Performing EDA on the interaction energies of this series, we can see that the total CT energy shows no trend with increasing the number of methyl groups on ammonia. EDA in the adiabatic framework is also in agreement with this although the relative magnitude of CT is much larger than the polarization (See Table B.4). This shows that understanding the enhancement in binding energy of these adducts in terms of increased Lewis basicity of substituted ammonia is incorrect. The CTA further supports this claim by not showing a clear trend in the CT energy associated with donation from NMe_pH_q to BH_3, which lies in the range of −138.1 to −127.7 kJ/mol, and does not become stronger with increasing number of methyl substituents at N. By contrast, the polarization energy increases with increasing number of methyl substituents and is thus the main origin of enhanced binding. The reason is likely because −Me is more polarizable than −H, and consequently more −Me groups cause larger polarization interactions.

Now, let us consider a chloride (−Cl) substitution at N instead of −Me substitution. The chloride, being an electron withdrawing group, reduces the electron density available for donation in forming an adduct. As the number of chloride groups on N increases, the binding energy of the adduct decreases from −133.1 kJ/mol in the case of BH_3–NH_3 adduct to just −58.2 kJ/mol for BH_3–NCl_3. EDA shown in Fig. (3.8a) shows that this is a result of the decrease in both polarization and charge transfer. The non-perturbative CTA reveals

two important trends upon chloride substitution: First, as expected, $NCl_pH_q \rightarrow BH_3$ CT decreases significantly from −138.1 kJ/mol to −95.6 kJ/mol on going from NH_3 to NCl_3. Second, $BH_3 \rightarrow NCl_pH_q$ CT energy increases upon chloride substitution. However, the latter increase is rather small and is eclipsed by the former leading to an overall decrease in the CT energy lowering. Thus, the non-perturbative CTA can elucidate the interplay of different pairwise CT energy components in order to explain the observed binding energies of borane-amine adducts.

Carbonyl complexes

Transition metal – carbonyl interactions are ubiquitous in transition metal chemistry and catalysis,[277, 278] such as in important intermediates in CO_2 reduction. [252] These are cases of synergic bonding with significant binding arising from both forward electron donation (carbonyl to metal complex) and backward (metal complex to carbonyl) donation. The Dewar-Chatt-Duncanson[229, 230] model has been invoked to explain bonding in metal carbonyls with differing degrees of success.[279] According to this model, the red shift in the carbonyl stretching frequency is understood as a consequence of back-donation from the metal to the anti-bonding $2\pi^*$ orbitals of CO, decreasing its bond strength, lengthening the C−O bond, and consequently decreasing its vibrational stretching frequency.

Free CO is IR active and its stretching frequency appears at 2143 cm^{-1}.[280] Consider the sequence of isoelectronic 3d transition metal hexacarbonyls: $V(CO)_6^-$, $Cr(CO)_6$, and $Mn(CO)_6^+$.[281, 282] The electron density on the metal decreases as we go from anionic $V(CO)_6^-$ to cationic $Mn(CO)_6^+$. As a consequence, the metal has less electrons to donate to the $2\pi^*$ orbitals of CO. Figure (3.9) shows that the non-perturbative $M(CO)_5 \rightarrow CO$ CT decreases from −172.9 kJ/mol in $V(CO)_6^-$ to −78.7 kJ/mol in $Mn(CO)_6^+$. The experimental red shift also decreases in the same order: 285 cm^{-1} for $V(CO)_6^-$, 140 cm^{-1} for $Cr(CO)_6$, and 37 cm^{-1} for $Mn(CO)_6^+$. The computed red shifts show a similar trend: 262 cm^{-1} for $V(CO)_6^-$, 127 cm^{-1} for $Cr(CO)_6$, and 11 cm^{-1} for $Mn(CO)_6^+$. As back-donation is a major component of the CT interaction, total CT energy also decreases as one moves to the right in the 3d transition metal sequence. The non-perturbative CTA also reveals the contrasting mechanisms dominating the nature of CT in $V(CO)_6^-$ and $Mn(CO)_6^+$: back donation is the major component of CT contributing about 73% of the CT energy in $V(CO)_6^-$, whereas forward donation is the major driving force accounting for 55% of the charge transfer energy in $Mn(CO)_6^+$. Thus, the CTA reveals the intricate interplay between different mechanisms driving CT in 3d metal hexacarbonyls. The variational forward-backward CTA[252] also shows that the adiabatic back-donation ($M(CO)_5 \rightarrow CO$) decreases upon moving from $V(CO)_6^-$ to $Mn(CO)_6^+$ while the forward ($CO \rightarrow M(CO)_5$) component decreases (See Table B.5).

This example also brings out the main limitation of the perturbative CTA, which is capable of capturing only ∼67% of the total CT energy. In the $V(CO)_6^-$ case, for example, the perturbative treatment underestimates the forward component by only 21.5 kJ/mol compared to the non-perturbative counterpart, but makes a much larger error in the backward component underestimating it by 62.8 kJ/mol. This inconsistency in the magnitude

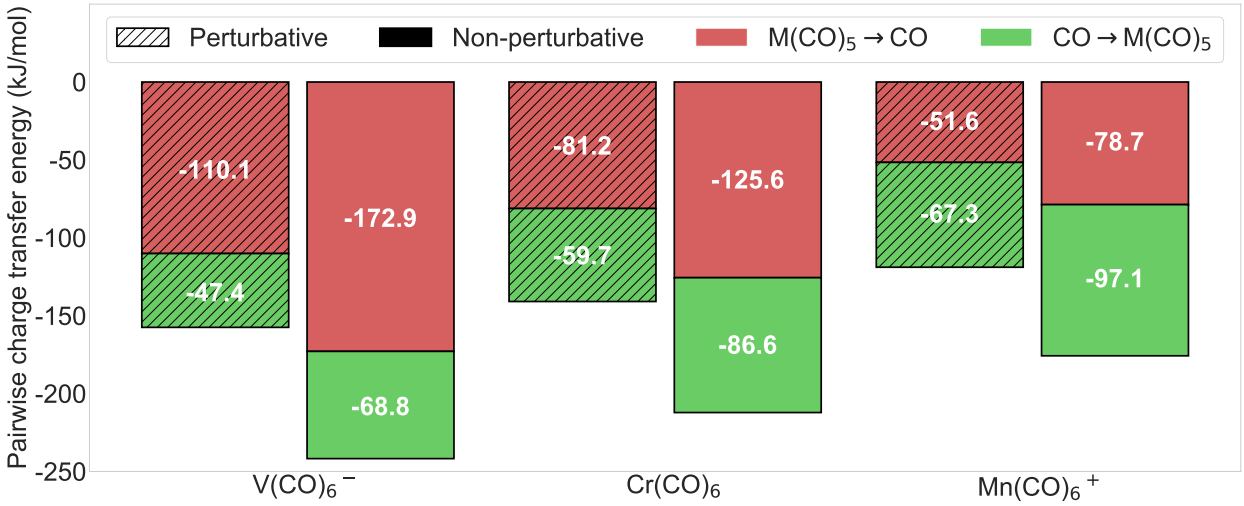


Figure 3.9: Stabilization energy associated with forward ($CO \rightarrow M(CO)_5$) and backward ($M(CO)_5 \rightarrow CO$) CT interactions in the perturbative and non-perturbative schemes for 3 isoelectronic transition metal hexacarbonyls. The $M(CO)_5 \rightarrow CO$ component is shown on the top and the $CO \rightarrow M(CO)_5$ is shown below it.

of underestimation in the perturbative scheme underscores the advantage of the new non-perturbative CTA.

Computational cost

As established above, the non-perturbative CTA is a superior alternative to perturbative analysis for studying charge decomposition. To address the computational cost of obtaining these improved results, we will briefly compare the cost of the non-perturbative scheme against its perturbative predecessor. The perturbative scheme computes $\mathbf{X}_{VO}^{\mathrm{RS}}$ by solving for the Roothaan step, Eq. (3.6), which involves only cubic-scaling matrix multiplications, where the dimension of these matrices involves the number of occupied (O) and virtual (V) orbitals of the complex. Given $\mathbf{X}_{VO}^{\mathrm{RS}}$, the pairwise energy decomposition follows by taking its trace with the Fock matrix constructed from the polarized ALMOs (Φ_{POL}). This Fock matrix is already constructed in the order to converge those orbitals and compute the energy of the polarized wavefunction ($E(\Phi_{\mathrm{POL}})$).

By contrast, the non-perturbative CTA involves solving Eq. (3.23) in order to compute the non-perturbative CT matrix, $\mathbf{X}_{VO}^{\mathrm{CT}}$. The process of solving this equation involves cubic scaling matrix diagonalization and multiplications of matrices whose dimension is the number of MOs, $N = O + V$ of the complex. Given $\mathbf{X}_{VO}^{\mathrm{CT}}$, the non-perturbative pairwise energy decomposition requires computation of the $\mathbf{F}^{\mathrm{CT}}$ using Eq. (3.27). This requires five different Fock matrices at different interpolating density points between Φ_{POL} and Φ_{CT}. Two of

these five Fock matrices ($\mathbf{F}(0)$ and $\mathbf{F}(\mathbf{X})$ in Eq. (3.27)) are available from computing the energy of the polarized wavefunction and energy of the fully-relaxed complex, leaving three additional Fock matrix constructions. Given that the complete ALMO-EDA scheme typically requires tens of Fock matrix constructions for determining the polarized and fully-relaxed wavefunctions, the addition of three more Fock matrix constructions only adds very little to the total computational cost.

3.5 Conclusions

We have introduced a non-perturbative approach that (numerically) exactly decomposes the variational energy lowering due to charge transfer (CT) in molecular complexes into pairwise additive terms in the context of ALMO-based energy decomposition analysis (ALMO-EDA) for density functional theory.[126, 202] The non-perturbative CT analysis (CTA) is a superior replacement for the existing perturbative CTA,[107] which relied on a perturbative approximation to incompletely and inexactly extract the pairwise contributions. As demonstrated here, the perturbative CTA decomposes different percentages of the total charge transfer energy at different points on a potential energy surface, and also has a dependence on the density functional used, typically underestimating CT for hybrid and range-separated density functionals while overestimating it for pure functionals.

This new method introduced in this work finds the generator, $\mathbf{X}_{VO}^{\mathrm{CT}}$, of the unitary transformation that transforms the polarized wavefunction into that of the fully-relaxed wavefunction. Next, an effective Fock-like matrix, $\mathbf{F}^{\mathrm{CT}}$, is constructed from Fock matrices computed at different density points connecting the polarized wavefunction and fully-relaxed wavefunction along the path given by $\mathbf{X}_{VO}^{\mathrm{CT}}$. Taking the trace of the product of the $\mathbf{F}_{OV}^{\mathrm{CT}}$ and $\mathbf{X}_{VO}^{\mathrm{CT}}$ along with appropriate orbital projectors sandwiched in the middle gives the non-perturbative charge decomposition in terms of donor and acceptor orbital pairs of the fragments comprising the complex. This completes the (numerically) exact pairwise energy decomposition scheme for CT. We also extended the CTA to decompose the charge flow in a pairwise fashion as well. Finally, singular value decomposition (SVD) of $\mathbf{X}_{VO}^{\mathrm{CT}}$ yields complementary occupied-virtual orbital pairs (COVPs), which are the most important orbitals involved in CT.

We have demonstrated the usefulness of the new CTA by applying it to various chemical systems with varying strengths of charge transfer interaction. Application to DNA base-pairs reveals the nature of hydrogen bonding in the thymine:adenine and guanine:cytosine complexes. Additionally, CTA reveals the effect of metallation on the hydrogen bonding patterns of the guanine:cytosine base-pair. Investigating the interaction energy of a series of borane adducts of the form $BX_3{-}NH_3$ (X = F, Cl, or Br) reveals that an increase in polarization and CT energy from NH_3 to BX_3 enhances the binding energy of the $BX_3{-}NH_3$ complex as we go down the halogen group. The CTA also revealed interesting aspects of the effect of methyl and halogen substitution on the nitrogen center. Additionally, the CTA provided insight into the delicate interplay of forward and backward donation in a series of isoelectronic transitional metal hexacarbonyls. It is likely to be useful for a wide variety of

other interpretive problems in intermolecular interactions, as well as other applications such as for training force-fields to account for pairwise decomposition of charge transfer.[108, 109]

Chapter 4

Benchmarking Density Functionals for Hydrogen Storage: The H2Bind275 Dataset

Reprinted with permission from Veccham, S. P.; Head-Gordon, M. Density Functionals for Hydrogen Storage: Defining the H2Bind275 Test Set with Ab Initio Benchmarks and Assessment of 55 Functionals *J. Chem. Theory Comput.* **2020**, *16*, 4963–4982

4.1 Introduction

Hydrogen (H_2) is a promising green alternative to the fossil fuel-based energy economy. H_2 is a clean fuel as the only byproduct it produces is water, which is non-toxic and unlike CO_2 has no deleterious effects on the environment. The gravimetric energy capacity of hydrogen is also about three times that of gasoline, providing a mass-efficient way to store and transport energy. One major problem with using H_2 as fuel is the low volumetric capacity of hydrogen which is one of the biggest hurdles in achieving a H_2-based economy.

One solution to this volumetric capacity obstacle is to store hydrogen in tanks under high pressure. Hydrogen fuel cell vehicles with high pressure carbon fiber-reinforced tanks are already on the market. However, safely storing H_2 under high pressure adds a substantial cost penalty, not to mention the additional safety complications associated with accidents. Nanoporous materials with high surface area offer a viable alternative to this storage problem by storing H_2 in adsorbed form. In principle, a given quantity of H_2 can be stored by adsorption at a much lower pressure in comparison to pure compression. Despite major synthetic and design efforts in the nanoporous framework community, none of the materials designed till date have been able to achieve all the targets for a viable storage material compiled by the U.S. Department of Energy (DOE). [283] While some recently synthesized materials have come close to the target,[128] novel design strategies and new material discoveries are still required to achieve the DOE targets.

Nanoporous sorbent materials, including Metal-Organic Frameworks (MOFs) and Covalent Organic Frameworks (COFs), represent the vast majority of storage materials proposed to date for hydrogen. For example, H_2 storage by binding to open metal sites with unsaturated coordination spheres in MOFs has successfully been explored in MOF-11,[284] M_2(dobdc),[285] and Mn-BTT.[286] Covalent Organic Frameworks which are robust both structurally and thermally, represent another paradigm for H_2 storage.[287] Enhancement of hydrogen storage capacities by impregnation with metal ions is an active area of research. Graphene, graphene oxide, and graphene origami are new carbon-based adsorbents that have been explored for hydrogen storage.[288–290] Given this huge diversity in the binding moeities of H_2, we feel it is necessary to have a dataset that captures this variety appropriately. The dataset compiled in this work aims to capture the diversity only in materials that store H_2 by adsorption, and not other materials which store H_2 by hydrogenation/dehydrogenation of molecules.[291]

In silico tools have become indispensible in the design of drugs, proteins, and materials with desired properties using both rational and high-throughput design strategies. In the context of designing materials for hydrogen storage, researchers are seeking frameworks with the highest usable capacity for a predetermined operating pressure range. The usable capacity is the difference between the amount of hydrogen adsorbed at high and low operating pressures. The usable capacity is completely determined by the adsorption isotherm of H_2, which in turn is influenced by the Gibbs free energy of adsorption, pore volume and surface area in the region where Henry's law is applicable. The Gibbs free energy of adsorption ΔG_{ads}, in turn, can be determined by the enthalpy (ΔH_{ads}) and entropy (ΔS_{ads}) of adsorption at a given temperature (T) according to Eq. (4.1)

$$\Delta G_{ads} = \Delta H_{ads} - T\Delta S_{ads} \tag{4.1}$$

Enthalpy and entropy are known to be roughly correlated, even though the exact relationship depends on the adsorbent material. [130, 131, 292] The most important component of enthalpy is the internal energy of binding at 0K which can be predicted with very good accuracy using different first principles tools of electronic structure theory. Common first principles tools that can be used to predict binding energy are Hartree Fock (HF), Density Functional Theory (DFT), second order Møller-Plesset Perturbation Theory (MP2), Configuration Interaction (CI), and different variants of Coupled Cluster theory (CC). These methods can predict binding energies with different degrees of accuracy. The associated computational cost is also very different with HF and DFT scaling as $\mathcal{O}(N^3)$, MP2 as $\mathcal{O}(N^5)$, and Coupled Cluster with Singles, Doubles and Perturbative Triples (CCSD(T)) as $\mathcal{O}(N^7)$. Of these methods, CCSD(T) provides the highest accuracy: typically yielding sub-kJ/mol accuracy for non-covalent interaction energies.[293] However, the need for thousands of calculations in a high-throughput screening along with its steep scaling in computational cost makes it impossible to use coupled cluster theory for designing materials.

DFT, with $\mathcal{O}(N^3)$ scaling, lies at a reasonable compromise between accuracy and cost for predicting binding energies. While in principle, DFT can provide an exact solution to

the Schrodinger equation under the Born-Oppenheimer approximation, the exact form of this functional remains unknown.[16] Over the last couple of decades, various groups across the world have proposed different approximations to this functional form using empirical, semi-empirical, and non-empirical arguments. There are hundreds of these approximations, termed density functional approximations(DFAs), and each of them provides different accuracies for different problems including non-covalent interactions, reaction barrier heights, thermochemistry, and ionization energies. Some density functionals have also been parameterized to predict properties of certain classes of chemical species limiting their applicability to other classes. In fact, different DFAs have previously been used to predict the deliverable capacity for various storage materials. [294–299] Even though the performance of DFAs have been thoroughly benchmarked [17, 18, 300, 301] for each of these properties (represented by one or multiple datasets), none of them thoroughly represents the hydrogen storage problem. In fact, to the best of our knowledge, none of the datasets even contain the hydrogen molecule. In this paper, we have compiled a dataset that captures the diversity of binding sites encountered in materials that adsorb hydrogen.

Another computational approach to determine the usable capacity of a material is using Grand Canonical Monte Carlo (GCMC) simulations which rely on force fields to accurately reproduce the interactions between the adsorbate (H_2) and the binding motifs. As these simulations need a large number of energy evaluations, they are usually performed using force fields. It is also possible to perform simulations directly using DFT at the expense of greater computational cost.[302] Parameterization of force fields require highly accurate ab initio data which can be generated, in principle, using density functional approximations.

Cluster modeling, that is reducing the size of the binding site to a few atoms at the immediate vicinity of the bound H_2 employing suitable chemical truncation, is often used to make extended adsorbent materials with well-defined binding sites computationally tractable.[295, 303, 304] The alternative to cluster modeling is to use periodic boundary conditions which would severely limit the usage of highly accurate wavefunction methods. In this work, we have a employed cluster models in order to obtain reference interaction energies with high accuracy. Small cluster models also allow for modeling of the binding site instead of a particular binding material, thereby increasing the transferability of the conclusions to other binding materials with similar binding sites.

This paper is organized as follows: The H2Bind275 dataset is introduced and its ability to comprehensively represent the diversity of H_2 binding sites and mechanisms is discussed. The protocol used for arriving at accurate reference interaction energies using coupled cluster theory with the focal point analysis is introduced. After a brief discussion of the performance of wavefunction methods, the performance of 55 density functionals is assessed and the best performing density functional approximations are identified. This assessment is performed for the whole H2Bind275 dataset and different chemical categories of the dataset, motivated by H_2 storage applications. The effect of addition of exact exchange and empirical dispersion corrections to DFAs is analyzed. The problems associated with using finite-sized basis sets for DFT calculations are discussed.

4.2 Computational details

All the density functional approximation interaction energies were computed using Q-Chem 5.0.[176] The reference interaction energies including MP2, CCSD, and CCSD(T) were also computed using Q-Chem 5.0. Post CCSD(T) computations, specifically CCSDT(Q), were performed using the MRCC program which is capable of performing arbitrary order coupled-cluster theory.[305–307] In order to ensure consistency across different quantum chemistry programs, we compared and verified HF and MP2 energies across both the programs. The CCSDT(Q) calculations were performed using 6-31G**(mod), a modified 6-31G** basis set where 0.35, 0.25, and 0.15 are the exponents of the f-type, d-type, and p-type polarization functions respectively. This basis set was found to give outstanding results for non-covalent interaction energies in the A24 dataset despite its small size.[308, 309]

Mechanism of H_2 binding

The factors involved in H_2 interaction with binding motifs can be broadly classified into two categories: (1) physisorption (2) chemisorption. Physisorption includes the electrostatic and induced electrostatic interactions of H_2. H_2 is an uncharged molecule with a permanent quadrupole moment. As a consequence, permanent electrostatics are expected to be very weak. However, a dipole can be induced in the presence of a strong electric field and H_2 can interact with the binding site through polarization. However, H_2's large HOMO-LUMO gap of 11.19 eV makes this interaction difficult and significant polarization only occurs in the presence of very strong electric fields, like those created by unscreened charges of open metal sites.[304] Another significant mechanism of interaction is charge transfer which occurs as a part of chemisorption. Charge transfer from the σ_g-bond orbital of H_2 to the binding site is the most common form of interaction as it can occur with any binding site. H_2 can also participate in a Kubas-like synergistic interaction wherein it interacts with transition metals through both forward ($\sigma_g(H_2) \rightarrow d(M)$) and backward donation ($d(M) \rightarrow \sigma_u^*(H_2)$). [310–312] While these Kubas-like interactions are too strong for H_2 storage applications, interaction tuning by changing the ligand framework around the transition metal can still make this mechanism important for storage applications.

H2Bind275 Dataset

As none of the existing non-covalent interaction energy datasets represent the H_2 binding problem comprehensively, we have compiled a new dataset which adequately represents the diversity in the H_2 binding motifs and mechanisms of interaction. Each geometry in the dataset, here after referred to as H2Bind275, consists of representative binding motifs interacting with one or multiple H_2 molecules. Development of materials with multiple H_2 binding at a single site is a promising strategy for making materials with enhanced uptake.[313, 314] The H2Bind275 dataset has many data points with multiple H_2s binding to a single site, and is consistent with this encouraging paradigm. Each geometry has been optimized using

the density functional ωB97M-V[44] in the def2-TZVPD basis set[173] in order to ensure maximum interaction with the binding the motif. The optimized geometry has been confirmed to be a minimum on the potential energy surface by ensuring that the hessian has no negative eigenvalues. All the geometries chosen are also in their ground spin state at equilibrium. We have also assessed that using ωB97M-V geometries does not induce any bias in the dataset (See Table C.4 and associated discussion for further information). The dataset can be broadly classified into four categories: (1) s-block ions (2) salts (3) organic ligands (4) transition metals.

s-block ions: This category consists of alkali and alkaline earth metal ions with an unscreened charge binding one or multiple H_2 molecules. Post synthetic modification is a common strategy to incorporate H_2 binding entities in porous materials like MOFs. Open metal sites consisting of s-block metal ions have been proposed as promising candidates for exhibiting an enhanced uptake of H_2. [295, 315]. With regard to this, we have included Li^+, Na^+, Mg^{2+}, and Ca^{2+} with one or multiple hydrogens bound as a representative sample of s-block ions constituting open metal sites. This category consists of 19 unique geometries with 77 different interaction energies.

Salts: This category consists of small inorganic salts like AlF_3, $CaCl_2$, CaF_2, and MgF_2 binding one or multiple H_2 molecules. While it would be highly desirable to incorporate open metal sites with unscreened charge in porous materials, it is generally hard because of their high reactivity. Metals ions are thus found in their screened forms, and these four inorganic salts form a good representation of such binding motifs.

Organic ligands: Porous materials like MOFs and Covalent Organic Frameworks consist of organic linker molecules connecting metal ions or clusters. While linkers might not be the main binding sites contributing to H_2 uptake, they represent a large surface area that the H_2 interacts with. This set consists of both aromatic and aliphatic organic ligands interacting with one H_2 molecule.

Transition metals: Many MOFs binding H_2 consist of transition metals or transition metal clusters at secondary binding units or open metal sites. This subset consists of monocationic 3d transition metals binding 1–4 H_2 molecules. For instance, Cr^+ bound to 1–4 H_2 belongs to this subset. Most of these species have been experimentally isolated and characterized and have also been theoretically well-studied. [316] In order to include species with screened charge, we have included species like hydrides, fluorides, and chlorides of copper, silver and gold, a few of which have been experimentally and theoretically studied earlier. [317–319] We have also included certain other species like CoF_3, $CrCl_2$, CuCN, $Ni(OH)_2$ in order to capture the diversity in the transition metal binding sites.

In summary, as shown in Table. C.5 and Fig. (4.1), the H2Bind275 dataset, consisting of 86 unique geometries and 275 interaction energies, represents the problem of hydrogen binding in porous frameworks reasonably comprehensively.

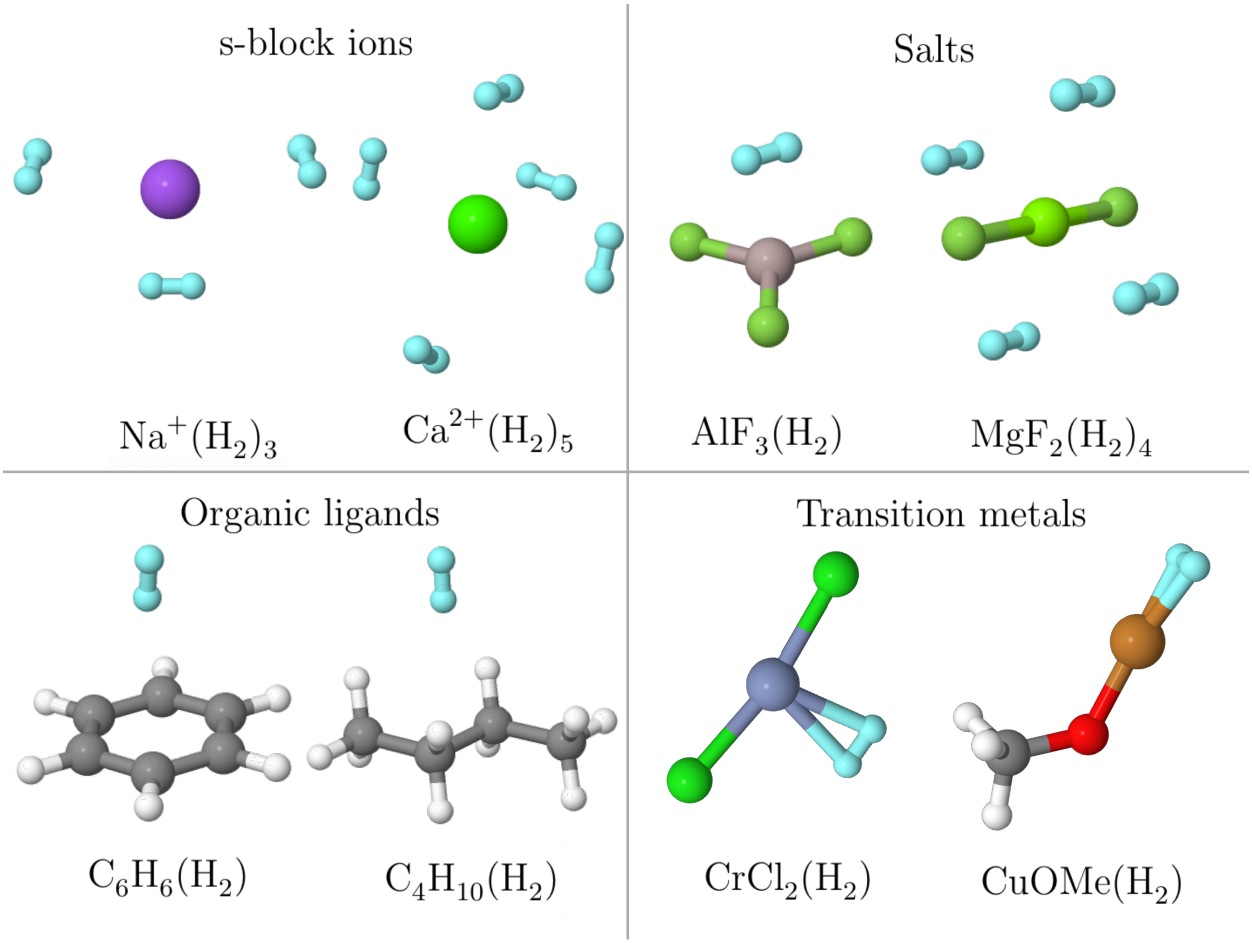


Figure 4.1: A figure showing instances of the H2Bind275 dataset categorized as s-block ions, salts, organic ligands, and transition metals.

Table 4.1: A table showing the number of unique geometries and interaction energies in each category of the dataset

	s-block ions	Salts	Organic ligands	Transition metals	Total
Geometries	19	13	5	49	86
Data points	77	44	10	144	275

Reference interaction energies

Coupled cluster theory was used to compute the reference interaction energies for this dataset. Coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) [60] is well-known to give high accuracy for non-covalent interactions and has been dubbed the "gold standard" of quantum chemistry.[293] Defining N as the number of basis functions, the steep $O(N^7)$ asymptotic scaling of computational cost of CCSD(T) together with the slow convergence of correlation energies with size of the basis set presents a huge hurdle to using CCSD(T) with large basis sets even for small molecules. In order to circumvent this problem, composite extrapolation procedures like the Gaussian-n models,[320–323] Weizmann-n models,[324–326] HEAT,[327] and ccCA[328] have been routinely employed. These procedures leverage the varying convergence rates of different components of interaction energy and additivity of different correlation energies to calculate molecular properties with sub-kJ/mol accuracy.

In this work, we use the composite CCSD(T) method as shown in (5.1).

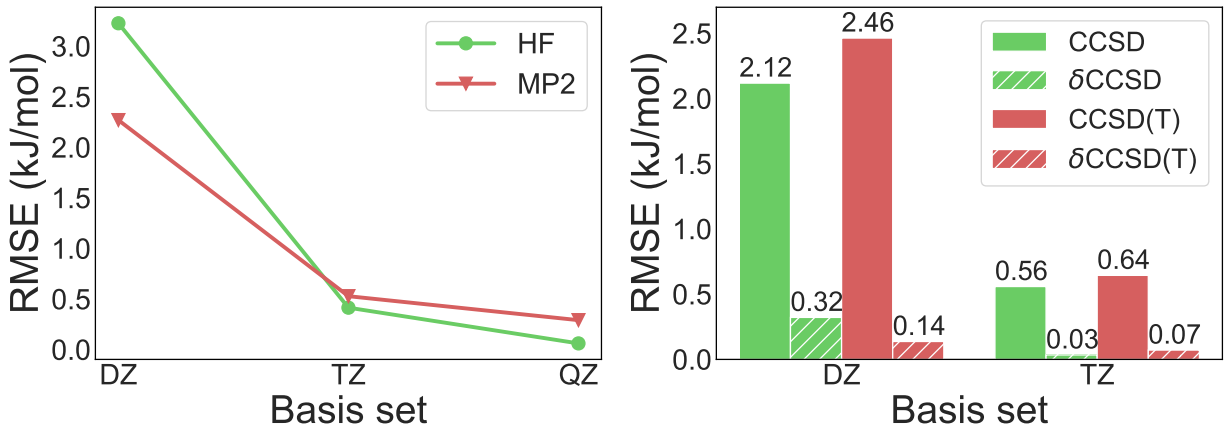
$$E_{\text{ref}} = E_{\text{HF/5Z}} + E_{\text{MP2/QZ}\rightarrow\text{5Z}} + \delta E_{\text{CCSD(T)/TZ}} + \delta E^{\text{core}}_{\text{MP2/TZ}} \tag{4.2}$$

$$\delta E_{\text{CCSD(T)/TZ}} = E_{\text{CCSD(T)/TZ}} - E_{\text{MP2/TZ}} \tag{4.3}$$

$$\delta E^{core}_{\text{MP2/TZ}} = E^{\text{core=0}}_{\text{MP2/TZ}} - E^{\text{core=n}}_{\text{MP2/TZ}} \tag{4.4}$$

In the composite scheme, the interaction energy is divided into the mean-field and correlation components. The mean-field component is computed using Hartree Fock (HF) theory with a basis set of quintuple-zeta quality and is labelled as $E_{\text{HF/5Z}}$. The correlation part is computed using the focal point analysis [329, 330] which is at the heart of many of the composite methods mentioned previously. The MP2 correlation energy is extrapolated to the basis set limit using quadruple-zeta and quintuple-zeta quality basis sets with the HKKN extrapolation formula[45] and is denoted as $E_{\text{MP2/QZ}\rightarrow\text{5Z}}$. The CCSD(T) correction to the MP2 correlation energy (labelled $\delta E_{\text{CCSD(T)/TZ}}$), defined as the difference between the CCSD(T) ($E_{\text{CCSD(T)/TZ}}$) and MP2 ($E_{\text{MP2/TZ}}$) correlation energies, was added at a triple-zeta quality basis. The effect of the core-valence electron correlation was computed using MP2 at a triple-zeta quality basis set and is denoted as $\delta E^{core}_{\text{MP2/TZ}}$. This effect is computed as the difference in the MP2 correlation energies with and without freezing core electrons (denoted as $E^{\text{core=n}}_{\text{MP2/TZ}}$ and $E^{\text{core=0}}_{\text{MP2/TZ}}$ respectively). Correlation consistent Dunning basis sets were employed to compute each of the components of the reference interaction energy. The cc-pVnZ[331, 332] family of basis sets were used when all core electrons were frozen in correlation calculations and the cc-pCVnZ[333, 334] family was used when only a part or none of the core was frozen. For transition metals, the cc-pwCVnZ[335] family was used with a neon frozen core. The effect of freezing core-valence correlations is discussed later in this section.

In this section, we attempt to isolate and characterize the different sources of error arising from using this protocol and subsequently justify the parameters used in this protocol. In the benchmark protocol fixed in Eq. (5.1), there can be three major sources of errors: (1) errors arising from using a finite basis set, (2) errors from neglecting higher order terms in



(a) (b)

Figure 4.2: (a) Basis set convergence properties of HF energy and MP2 correlation energies at double-zeta (DZ), triple-zeta(TZ), and quadruple-zeta (QZ) quality basis sets. Errors are with respect to quintuple-zeta basis sets (5Z) (b) Basis set convergence properties of CCSD and CCSD(T) terms contrasted with the convergence properties of δCCSD and δCCSD(T) terms at double-zeta (DZ) and triple-zeta (TZ) quality basis sets. The errors are with respect to the quadruple-zeta (QZ) quality basis sets.

coupled cluster theory, and (3) errors from using the frozen core approximation for correlated calculations.

In order to ensure that each of the terms in Eq. (5.1) is converged with respect to the size of the basis set, we investigated the convergence properties of each of these terms over a small subset of the H2Bind275 dataset as shown in Fig. (4.2). The SCF interaction energy converges roughly exponentially with the cardinality of the basis set which can be seen from Fig. (4.2)(a).[336] The RMSD between the HF interaction energies at quadruple-zeta and quintuple-zeta basis sets is only 0.06 kJ/mol and hence we have used the HF at quintuple-zeta quality basis set without any further extrapolation. The MP2 correlation energy, however, converges only polynomially with cardinal number, leading to a substantially larger difference of 0.29 kJ/mol between quadruple-zeta and quintuple-zeta basis sets. The HKKN formula used to extrapolate the MP2 correlation energy recovers an additional RMS interaction correlation energy of 0.30 kJ/mol that is not captured at the 5Z basis level. The convergence of the CCSD and CCSD(T) correlation energies are also slow with respect to the size of the basis set as shown in Fig. (4.2)(b). The deviation in the CCSD(T) interaction correlation energies at triple-zeta quality basis set is 0.64 kJ/mol (RMSD) from those at the quadruple-zeta quality basis set. The focal point analysis method used in Eq. (5.1) circumvents this problem by computing only the difference in correlation energy between CCSD(T) and MP2 ($\delta E_{\mathrm{CCSD(T)}}$ term). This difference term converges much more quickly with respect to the size of the basis set with a difference of only 0.07 kJ/mol between the triple-zeta and

quadruple-zeta quality basis sets. Of the binding motifs investigated in this study, the Mg^{2+} species binding one or multiple hydrogens will exert the strongest electric field on hydrogen(s) because of its large charge density. This strong electric field would distort and shift the density on H_2 more than other binding motifs. Diffuse basis functions are required to capture this shift in the density, and Mg^{2+} ion binding motifs should show the maximum deviation in the interaction energies between basis sets containing diffuse functions (aug-cc-pCVnZ) and those that do not (cc-pCVnZ). For the four Mg^{2+} species considered, the RMSD in the interaction energies is only 0.10 kJ/mol with a maximum percent deviation of 0.015%. This represents a soft upper bound to the expected deviation in the interaction energies if the cc-pCVnZ basis sets are replaced with their augmented counterparts. Given that this deviation is almost negligible and the heightened computational cost involved in using aug-cc-pCVnZ type basis sets, we have used cc-pCVnZ type basis sets for all the reference calculations in this work.

The effect of higher order excitations that were neglected in CCSD(T) was analyzed using CCSDT(Q) in the 6-31G**(mod) basis set which has been shown to yield accurate results for computing the $\delta E_{\text{CCSDT(Q)}}$ terms computed as shown in Eq. (4.5).

$$\delta E_{\text{CCSDT(Q)/6-31G**(mod)}} = E_{\text{CCSDT(Q)/6-31G**(mod)}} - E_{\text{CCSD(T)/6-31G**(mod)}} \tag{4.5}$$

For the 118 species whose HF solution matches up between MRCC and Q-Chem, the maximum $\delta E_{\text{CCSDT(Q)}}$ correction was found to be -1.8 kJ/mol with an average correction of 0.2 kJ/mol. The maximum $\delta E_{\text{CCSDT(Q)}}$ correction is seen for the Cu^+ species bound to 1 H_2 whose reference interaction energy is 66.9 kJ/mol. The correction of 1.8 kJ/mol is only 2.8% of the total interaction energy. CCSD(T) shows systematic underbinding for all cases except a few ones like Cu salts and Zn^+ which show a positive value of $\delta E_{\text{CCSDT(Q)}}$. As shown in Fig.C.1 of the Chapter C, we can see that most of the $\delta \text{E}_{\text{CCSDT(Q)}}$ corrections are nearly zero, justifying the neglect of their contribution in fixing the final reference interaction energies.

Core-valence correlations contribute a non-negligible amount to the interaction energy. Just considering the salts subset, the core-valence correlations contribute 3.0 kJ/mol to the reference interaction energy. This quite large contribution of the core-valence correlation energy to the interaction energy was very surprising. In order to account for this effect, the size of the frozen core was made as small as computationally feasible. For the s-block ions and salts subsets, no core electrons were frozen, making the term $\delta E^{core}_{\text{MP2/TZ}}$ zero. For the organic subset, the effect of core-valence correlations was found to be very small (0.04 kJ/mol) and was therefore taken care of in an additive manner as in equation Eq. (5.3) using MP2 and a triple-zeta quality basis set. For the transition metals subset, a neon core was frozen and the effect of freezing it was estimated using MP2 at a triple-zeta quality basis set.

All the calculations were done on the ground electronic and spin state of the species. The s-block ions, salts, and organic categories consist only of closed-shell singlet species with unambiguous ground spin states. For the transition metal category, ground spin state was determined using CCSD(T)/TZ. In cases where experimental references were found,

Table 4.2: A table showing the different error metrics used in the various interaction energy regimes specifically catered to the H_2 binding problem. The error, ΔE, is defined as $\Delta E = |E - \hat{E}|/\tilde{E}$, where $\hat{E}$ is the reference interaction energy and $\tilde{E}$ is the regularizer.

	Energy range	Error metric	Error Formula	Regularizer $\tilde{E}$
Strong	–200 to –25	Absolute error	$\Delta E = \lvert E - \hat{E} \rvert$	–1
Favorable	–25 to –15	Relative error	$\Delta E = \lvert \left(\frac{E-\hat{E}}{\hat{E}}\right) \times 100 \rvert$	$\frac{\hat{E}}{100}$
Weak	–15 to 0	Regularized Relative error	$\Delta E = \lvert \left(\frac{E-\hat{E}}{-12}\right) \times 100 \rvert$	$\frac{-12}{100}$

the CCSD(T)/TZ ground spin state agreed with experiments. A table showing the spin state chosen and their experimental references can be found in Table C.2. Transition metal containing systems are often tricky to handle as they might exhibit multireference character. Spin-symmetry breaking at the Hartree-Fock level is often required for describing systems with multireference character and hence the existence of spin-symmetry breaking can be used as a diagnostic tool for it.[337, 338] Of the binding motifs chosen, the maximum deviation of $\langle S^2 \rangle$ deviation of 0.018 is exhibited by $Co^+(H_2)_4$. This deviation is negligible and hence these orbitals can reliably be used for correlation calculations.

DFT interaction energies were computed using a quadruple-zeta quality Karlsruhe basis, def2-QZVPPD[173] and a quadrature grid consisting of 99 Euler-MacLaurin radial points and 590 Lebedev angular points. SG-1 was used for integrating the non-local VV10 part in functionals that employ it.[175] There are two additional degrees of freedom for double hybrid density functionals: (1) employment of the density fitting approximation, (2) frozen core approximation. We have used these density functionals as they were trained originally. In cases where the authors did not specify how they were trained, we have used the parameters that give the least error and these details have been documented in Table C.1 The reference and density functional interaction energies were computed as the difference between the complex, the substrate, and H_2, as shown in Eq. (4.6)

$$\Delta E_{int} = E\big(\text{M–}(H_2)_n\big) - E\big(\text{M–}(H_2)_{n-1}\big) - E\big(H_2\big) \tag{4.6}$$

where ΔE_{int} is the interaction energy binding one of the H_2s to the substrate, $E\big(\text{M–}(H_2)_n\big)$ is the energy of the substrate bound to one or multiple H_2s, $E\big(\text{M–}(H_2)_{n-1}\big)$ is the energy of the substrate bound to $(n-1)$ H_2s at the M–$(H_2)_n$ geometry, and $E\big(H_2\big)$ is the energy of the H_2 at the M–$(H_2)_n$ geometry. This is commonly referred to as the "vertical" interaction energy as the geometries of the subsystems are not allowed to relax. Another way to compute interaction energy is to account for the geometric distortion of the substrate and H_2 upon complex formation. In this method, the relaxed geometry for the substrate and H_2 are used in (4.6). This method of computing interaction energy is called "adiabatic" interaction energy. In this work, we compute the interaction energy for each binding moeity using both the adiabatic and vertical methods.

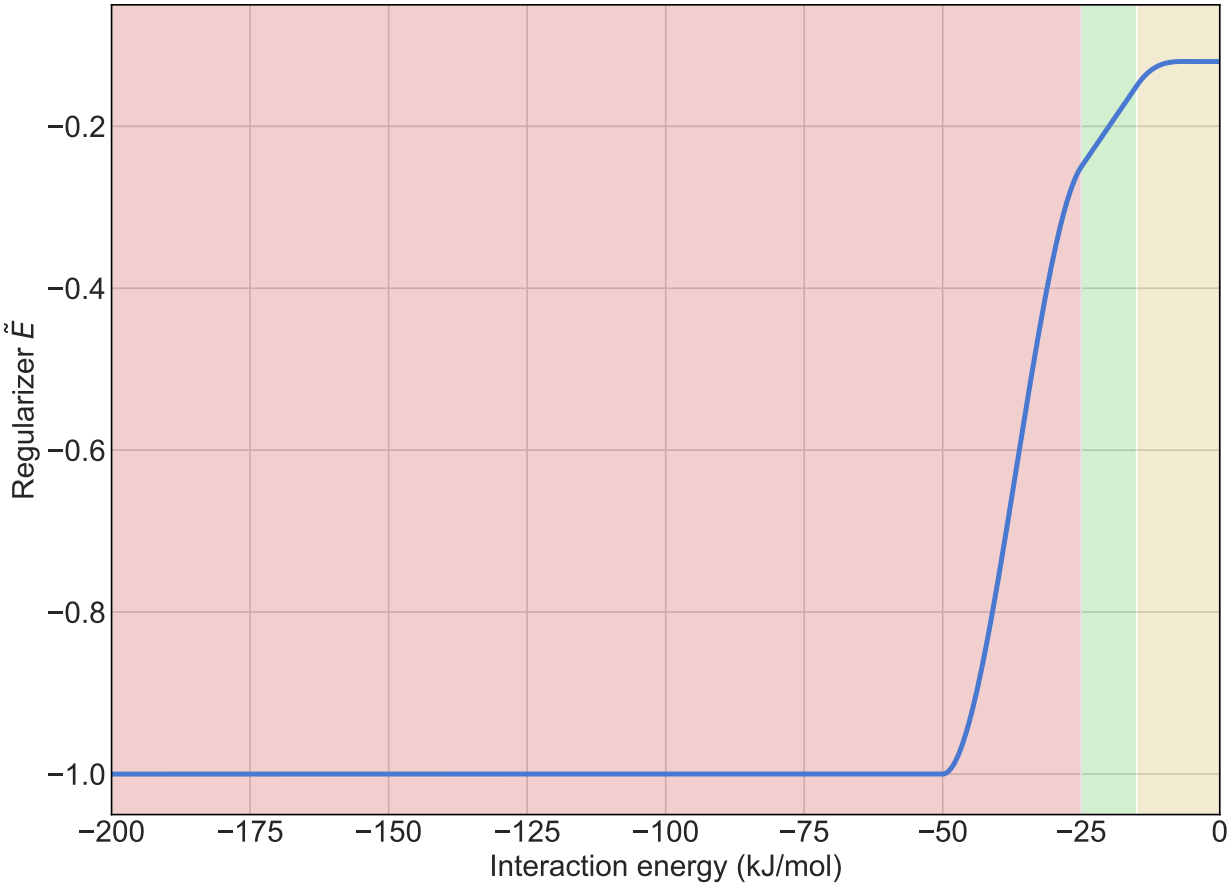


Figure 4.3: A figure showing the regularizer in the interaction energy regime of –200 to 0 kJ/mol.

The reference binding energies for the H2Bind275 dataset are spread over two orders of magnitude (*vide infra*). While all the species in the dataset inform us about the performance of density functionals for H_2 binding, not all of them give us the same amount of information. In H_2 storage applications, it is desired to have binding sites with an interaction energy of –15 kJ/mol to –25 kJ/mol for maximum uptake when operated between the pressures of 5 and 100 bar.[130, 339, 340] We have designed our error metric so as to give more weight to species in the interaction energy regime of –25 to –15 kJ/mol. The total interaction energy regime, and consequently the binding moieties, were divided into three categories: (1) strong, (2) favorable, and (c) weak. The energy ranges and error metrics used are summarized in Table. (4.2). The strong category consists of species with binding energies stronger than –25 kJ/mol. The DFT errors in this regime should have smaller weights and hence absolute error was used in this region making the actual magnitude of the errors small. This would imply a regularizer of $\tilde{E} = -1$ in this regime. The favorable regime consists of species

with an interaction energy between –25 kJ/mol and –15 kJ/mol. Large weights should be associated with errors in this region as they are of most importance from the H_2 storage perspective. Absolute percentage error is used in this region making the absolute magnitude of the errors relatively large. A regularizer of $\tilde{E} = \hat{E}/100$ was used for this range. Errors in the > -15 kJ/mol region are also important as these represent secondary H_2 binding sites like H_2 binding to organic linker molecules in MOFs. While these sites were not specifically designed to bind H_2, they serve as important structural units that keep the binding sites together. It is important to accurately model these interactions as they also contribute to the overall uptake of the material. However, small values of reference interaction energy in the denominator could cause the error to blow up and get numerically large weights. This was avoided by using a constant regularized reference of –12 kJ/mol, which implies $\tilde{E} = -12/100$. The comprehensive error metric, Regularized Mean Absolute Percentage Error (RegMAPE), smoothly interpolates between absolute error and percentage error while transitioning from strong to favorable regime and again smoothly interpolates between the percentage error and regularized percentage error while moving from the favorable to weak regime. The regularizer, $\tilde{E}$, is plotted in Fig. (4.3).

4.3 Results and Discussion

Reference Interaction Energies

The computed reference interaction energies range from –189.0 kJ/mol to –2.3 kJ/mol spanning more than two orders of magnitude. The distribution of the interaction energies can be seen in Fig. (5.1) which shows the distribution of the reference interaction energies by chemical categories of the dataset. s-block ions, consisting of 77 data points are concentrated in the –10 to –50 kJ/mol region. In contrast, the salts category has a much smaller range of –11 kJ/mol to –25 kJ/mol with 18 data points in the favorable regime. Furthermore, the organic category has an even smaller range with all the species lying between –2.3 kJ/mol and –5.6 kJ/mol. Transition metals, by far, show the most diversity in the interaction energies with values ranging from –5.3 kJ/mol to –189.0 kJ/mol. However, majority of the transition metal containing species have interaction energies of less than 100 kJ/mol. Of the total 275 data points, 50 of them lie in the favorable range for hydrogen storage and these points are given the most weightage by the error metric.

Performance of wavefunction methods

Before looking into the performance of different density functionals, we will briefly examine the performance of different wavefunction methods. Hartree Fock binding energies have a RegMAPE of 45% which highlights the importance of correlation energy for binding, as captured by coupled cluster theory. HF performs better than only the LDA density functionals. HF underbinds all the cases with a mean signed error of 19.6 kJ/mol. HF is

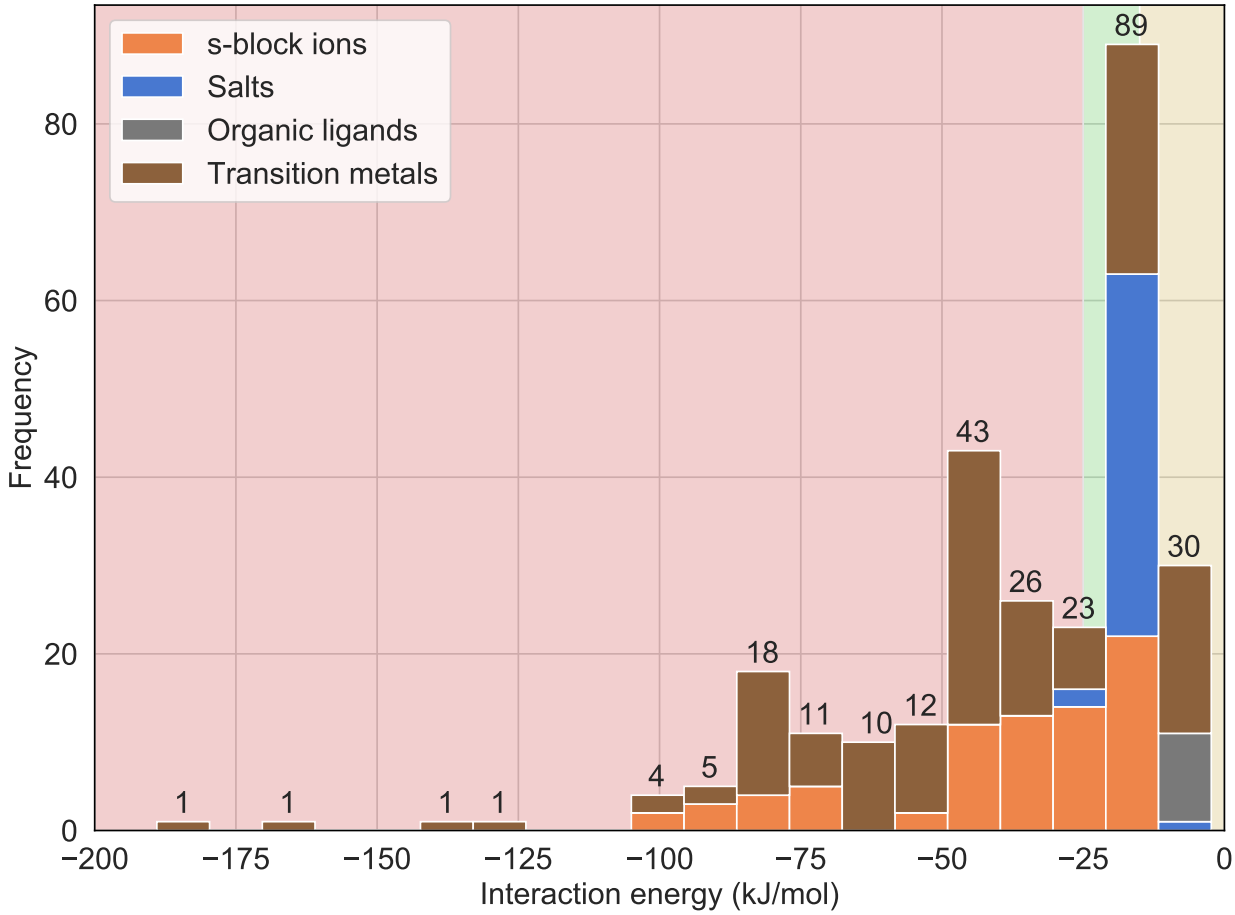


Figure 4.4: A figure showing the distribution of interaction energies for the whole H2Bind275 dataset.

unable to capture dispersion and predicts that all the dispersion-bound complexes like those in the organic ligands subset are unbound. MP2 at the quintuple-zeta quality basis set is able to capture most of the correlation binding energy and has a RegMAPE of only 5.9%. MP2 binding energies extrapolated to the complete basis set (CBS) limit perform slightly better with a RegMAPE of 5.2%. MP2 also slightly overbinds with a mean signed error of –1.8 kJ/mol which is consistent with the conventional wisdom in quantum chemistry that MP2 overestimates dispersion interactions.[341, 342]

Choice and overview of density functionals

While DFT is formally exact within the Born-Oppenheimer approximation, in practice, the exact form of the exchange-correlation functional remains unknown. As there is no

Table 4.3: A table showing all the density functional approximations benchmarked in this work classified by rung of the Jacob's ladder.

Rung #	Rung name	Functionals
Rung 5	Double Hybrid	ωB97M(2), DSD-PBEPBE-D3(BJ), XYG3, B2PLYP-D3(BJ), PTPSS-D3(0), PBE0-DH, XYGJOS
Rung 4	Hybrid	B3LYP, B3LYP-D3(0), PBE0, PBE0-D3(BJ), MN15, TPSSh, TPSSh-D3(BJ), MVSh, SCAN0, M06, M06-2X, M06-2X-D3(0), revM06, ωB97M-V, ωB97X-D, ωB97X-D3, ωB97X-V, M11, revM11, HSE-HJS, MN12-SX
Rung 3	meta-GGA	TPSS, TPSS-D3(BJ), revTPSS, SCAN, SCAN-D3(BJ), MS2, MS2-D3(op), B97M-V, B97M-rV, M06-L, MN15-L, mBEEF
Rung 2	GGA	PBE, PBE-D3(0), PBE-D3(op), RPBE, revPBE, revPBE-D3(op), BLYP, BLYP-D3(op), PW91, GAM, B97-D3(0), B97-D3(BJ)
Rung 1	LDA	SPW92, SVWN5

systematic recipe for improving DFT, various group all over the world have come up with multiple DFAs. As each of these approximations perform differently for various problems, in this section we characterize the performance of 55 DFAs for their ability to predict binding energies of H_2 to different binding substrates accurately. The 55 functionals were chosen based on different criteria. One of the criteria was their superior performance in non-covalent interaction energy databases. [18] Density functionals were also chosen to represent distinct functional families being developed by leading research groups. We have also included some functionals that are frequently used in the H_2 storage modeling community. While there are different methods to classify density functionals, Perdew's metaphorical Jacob's ladder is the most popular one.[20] As one moves up the Jacob's ladder from the Hartree world to the heaven of chemical accuracy, one can expect density functionals to get more and more accurate. However, this might not necessarily be true as increasing the degrees of freedom in density functionals as one moves up the Jacob's ladder can also result in overfitting of parameters and subsequently poor transferability. For further information about the functional forms of these density functionals, the number of parameters, and the datasets they were trained, we refer the reader to these comprehensive reviews of density functional development.[17, 18]

Table (4.3) shows all the functionals chosen for benchmarking in this work. Rung one functionals, SVWN5[22] and SPW92,[23] contain Slater exchange and different parameterizations for correlation energy and depend only on electron density ρ. The 12 Generalized Gradient Approximation (GGA) density functionals chosen, depend on density and the gradient of density (ρ and $\nabla\rho$), can be classified into 6 families. The PBE family of density functionals represented in this paper consist of the original PBE density functional,[24] PBE-D3(0),[37] RPBE[343] and their revised counterparts revPBE[344] and revPBE-D3(op).[345]

Other families represented include the BLYP family consisting of BLYP[25, 26] and BLYP-D3(op),[345] and the B97[346] family comprising B97-D3(0)[37] and B97-D3(BJ).[347] Other successful GGA functionals included are GAM,[348] PW91,[349] BP86-D3(BJ).[25, 347, 350] Rung 3 functionals are called meta-GGA functionals and they depend on the density, gradient of density, and the kinetic energy density (ρ, $\nabla\rho$, and τ). Meta-GGAs tested in this study consist of the non-empirically developed TPSS,[27] TPSS-D3(BJ),[347] and revTPSS.[351] The more recently developed SCAN[352] and its dispersion-corrected counterpart (SCAN-D3(BJ)[353]) also belong to this rung. B97M-V is a combinatorially optimized semi-local density functional with VV10 non-local correction developed by Mardirossian and Head-Gordon.[28] B97M-rV[354] has the rVV10 non-local correction[355] refit to the parent B97M-V functional and can be efficiently implemented in periodic codes. The mBEEF functional developed by Bligaard and co-workers is expected to give superior results for surface science and catalysis problems. [356] Other meta-GGA functionals included in this study are the Minnesota functionals developed by Truhlar and co-workers (M06-L[357] and MN15-L[358]), MS2,[359] and MS2-D3(op).[345] One of the biggest deficiencies of the semi-local density functionals (rung 1, 2, and 3) is self-interaction error which can be partially offset by adding some fraction of Hartree Fock exchange to the exchange-correlation functional.[360] Popular hybrid functionals are PBE0[361] and PBE0-D3(BJ)[347] from the PBE family and TPSSh[362] and TPSSh-D3(BJ)[347] from the TPSS family and SCAN0[363] from the SCAN family of density functionals. B3LYP[29] and B3LYP-D3(BJ)[347] are by far the most widely used density functionals today. Minnesota global hybrid functionals considered in this study are M06,[364] M06-2X,[364] M06-2X-D3(0),[37] revM06,[365] and MN15.[366] MVSh is another hybrid density functional developed by Perdew and co-workers evaluated for H_2 binding.[367] The aforementioned functionals, containing a fixed amount of exact exchange, only partially alleviate the problem of self-interaction error. A modified class of hybrid density functionals called range-separated hybrids present a more sophisticated approach to eliminating self-interaction error by treating exchange differently in long and short ranges.[368, 369] Short-range exchange is treated using both DFT exchange and exact exchange and only exact exchange is used in the long-range. ωB97X-D[258] is a reparameterization of the original ωB97X[370] functional to include atom-atom dispersion and ωB97X-D3[371] is a further reparameterization to include Grimme's D3 dispersion corrections.[37] ωB97X-V[372] and ωB97M-V[44] are range-separated hybrid GGA and meta-GGA functionals derived from the combinatorial approach. These functionals have been shown to perform well for a wide range of chemical problems previously.[18] M11[373] and its recently published revised version revM11[374] are other popular range-separated hybrids. Another strategy is to use Hartree Fock exchange in the short-range and DFT exchange in the long-range. Such density functionals are amenable to usage in solid-state calculations, and are called screened exchange functionals. Two of such functionals, HSE-HJS[375, 376] and MN12-SX,[377] are included in this study. The fifth rung of Jacob's ladder consists of double hybrid density functionals which have some fraction of correlation energy from second order Møller-Plesset Perturbation Theory perturbation theory. Addition of a fraction of correlation from perturbation theory can be justified based on Görling-Levy perturbation

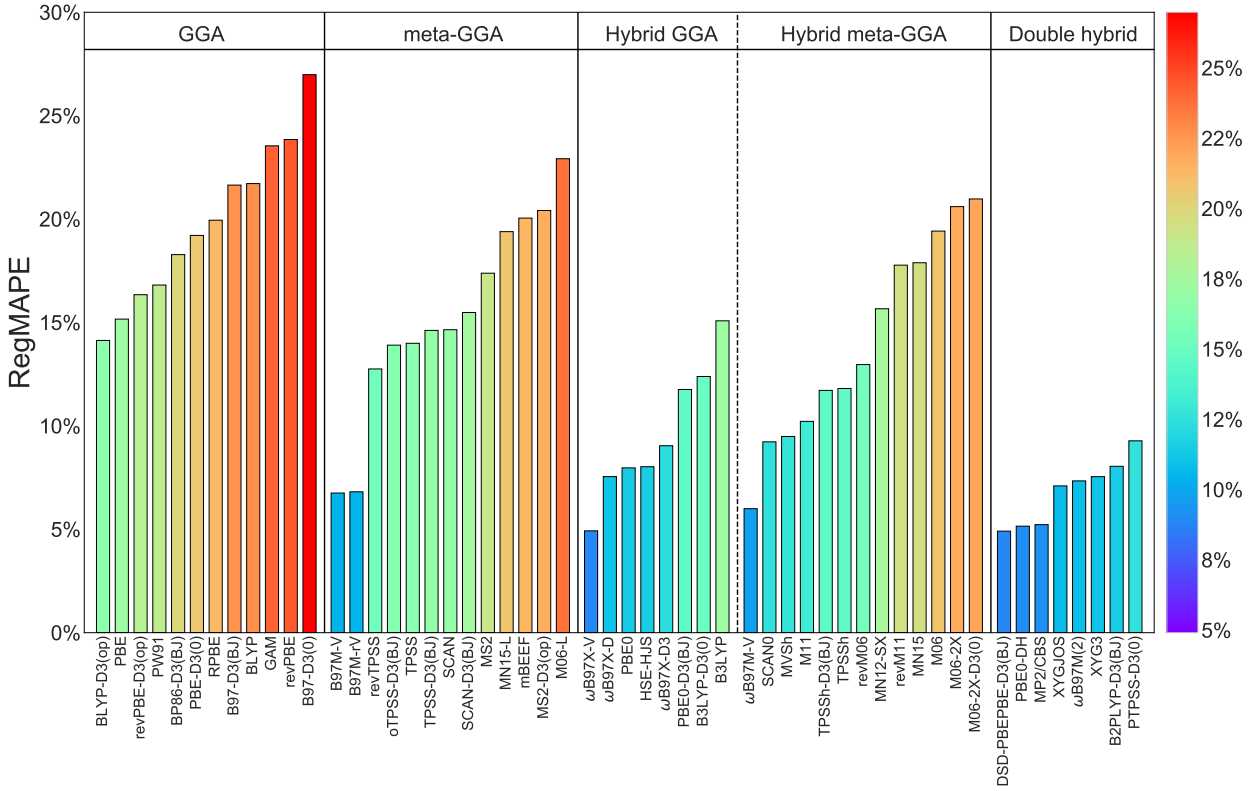


Figure 4.5: A figure showing the performance of the density functional approximations by rung of the Jacob's ladder. The LDA functionals, SPW92 and SVWN5, have an error of 60.06% and 60.08% respectively, and have not been shown in this figure. MP2 interaction energies extrapolated to the complete basis set limit has also been shown for comparison.

theory.[378] B2PLYP is the first density functional to use Kohn-Sham orbitals to compute a perturbation theory correction.[36] PBE0-DH is a non-empirical double-hybrid belonging to the PBE family. [379] XYG3 double hybrid density functional introduced a new class of double hybrid density functionals which uses orbitals from a successful lower-rung density functional (B3LYP orbitals in the case of XYG3) to perform a single-shot computation of the exchange-correlation and PT2 energies.[33] Other successful double hybrid density functionals following this approach are ωB97M(2)[34], the combinatorially-optimized double hybrid density functional using ωB97M-V orbitals and XYGJ-OS[380], the opposite-spin equivalent of XYG3. Other popular double hybrid density functionals like DSD-PBEPBE-D3(BJ)[381] and PTPSS-D3(0)[382] are also included for comparison.

Rank	s-block ions	Salts	Organic ligands	Transition metals	Total	DFA
1	1.5	1.3	2.0	4.1	3.1	PBE0-DH
2	1.4	0.7	0.2	4.5	3.3	ωB97X-V
3	1.0	0.8	0.3	5.1	3.7	ωB97M-V
4	0.8	1.0	0.2	5.2	3.8	DSD-PBEPBE-D3(BJ)
5	4.0	1.4	0.4	4.4	3.8	ωB97X-D
6	3.1	2.6	3.1	4.8	4.0	MVSh
7	2.3	2.2	2.7	5.3	4.1	PBE0
8	3.3	1.2	0.6	5.1	4.1	ωB97X-D3
9	2.4	2.0	2.6	5.4	4.2	HSE-HJS
10	1.7	2.9	0.9	5.5	4.2	XYG3
11	2.3	1.5	0.6	5.6	4.3	XYGJOS
12	0.9	1.0	0.4	6.4	4.6	PTPSS-D3(0)
13	1.7	2.3	0.4	6.4	4.8	ωB97M(2)
14	3.1	0.9	0.5	6.6	5.0	B2PLYP-D3(BJ)
15	2.2	4.4	4.5	6.4	5.2	TPSSh
16	3.7	1.1	0.4	6.6	5.2	PBE0-D3(BJ)
17	0.9	1.4	1.5	7.3	5.3	SCAN0
18	2.1	2.3	1.1	7.4	5.5	revM06
19	5.0	4.4	1.9	6.4	5.6	revM11
20	2.8	1.2	0.2	7.7	5.8	B3LYP-D3(0)
21	3.3	0.7	0.6	7.7	5.9	TPSSh-D3(BJ)
22	2.8	5.6	5.4	7.3	6.0	B3LYP
23	5.9	4.2	2.8	7.0	6.2	mBEEF
24	1.1	0.8	0.2	8.7	6.3	B97M-V
25	3.1	3.3	2.6	8.2	6.3	MN15
26	1.0	0.7	0.1	8.8	6.4	B97M-rV
27	2.0	3.4	3.9	8.6	6.5	revTPSS
28	2.8	2.2	1.1	8.9	6.6	oTPSS-D3(BJ)
29	2.5	4.8	4.6	8.6	6.7	TPSS
30	5.0	4.3	4.2	8.1	6.7	MN15-L
31	4.1	1.6	0.4	9.3	7.1	revPBE-D3(op)
32	3.3	8.7	5.7	8.8	7.5	RPBE
33	3.9	2.5	0.8	10.1	7.7	BLYP-D3(op)
34	3.8	0.8	0.6	10.4	7.8	TPSS-D3(BJ)
35	4.3	10.0	7.1	8.5	7.8	revPBE
36	2.0	2.6	0.5	11.1	8.1	M11
37	1.8	3.1	0.5	11.1	8.2	B97-D3(BJ)
38	3.9	8.5	7.2	9.7	8.2	BLYP
39	3.7	1.2	0.7	11.6	8.6	MN12-SX
40	3.1	2.8	0.6	12.2	9.1	M06
41	3.6	2.5	2.7	12.4	9.3	PBE
42	2.2	3.2	0.4	12.9	9.5	B97-D3(0)
43	1.8	2.1	1.9	13.3	9.8	BP86-D3(BJ)
44	4.6	1.1	1.0	13.5	10.1	PBE-D3(0)
45	3.5	2.4	0.8	13.8	10.2	M06-L
46	4.6	1.6	1.8	13.6	10.2	PW91
47	1.6	2.8	0.4	14.7	10.8	M06-2X
48	1.7	2.9	0.7	14.7	10.8	M06-2X-D3(0)
49	1.3	1.2	1.6	15.0	10.9	SCAN
50	2.1	1.2	0.3	15.6	11.3	GAM
51	1.4	1.6	0.5	15.6	11.3	SCAN-D3(BJ)
52	2.9	1.5	0.9	15.5	11.4	MS2
53	3.5	2.4	1.1	16.0	11.7	MS2-D3(op)
54	11.7	10.2	3.9	39.6	29.6	SPW92
55	11.7	10.2	3.9	39.6	29.6	SVWN5

Double hybrid | Hybrid | meta-GGA | GGA | LDA

Figure 4.6: A figure showing the Root Mean Square Errors for each category of the dataset as well as the total RMSE. The functionals are arranged in ascending order of the total RMSE. The best performing density functional in each rung has been highlighted.

Rank	s-block ions	Salts	Organic ligands	Transition metals	Total	DFA
1	1.8	4.4	1.4	7.0	4.9	DSD-PBEPBE-D3(BJ)
2	2.2	3.5	1.3	7.1	4.9	ωB97X-V
3	1.9	5.9	16.6	5.9	5.2	PBE0-DH
4	1.2	3.8	2.0	9.5	6.0	ωB97M-V
5	2.7	3.7	1.3	10.2	6.8	B97M-V
6	2.5	3.9	0.7	10.4	6.8	B97M-rV
7	5.2	8.8	4.6	7.7	7.1	XYGJOS
8	2.4	13.9	3.6	8.2	7.3	ωB97M(2)
9	2.6	16.8	7.4	7.4	7.5	XYG3
10	6.3	7.5	3.0	8.6	7.5	ωB97X-D
11	2.6	12.4	22.6	8.5	8.0	PBE0
12	2.7	11.4	21.6	8.9	8.0	HSE-HJS
13	5.6	4.2	3.7	10.8	8.0	B2PLYP-D3(BJ)
14	7.1	6.5	4.1	11.2	9.0	ωB97X-D3
15	2.1	8.1	12.1	13.2	9.2	SCAN0
16	2.7	5.6	3.1	14.3	9.3	PTPSS-D3(0)
17	9.8	13.5	26.1	6.9	9.5	MVSh
18	5.3	14.9	4.1	11.8	10.2	M11
19	6.2	3.4	4.4	17.7	11.7	TPSSh-D3(BJ)
20	7.8	6.5	2.8	16.1	11.8	PBE0-D3(BJ)
21	4.2	28.2	36.9	9.1	11.8	TPSSh
22	2.9	5.6	1.5	20.3	12.4	B3LYP-D3(0)
23	3.4	21.6	32.6	13.7	12.8	revTPSS
24	4.9	13.6	8.7	17.4	13.0	revM06
25	4.9	13.4	8.1	19.3	13.9	oTPSS-D3(BJ)
26	4.6	30.5	38.1	12.3	14.0	TPSS
27	5.4	13.1	6.7	19.7	14.1	BLYP-D3(op)
28	7.1	4.0	4.2	22.6	14.6	TPSS-D3(BJ)
29	3.7	7.2	13.5	22.9	14.7	SCAN
30	6.2	35.5	44.9	11.5	15.1	B3LYP
31	4.5	14.8	22.1	20.5	15.2	PBE
32	2.7	8.8	4.2	25.1	15.5	SCAN-D3(BJ)
33	6.4	6.6	4.4	24.2	15.7	MN12-SX
34	8.4	8.6	2.6	23.9	16.3	revPBE-D3(op)
35	7.3	8.6	14.2	24.6	16.8	PW91
36	5.7	8.9	7.4	26.9	17.4	MS2
37	12.8	28.2	15.7	17.4	17.8	revM11
38	7.8	20.2	21.1	22.4	17.9	MN15
39	5.1	13.0	15.3	26.9	18.3	BP86-D3(BJ)
40	8.1	6.3	8.3	29.9	19.2	PBE-D3(0)
41	11.9	24.0	34.9	20.9	19.4	MN15-L
42	6.4	15.3	3.7	28.7	19.4	M06
43	7.6	55.4	46.3	13.9	19.9	RPBE
44	17.3	27.5	23.3	19.0	20.0	mBEEF
45	8.2	14.8	9.0	29.4	20.4	MS2-D3(op)
46	3.2	16.5	2.9	32.4	20.6	M06-2X
47	3.4	17.5	4.9	32.5	21.0	M06-2X-D3(0)
48	2.7	18.6	3.3	34.0	21.6	B97-D3(BJ)
49	9.9	54.5	59.1	15.4	21.7	BLYP
50	6.1	13.3	6.0	36.0	22.9	M06-L
51	3.6	6.2	1.9	41.0	23.5	GAM
52	12.0	64.7	58.2	15.3	23.8	revPBE
53	6.2	16.1	3.0	43.1	27.0	B97-D3(0)
54	25.8	66.7	32.5	78.3	60.1	SPW92
55	25.8	66.7	32.4	78.3	60.1	SVWN5

Double hybrid | Hybrid | meta-GGA | GGA | LDA

Figure 4.7: A figure showing the RegMAPE for each category and the total RegMAPE of the dataset. The functionals are arranged in ascending order of the total RegMAPE. The best performing density functional in each rung has been highlighted.

Performance of density functional approximations

Fig. (4.5) shows that density functionals exhibit varied performance for H_2 binding applications. In general, the performance of the density functionals improves as we climb up Jacob's ladder. However, this rule has many notable exceptions. Using RegMAPE as the error metric, we see that DSD-PBEPBE-D3(BJ), a spin-component-scaled double hybrid density functional with empirical dispersion correction provides the best performance with a RegMAPE value of 4.9%. DSD-PBEPBE-D3(BJ) also has the fourth least RMSE of 3.8 kJ/mol and both the error metrics show that this density functional is the best performing for H_2 storage as shown in Fig. (4.6) and Fig. (5.2). The second best performing functional is the combinatorially optimized range-separated hybrid GGA, ωB97X-V, which shows a RegMAPE of 4.9%, which is comparable to that of DSD-PBEPBE-D3(BJ). ωB97X-V also yields the lowest RMSE of 3.3 kJ/mol. It is followed by the non-empirical general purpose double hybrid density functional DSD-PBEPBE-D3(BJ) which shows an error of 5.2%. The 4^{th} best performing density functional, ωB97M-V, has a gap of about 0.8% showing an error of 5.7%. ωB97M-V functional is the 3^{rd} best performing density functional by RMSE and was also the best performing hybrid functional on a main group chemistry database of nearly 5000 data points. [18] The B97M-V and the B97M-rV functionals closely follow with errors of 6.8%. This is especially interesting as B97M-V and B97M-rV are semi-local functionals which are much less expensive than hybrid functionals for both cluster and periodic computations.[70, 71, 223] B97M-rV uses the rVV10 non-local functional, which allows efficient evaluation in plane wave basis codes. [355] Although these two functionals have a large RMSE of 6.3 and 6.4 kJ/mol, their erroneous predictions of binding energies seem to be largely concentrated in the region outside the interesting regime of –15 to –25 kJ/mol. Other double hybrid density functionals providing providing competitive performance are the XYGJOS, ωB97M(2), and XYG3 functionals. ωB97X-D, PBE0, and HSE-HJS hybrid functionals also show low RegMAPEs of 7.5%, 8.0%, and 8.0% respectively. MP2 interaction energy extrapolated to the complete basis set limit shows a low RegMAPE of 5.2% and is comparable to the performance of the best density functionals. This is further interesting given that double hybrid functionals are as expensive as MP2.

It is interesting to note that commonly used density functionals like B3LYP and PBE show very poor performance (ranked 30 and 31 respectively). The dispersion-corrected version of B3LYP, B3LYP-D3(0), shows some slight improvement in performance while that of PBE, PBE-D3(0) shows worse performance, jumping down 9 places. Effect of addition of dispersion correction is discussed in detail in a later subsection. Even newly developed functionals like SCAN show disappointing performance. The screened-exchange density functional HSE-HJS, which has been suggested for use in solid-state calculations, also shows excellent performance with a RegMAPE of 8.0% and RMSE of 4.2 kJ/mol. While MVSh has a very low RMSE of 4.0 kJ/mol reflecting its good performance over the entire dataset weighted equally, its corresponding RegMAPE of 9.5% is not that impressive suggesting that it is not very suitable for usage in the interesting H_2 storage regime. XYG3, the first xDH density functional, shows comparable performance in terms of both RMSE and RegMAPE.

Performance by dataset category

While the goal of this work is to rank functionals based on their performance for the H_2 storage problem as represented by the entire H2Bind275 dataset, it would be useful to analyze their performances by chemical categories of the data in order to characterize the origin of errors. While the category errors are not designed to sum up to the total in Tables (4.6) and (5.2), they can be contrasted with each other in order to get a relative sense. Another error metric, the regularized maximum absolute percentage error, is shown in Table C.6

The errors from the transition metals category are the largest in both absolute (as indicated by RMSE) and relative error metrics (as indicated by RegMAPE). This is not surprising considering that the majority of the semi-empirical density functionals are trained on main-group molecular properties. The transferability of DFAs trained on main-group element properties to those of transition metals has been of significant interest in the DFA development community.[19] For the transition metal category, the PBE0-DH functional performs best with a RegMAPE of 5.9% and RMSE of 4.1 kJ/mol. MVSh performs the second best with a RegMAPE of 6.9% and is also the best performing hybrid functional. B97M-V is the best performing semi-local functional (as with the entire dataset).

The s-block ions category is second largest with 77 data points. This category, consisting of s-block monocations and dications, represent the ability of the density functionals to capture electrostatics and polarization interactions accurately. ωB97M-V outperforms all other functionals in this category with a RegMAPE of 1.2%. The second best performing functional is DSD-PBEPBE-D3(BJ) with a RegMAPE of 1.8% and this is closely followed by PBE0-DH. B97M-rV, a meta-GGA density functional, is the best performing semi-local functional with a RegMAPE of 2.5%.

The organic ligands category is the smallest (only 10 data points), and tests the ability of density functionals to describe van der Waal's interactions correctly. It is a little surprising to see that the local functional B97M-rV outperforms its hybrid and double hybrid counterparts. Evidently these functionals, with their VV10 non-local corrections, are well equipped to deal with such dispersion dominated interactions. The best performing hybrid is the ωB97X-V functional which appears 2^{nd} in the ranking. PBE0-DH consistently underbinds all the 10 data points with a MSE of 2.0 kJ/mol leading to an RMSE of 2.0 kJ/mol and a quite large RegMAPE of 16.6%. This underestimation of dispersion reflects the fraction of PT2 which is substantiated by poorer performance of its hybrid and semi-local counterparts, PBE0 and PBE, which exhibit high RegMAPEs of 22.6% and 22.1% respectively. Upon addition of dispersion corrections via the PBE0-D3(BJ) and PBE-D3(0), their errors decrease to a mere 2.8% and 8.3% respectively.

The performance of DFAs in the salts category is also somewhat surprising because of the larger magnitude of errors in comparison to the other main group categories as the majority of the DFAs are trained on main-group chemistry properties. TPSSh-D3(BJ) is the best performing for the salts category with RegMAPE of 3.4% and RMSE of 0.7 kJ/mol. B97M-V is the best semi-local functional with a RegMAPE of 3.7% and RMSE of 0.8 kJ/mol. In this category, the best hybrid and semi-local functionals outperform the best double hybrid

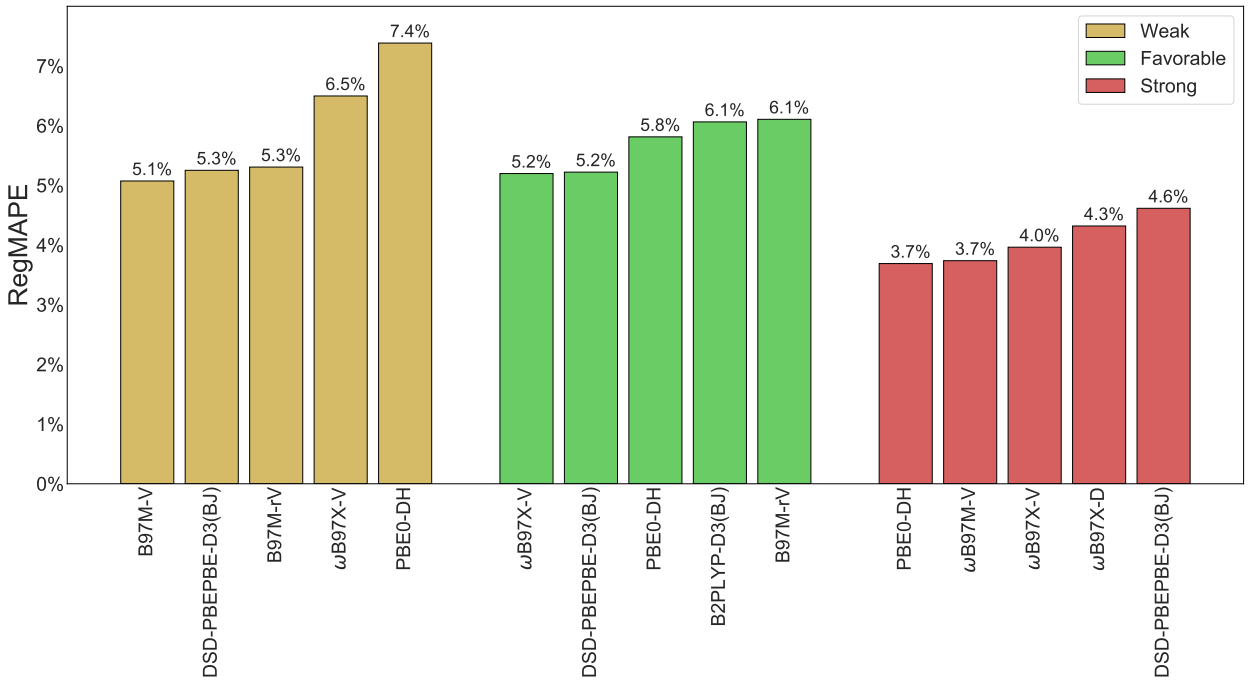


Figure 4.8: A plot showing the performance of the top five density functional approximations in the three interaction energy ranges relevant to H_2 binding applications: (a) Weak (less than –15 kJ/mol) (b) Favorable (–15 to –25 kJ/mol) (c) Strong (larger than –25 kJ/mol)

functional B2PLYP-D3(BJ).

Performance by interaction energy range

Hydrogen storage applications usually target materials that can bind H_2 with an interaction energy of –15 to –25 kJ/mol. As a consequence, binding moeities with interaction energies outside this range are less important but cannot be disregarded altogether as they might appear in modified forms or as secondary binding sites in storage materials. Our regularized error metric, RegMAPE is designed to weigh interactions in the favorable regime more than others. Semi-local functional like B97M-V outperforms others in the low interaction energy regime. The second best functional for this region is a double hybrid with D3(BJ) dispersion correction, DSD-PBEPBE-D3(BJ), with a RegMAPE of 5.3%. The best hybrid functional in this domain, ωB97X-V, has a RegMAPE of 6.5%. In the favorable regime where the data points have the most weight, the performance is reminiscent of the performance over the whole dataset with a few notable exceptions. ωB97X-V, DSD-PBEPBE-D3(BJ), and PBE0-DH appearing in the 1^{st}, 2^{nd}, and 3^{rd} spots also appear in the top ten density functional list in the overall performance. This further illustrates that RegMAPE is performing as it was expected to - that is, it gives higher weights to data points within the favorable interaction

Table 4.4: A table showing the best performing density functionals in each rung of the Jacob's ladder along with their overall ranking, RegMAPE, and RMSE.

Rung	Rung name	Functional	Rank	RegMAPE (%)	RMSE (kJ/mol)
1	LDA	SPW92	54	60.1	29.6
2	GGA	BLYP-D3(op)	27	14.1	7.7
3	meta-GGA	B97M-V	5	6.8	6.3
4	Hybrid	ωB97X-V	2	4.9	3.3
5	Double hybrid	DSD-PBEPBE-D3(BJ)	1	4.9	3.8

energy regime. One notable exception is the double hybrid B2PLYP-D3(BJ) functional which is ranked 4$^{\text{th}}$ in the favorable regime with a RegMAPE of 6.1% but does not even appear in the top 10 density functional list due to large errors in the strong interaction energy regime. Fig. (4.8) shows smaller RegMAPE values for the strong regime as absolute errors are used. PBE0-DH gives the best predictions in this regime with a RegMAPE of 3.7% and is closely followed by ωB97M-V and then by other range-separated hybrids like ωB97X-V and ωB97X-D. It is also interesting to note the poor performance of semi-local functionals. The best performing semi-local functional is MN15-L with a RegMAPE of 7.7%. Even functionals like B97M-V and B97M-rV, which provide very good performance in other categories, fail in this regime.

Performance by rung of Jacob's ladder

Density functionals can be classified on the basis of Jacob's ladder as was summarized in Table (4.3). It is generally expected that the density functionals perform better as one climbs the rungs of the Jacob's ladder. Rung 1 of the Jacob's ladder, consisting of density functionals that depend only on the density ρ, perform the worst among all the ones tested with a RegMAPE of 60.1%. The performance increases significantly upon moving from Rung 1 to Rung 2 with BLYP-D3(op) showing an error of 14.1% (RMSE of 7.7 kJ/mol). Other notable GGA functionals that show comparable performance are PBE and the revPBE-D3(op) functional. The revPBE-D3(op) also gives the lowest RMSE among the GGA functionals.

Upon moving up another rung from GGAs to meta-GGAs, the density functionals depend on the kinetic energy density (τ) in addition to the density and its gradient (ρ and $\nabla\rho$). The best meta-GGA functional, B97M-V, is also one of the best performing functionals of this work with a RegMAPE of 6.8%. This represents a significant improvement over the best GGA functional with a RegMAPE of 14.1%. However, in terms of RMSE, the root mean square error of B97M-V is 6.3 kJ/mol. The performance of B97M-rV functional closely follows that of B97M-V, but the performance of meta-GGAs considerably worsens after this with the third best meta-GGA showing a RegMAPE of 12.8%, which is almost double that of B97M-V and B97M-rV. SCAN and its dispersion-corrected version, SCAN-D3(BJ), give larger errors of 14.7% and 15.5% respectively despite being recommended as one of

the best performing meta-GGA density functionals in a comprehensive DFA performance assessment.[17] The mBEEF functionals, which is expected to give good performance for surface science and catalysis, performs very poorly with a RegMAPE of 20.0%, though it has a comparatively lower RMSE of 6.2 kJ/mol. While the best performing meta-GGA beats the best performing GGA functional, other meta-GGAs like SCAN, MS2, and MN15-L show poorer performance than the best GGA, BLYP-D3(op).

Hybrid functionals perform very well with ωB97X-V, showing the best performance with a RegMAPE of 4.9% (RMSE of 3.3 kJ/mol). It is closely followed by its hybrid meta-GGA counterpart, ωB97M-V, which shows a RegMAPE of 6.0% (RMSE of 3.7 kJ/mol). The third best performing mega-GGA functional is the long-range corrected ωB97X-D functional (RegMAPE of 7.5% and RMSE of 3.8 kJ/mol). Out of the top five hybrid functionals, four of them are range-separated, indicating their success relative to global hybrids having a constant fraction of exact exchange. The best performing Minnesota functional is M11 with a RegMAPE of 10.2% (RMSE of 8.1 kJ/mol). Other Minnesota hybrid meta-GGAs like revM11, MN15, M06, M06-2X, and M06-2X-D3(0) tested provide poor performance for H_2 binding energies.

Double hybrid functionals, with some fraction of MP2 correlation energy, consistently perform the best with five of the seven functionals tested appearing in the top 10 list of the best density functional approximations. Leading the pack is the DSD-PBEPBE-D3(BJ) functional with a RegMAPE of 4.9% and RMSE of 3.8 kJ/mol. This density functional also happens to be the best performing among the 55 functionals tested in this study. This is closely followed by the PBE0-DH functional. The combinatorially optimized ωB97M(2) double hybrid does not outperform other functionals in rung 5 unlike its meta-GGA and hybrid counterparts. Even the worst performing double hybrid, PTPSS-D3(0), only has a RegMAPE of 9.3% (RMSE of 4.6 kJ/mol). Double hybrids, which contain some fraction of MP2 correlation energy, reflect the good performance of MP2 itself (RegMAPE of 5.9%).

Performance upon addition of HF exchange

Table S3 and Fig. (4.9) show the performance of density functional approximations belonging to the same family with and without exact exchange in order to quantify the effect of adding exact exchange to semi-local functionals. Addition of Hartree Fock exchange, also known as exact exchange, has been known to partially alleviate the problem of self-interaction error in density functionals. However, computation of exact exchange is the bottleneck in the computation of the fock matrix for hybrid density functionals. In all the six families of densities functionals shown in Fig. (4.9), we see that addition of exact exchange improves the performance of the DFAs. All the hybrid density functionals considered in Fig. (4.9) are global hybrid functionals, meaning they contain a constant fraction of exact exchange for all inter-electronic distances. This is in contrast to range-separated hybrids which vary the amount of exact exchange with inter-electronic distance. The TPSSh functional contains a small amount of exact exchange (10%) and subsequently shows only a small improvement with RegMAPE going down by 2 percentage points. There is a significant improvement

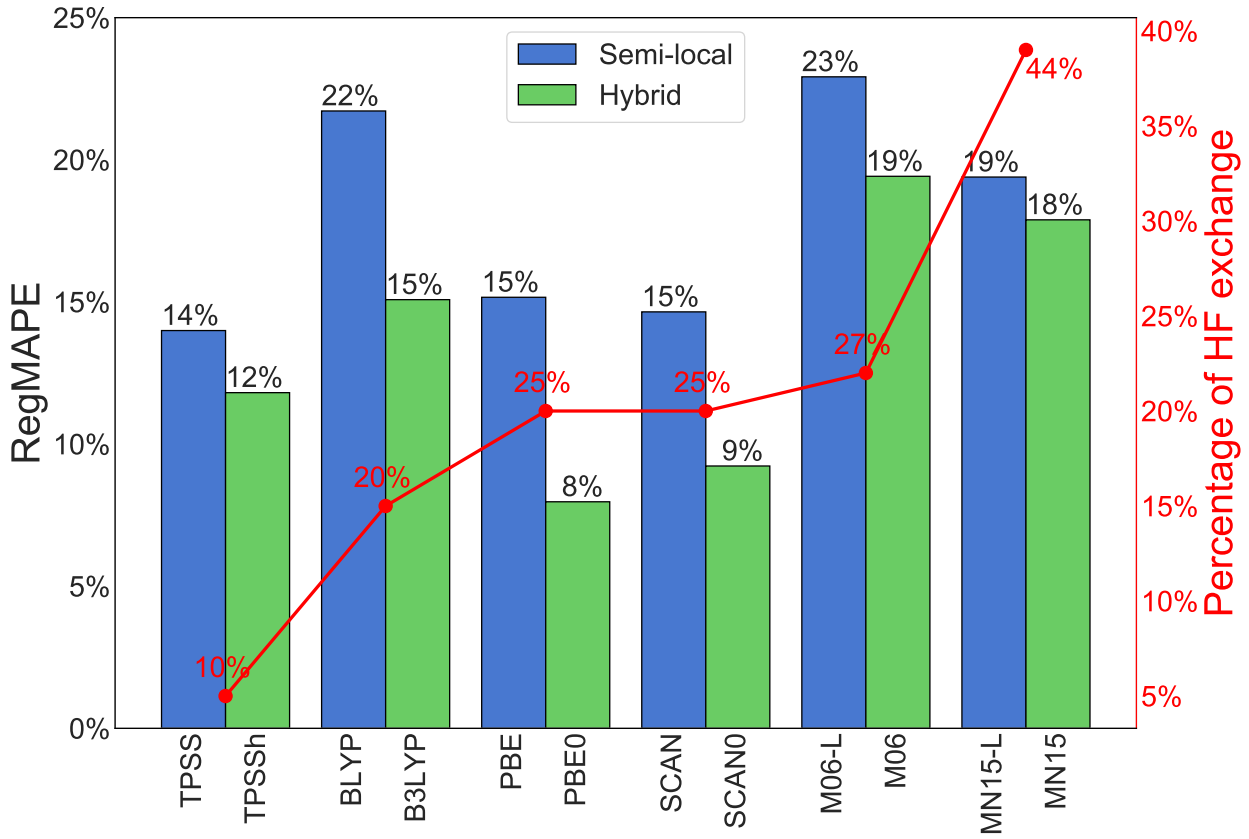


Figure 4.9: A graph showing the performance of density functional approximations with and without exact exchange. The right hand y-axis shows the percentage of HF exachange contained in each of the hybrid functionals.

in the performance of BLYP, PBE, and SCAN functionals upon addition of a considerable amount of HF exchange. For example, the RegMAPE of PBE (15.2%) goes down by a factor of two upon addition of 25% exact exchange to make PBE0. However, in the case of the two Minnesota functionals examined, M06 family and MN15 family, we see that even addition of a large fraction of exact exchange does not improve the performance considerably. The M06-2X functional, which contains twice the amount of exact exchange (54%) in comparison to the M06 functional, also does not show significant improvement in H_2 binding energies with a RegMAPE of 20.6%. The additional computational cost in computing exact exchange is probably not justified for the minute improvement in the performance in the case of the Minnesota functionals. Range-separation represents a more successful strategy for combining HF exchange in the long range with DFT exchange in the short range. While the best global hybrid functional PBE0 ranks 11^{th}, range-separated hybrids take the 2^{nd}, 4^{th}, and 10^{th} spots. Another possible technique for range-separation is the screened exchange method,

Table 4.5: A table showing the Mean Signed Error(MSE), Root Mean Squared Error (RMSE), RegMAPE, and rank of the density functionals with and without dispersion correction.

No Dispersion correction					With Dispersion correction				
Functional	MSE	RMSE	RegMAPE	Rank	Functional	MSE	RMSE	RegMAPE	Rank
PBE	-4.9	9.3	15.2	31	PBE-D3(0)	-7.0	10.1	19.2	40
revPBE	4.3	7.8	23.8	52	revPBE-D3(op)	-3.6	7.1	16.3	34
BLYP	2.4	8.2	21.7	49	BLYP-D3(op)	-2.4	7.7	14.1	27
TPSS	-1.3	6.7	14.0	26	TPSS-D3(BJ)	-4.7	7.8	14.6	28
SCAN	-6.0	10.9	14.7	29	SCAN-D3(BJ)	-6.7	11.3	15.5	32
MS2	-7.3	11.4	17.4	36	MS2-D3(op)	-8.1	11.7	20.4	45
B3LYP	2.4	6.0	15.1	30	B3LYP-D3(0)	-1.1	5.8	12.4	22
PBE0	-1.2	4.1	8.0	11	PBE0-D3(BJ)	-3.6	5.2	11.8	20
TPSSh	-0.1	5.2	11.8	21	TPSSh-D3(BJ)	-3.3	5.9	11.7	19
M06-2X	-1.9	10.8	20.6	46	M06-2X-D3(0)	-2.0	10.8	21.0	47

which uses HF exchange in the short range and DFT exchange in the long range. HSE-HJS and MN12-SX are screened exchange density functionals tested in this study. While there are no counterparts to compare against, we can see that the HSE-HJS functional performs reasonably well with a RegMAPE of 8.0%. On the other hand, the MN12-SX functional shows very poor performance.

Effect of dispersion corrections

In the absence of a strong electric field, dispersion is an important mechanism of interaction between H_2 and the binding site. These types of dispersion-dominated interactions are important in H_2-organic linker interactions represented by the organic ligands category of H2Bind275. The inherent semi-local parameterization of DFAs make it difficult for them to capture long-range dispersion effects without additional corrections. Several methods have been introduced to calculate the effect of dispersion self-consistently and as a single-shot computation. Of the self-consistent methods introduced, the vdW-DF method (BEEF-vdW and optB88-vdW), the VV10 and rVV10 non-local correlation functional (ωB97M-V, ωB97X-V, B97M-V, and B97M-rV) have been used in different DFAs in this paper. The DFAs which have the VV10 correction for dispersion interactions are, in fact, the most successful ones as they incorporate the correct interaction physics via VV10 parameters that are set consistently with all other parameters related to semi-local XC and exact exchange to avoid double counting. Another inexpensive approach to incorporating the effect of dispersion is using Grimme's empirical DFT-D suite of methods which is a damped atom-atom C_6 potential. [35, 37, 383] Density functionals containing the original DFT-D3 scheme with the CHG-style damping function[258] are suffixed by D3(0). The DFT-D3 scheme, combined with the damping function of Becke and Johnson are suffixed by D3(BJ).[347] Witte *et al.*

further generalized the Becke and Johnson damping function by optimizing the exponent of damping and this combination is termed D3(op).[345]

In this section, we analyze the effect of addition of DFT-D type dispersion correction on the H_2 interaction energies. Dispersion corrections are always negative and will only make binding energies stronger. The effect of addition of dispersion to 10 DFAs is shown in Fig. S2 and Table (4.5). The addition of dispersion improves the performance of some DFAs while it worsens the performance of other DFAs. Addition of empirical dispersion to PBE, TPSS, SCAN, MS2, PBE0, and M06-2X makes their performance worse. In these cases, the parent density functional is already overbinding H_2(s) without dispersion corrections as shown by their negative Mean Signed Errors (MSE) in Table (4.5) and it is hence not very surprising to note that the dispersion-corrected results worsen. PBE, SCAN, and MS2 density functionals severely overbind with MSEs of –4.9, –6.0, and –7.3 kJ/mol. Addition of dispersion increases overbinding showing MSEs of –7.0, –6.7, and –8.1 kJ/mol. Addition of dispersion to PBE0 actually takes it from a top performing functional (ranked 11^{th}) to a mediocre one (ranked 20^{th}). Of the ten DFAs and their corresponding dispersion tails investigated in this section, MS2 is the most overbinding with a MSE of –7.3 kJ/mol, which only worsens slightly to –8.1 kJ/mol upon addition of dispersion correction. The correction is highly damped. A similar phenomenon is seen for M06-2X whose MSE increases only slightly from –1.9 to –2.0 kJ/mol. TPSSh is the only functional of the ten investigated that has neither an underbinding nor an overbinding problem with a MSE of –0.1 kJ/mol. Though addition of dispersion increases the MSE from –0.1 to –3.3 kJ/mol and RMSE from 5.2 to 5.9 kJ/mol, the corresponding RegMAPE changes slighly from 11.8% to 11.7% indicating that the dispersion corrections affects the interesting regime of H_2 binding only slightly.

Empirical dispersion corrections actually improve H_2 binding energy estimates for the revPBE, BLYP, and B3LYP functionals. Dispersion corrections enhance the performance of B3LYP from a RegMAPE of 15.1% to 12.4%. The parent density functional systematically underbinds with a MSE of 2.4 kJ/mol. Upon addition of dispersion, the B3LYP-D3(0) functional systematically overbinds with a MSE of –1.1 kJ/mol. Addition of D3(op) correction to revPBE improves its performance from one of the worst performing functionals (ranked 52^{nd}) to a mediocre performing functional (revPBE-D3(op) is ranked 34^{th}). While the improvement in RMSE from 7.8 kJ/mol to 7.1 kJ/mol is not large, most improvement occurs in the interesting regime for H_2 binding, with RegMAPE improving significantly from 23.8% to 16.3%. Similarly, the dispersion-corrected BLYP functional, BLYP-D3(op), is the best performing GGA functional (ranked 26^{th}) while its uncorrected counterpart is ranked 49^{th}. Again, the improvement is concentrated in the interesting regime for H_2 storage materials.

In this section, we briefly discuss the performance of density functional approximations containing the vdW-DF non-local correlation functional. This series of non-local functionals were initially developed to study layered materials, but has ever since been used to study a wide range of materials with dispersion interactions. The vdW-DF-04 non-local correlation functional[40] paired with optB88 exchange functional and LDA correlation functional (optB88-vdW)[384] has been widely used in modeling the hydrogen storage properties of different materials.[385–388] However, the vdW-DF functional form has been defined only in

Table 4.6: A table showing the performance of density functionals containing the non-local correlation vDW-DF-04 density functional of Lundqvist and Langreth relative to the performance of the best performing functionals for the spin unpolarized subset of the H2Bind275 dataset. OptB88-vDW and BEEF-vDW, density functionals containing vDW-DF-04 non-local correlation, exhibit poor performance.

Functional	RegMAPE	RMSE	Rank
ωB97M-V	2.7	3.1	1
ωB97X-V	2.8	2.6	2
DSD-PBEPBE-D3(BJ)	2.9	2.0	3
B97M-rV	3.6	4.1	4
B97M-V	3.6	4.1	5
optB88-vDW	9.7	5.9	37
BEEF-vDW	15.4	12.7	48

the spin unpolarized form severely limiting its applicability. We have limited our analysis to two vdW-DF containing functionals: BEEF-vdW, a density functional within the bayesian error estimation framework, and optB88-vdW, containing the B88 exchange functional with parameters optimized to best reproduce the interaction energies of the S22 dataset. We find that both these functionals show poor performance in terms of all error metrics used. The optB88-vdW and BEEF-vdW functionals show a RegMAPE of 9.7% and 15.4% ranking 37^{th} and 48^{th} respectively. The magnitude of this error should be contrasted to ωB97M-V, the best performing DFA for this spin unpolarized subset which gives a RegMAPE of 2.7%. This subset also contains the easier portion of the dataset as the errors for this subset are much smaller than those for the entire dataset. For example, the ωB97M-V functional gives an error of 6.0% for the entire dataset but gives only an error of 2.7% for this subset. This further exacerbates the failure of the vdW-DF functionals to appropriately capture the physics of interaction between H_2 and binding moieties. Both DFAs show systematic underbinding suggesting that dispersion interactions have not been captured completely. We also find that the large magnitude of error does not originate from a couple of outliers, but from their systematic inability to describe interaction physics correctly in a wide range of chemical species.

Top five best performing density functionals

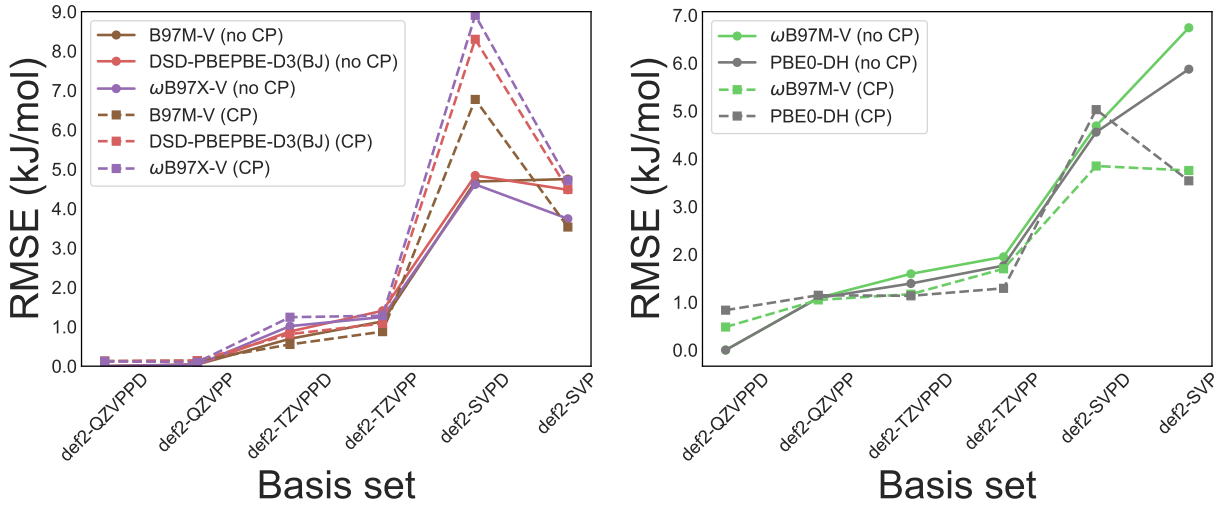
In this section, we will attempt to understand the sources of error in the top five best performing density functionals in the entire H2Bind275 dataset: DSD-PBEPBE-D3(BJ), ωB97X-V, PBE0-DH, ωB97M-V, and B97M-V. Having identified these functionals for overall good performance for H_2 binding applications, it will be very informative to understand the origin of errors in these functionals and hence the shortcomings of currently available density functional approximations. A systematic understanding of the failure of the best

density functionals could provide new avenues for development of better density functional approximations for H_2 storage and other applications. The top five functionals chosen also cover rungs 3–5 of Jacob's ladder with two double hybrids, two hybrids, and one semi-local meta-GGA. Each of these categories is also representative of the different costs associated with evaluating the density functionals with double hybrids being the most expensive, followed by hybrids, and lastly by semi-local functionals. Reduction of the computational cost of hybrid functionals to that of semi-local functionals is an active area of research. [96, 97, 132]

The largest contributors of error for the DSD-PBEPBE-D3(BJ) functional are the Ti^+ and Sc^+ species. Other species like Zn^+, Cr^+, and $CaCl_2$ are also leading sources of error. For the ωB97X-V functional, most errors come from transition metal species with predictions incorrect by about 46% for some Ti^+ species. Zn^+ and Fe^+ are some other species exhibiting large errors from 10% to 25%. The PBE0-DH also fails to predict accurate interaction energies for the Ti^+ species making errors of up to 56%. Zn^+ and Fe^+ are some difficult species for the the PBE0-DH functional. PBE0-DH also fails to predict the interaction energies with organic species accurately because of the lack of dispersion corrections as noted earlier. It is rather unsatisfying to see the failure of PBE0-DH for organic species, which are supposed to be easy problems for density functional approximations. However, a simple addition of Grimme-type dispersion correction cannot fix this issue as PBE0-DH systematically overbinds (MSE of –0.7 kJ/mol). ωB97M-V is the 4^{th} best performing density functional, and its major contributors of error are Ti^+, Zn^+, and Fe^+ species. B97M-V also incorrectly predicts the interaction energies of Ti^+ species by about 66%. There are a few common denominators among the major contributors of error for the top five best performing functionals. These species mostly belong to the transition metal category. This is rather expected as majority of them are trained on main group chemistry properties. While some density functionals like MN15 have been trained on transition metal properties,[366] it is interesting to see density functionals like ωB97M-V and B97M-rV, which were trained on main-group chemistry properties providing comparable performance to that of MN15 for transition metal properties.[19]

Semi-local and hybrid DFT calculations scale as the third power in the size of the one-particle basis set. Double hybrids, with the density fitting approximation for the MP2 piece, scale as the fifth power of the size of the basis set. While this scaling is much more favorable than that of CCSD(T), it limits the applicability of DFAs to systems of larger sizes. In the context of using DFAs to screen potential H_2 adsorption materials, it can limit the number of screenings possible in a given period of time. All interaction energies computations using finite basis sets suffer from two distinct kinds of error: Basis Set Superposition Error (BSSE) and Basis Set Incompleteness Error (BSIE). In finite basis set computations, these two errors may partially cancel. Basis set superposition error can be reduced by the counterpoise correction.[389]

Fig. (4.10) shows the errors in finite basis set interaction energy calculations for the top five DFAs in this study. All calculations were performed only on the vertical interaction energy subset as performing basis set superposition error corrections to adiabatic interaction



(a) (b)

Figure 4.10: (a) A figure showing the errors in the counterpoise corrected and uncorrected interaction energies for ωB97X-V, ωB97M-V, and B97M-V with counterpoise corrected def2-QZVPPD interaction energies as the reference. (b) A figure showing the errors in the counterpoise corrected and uncorrected interaction energies for DSD-PBEPBE-D3(BJ), and PBE0-DH with counterpoise corrected def2-QZVPPD interaction energies as the reference.

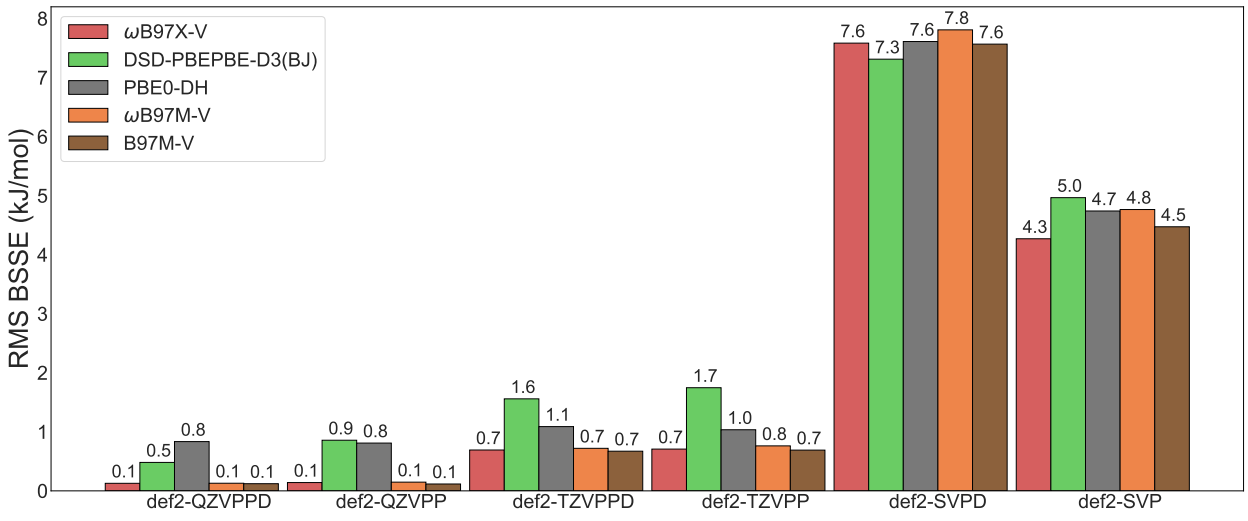


Figure 4.11: A figure showing the basis set superposition error for the top five best performing DFAs for different Karlsruhe basis sets.

energies is not straightforward. All the errors are computed with respect to the counterpoise corrected def2-QZVPPD interaction energy for the respective DFA so as to isolate the basis set errors from the DFA errors. BSSE decreases with increase in the basis set size for all the DFAs except for def2-SVPD/def2-SVP. In the case of the small def2-SVPD basis set, monomers can hugely benefit by borrowing the diffuse functions belonging to the other monomers. While this is possible even in the def2-TZVPPD and def2-QZVPPD case, the monomer's own basis sets are large enough that they do not gain much from borrowing a neighbor's diffuse function. Note the large magnitude of BSSE for the double hybrid density functionals in triple and quadruple-zeta basis sets in comparison to other DFAs. Even in the large quadruple-zeta def2-QZVPPD basis set, the double hybrids DSD-PBEPBE-D3(BJ) and PBE0-DH have a BSSE of 0.5 and 0.8 kJ/mol, while the BSSE of non double hybrid functionals is limited to 0.1 kJ/mol. This can be attributed to the fact that MP2 component in the double hybrids converges slowly with the basis set size.

Upon reducing the size of the basis set from def2-QZVPPD, both BSSE and BSIE cause overestimation and underestimation of interaction energies respectively. In the def2-QZVPP basis set, the counterpoise corrected and counterpoise uncorrected interaction energies show performance comparable to that of def2-QZVPPD with counterpoise corrected energies being slightly better. In the def2-TZVPPD basis set, the counterpoise uncorrected interaction energies perform slightly better. Counterpoise uncorrected interaction energies in the def2-TZVPP basis are the best compromise between accuracy and cost with errors of about 1 kJ/mol for the semi-local and hybrid functionals and 1.5 kJ/mol for the double hybrids. The opposite trend is seen in the def2-SVPD basis set with counterpoise corrected energies being better than then their uncorrected counterparts. The error cancellation is much more effective for double hybrids in this small basis set than the semi-local or hybrid functionals in which the BSSE dominates, leading to a systematic underbinding in the counterpoise uncorrected case. The error cancellation is even more effective in the def2-SVP case with with RMSE of 3.5 to 4.5 kJ/mol.

4.4 Conclusions

In the search for improved H_2 storage materials, density functional approximations could potentially be used to screen and simulate adsorption frameworks. In this work, we have created a dataset of 275 different chemical moeities, binding one or multiple H_2s to capture the different physical and chemical modes of hydrogen activation in various adsorbent paradigms. We have compiled highly accurate reference interaction energies for the dataset using coupled cluster theory with singles, doubles, and perturbative triples with the focal point analysis scheme. We have assessed the performance of 55 density functional approximations from all rungs of the Jacob's ladder using an error metric specifically designed to give larger weight to interaction energies in the range of –15 to –25 kJ/mol. We have examined the effect of exact exchange and empirical dispersion corrections on the performance of density functionals. We have identified problematic cases for density functionals, and

have recommended efficient techniques to predict hydrogen interaction binding energies.

Of the 55 DFAs assessed in this study, we have identified five density functionals for providing the best performance. These five functionals are ωB97X-V and ωB97M-V in the hybrids category, DSD-PBEPBE-D3(BJ) and PBE0-DH in the double hybrids category, and B97M-V in the semi-local meta-GGA category. DSD-PBEPBE-D3(BJ) performs the best over the 275 data points with an error of 4.9%. B97M-V is the best performing semi-local functional with a error of 6.8%. The B97M-rV functional also provides equally competitive performance, and with the rVV10 modification, it can also be efficiently used in periodic codes. Modern density functional approximations fail to accurately predict interaction energies of transition metal containing systems causing the biggest errors in that category. The best performing DFA at each rung of the Jacob's ladder surpasses the best DFA from the previous rung. We have also identified that the addition of the exact exchange usually helps with the performance of the density functionals. However, there is no systematic improvement in performance with the fraction of exact exchange present in DFAs. Addition of empirical dispersion correction can immensely benefit systematically underbinding DFAs like revPBE, BLYP, and B3LYP. However, empirical dispersion corrections should be utilized with caution as they will only worsen the performance if the parent DFA is already overbinding. We have also illustrated the interplay between the two major sources of error associated with using finite basis sets: basis set superposition error and basis set incompleteness error. Taking advantage of the error cancellation between these two sources, we have shown that using def2-TZVPP basis without counterpoise corrections provides a good compromise between accuracy and cost.

With the identification of density functional approximations that can accurately predict interaction energies of H_2 with a wide variety of binding moieties, we have expanded the set of in silico tools that can accelerate the discovery and validation of new hydrogen storage materials. These density functionals can be used to screen potential binding sites in a high-throughput fashion or can be used to train framework-specific force fields for use in molecular dynamics and/or Monte Carlo simulations. The H2Bind275 dataset contains only geometries in which H_2 is at the minimum of the potential energy curve with respect to the binding moeity, thereby maximizing the interaction energy. It would be useful for the H_2 storage community to see if these conclusions are extensible to other points on the potential energy curve using the strategies similar to the S22x5[390] and S66x8 dataset.[391] One could also conceive of similar assessments for properties like infrared frequencies. It would be useful to assess the performance of the best performing density functionals found in this work on real-life H_2 storage materials like MOFs, given the availability of highly accurate reference interaction energies. This will require careful consideration of the zero-point vibrational energies. Finally, future density functionals can be evaluated on this dataset to assess their suitability for hydrogen storage applications, or as part of testing functional performance for non-covalent interactions.

Chapter 5

Benchmarking Density Functionals for Predicting Potential Energy Curves in Hydrogen Storage Applications

5.1 Introduction

Hydrogen (H_2) is a favorable substitute for fossil fuels as the only by-product of hydrogen fuel cell engines is water and the efficiency of a fuel cell is significantly higher than an internal combustion engine. However, H_2 is a light gas with low volumetric and gravimetric energy densities. This poses a significant hurdle to storage and transportation of H_2. Storing H_2 reversibly in adsorbed form on porous materials is a promising solution to this problem.[392–395] Ideally, such materials should adsorb H_2 at high pressure and release it at low pressure so that the released H_2 can be used for operating a fuel cell. Designing materials with this property, while simultaneously not compromising on high volumetric and gravimetric storage capacities, is an active area of research.[283, 396, 397]

While multiple porous materials like Metal-Organic Frameworks (MOFs), Covalent Organic Frameworks (COFs), graphene, and other amorphous materials have been shown to adsorb H_2, none of these materials meet all the target criteria proposed by the U.S. Department of Energy for an ideal storage material.[283] As experimental synthesis and characterization of potential H_2 storage materials is expensive and time-consuming, computational modeling and screening of materials has emerged as a viable alternative to it.[296, 398, 399] Computational techniques can be used in two different, potentially complementary ways. First, molecular modeling can be used to understand the mechanism of H_2 binding in different porous materials and this understanding can be used to systematically tune materials to achieve target properties.[283, 285, 295] Second, computational techniques can be used to screen materials in a high-throughput manner to select only a handful of potentially viable materials for synthesis and characterization.

The ability of a material to store H_2 is characterized by its usable capacity, which is

defined as the amount of H_2 stored at the high operating pressure that is released when the pressure is reduced to the low operating pressure. Optimizing the usable capacity for typical fixed operating pressures of 5 bar and 100 bar gives an optimal value for the Gibbs free energy of adsorption (ΔG_{ads}). Assuming a correlation between enthalpy and entropy of adsorption in porous materials gives a range of -15 to -25 kJ/mol for the optimal value for enthalpy of adsorption (ΔH_{ads}).[130, 339, 340] The internal energy of binding, which is the largest component of ΔH_{ads}, can be computed using different quantum chemistry methods, including, but not limited to, density functional theory (DFT),[285, 295, 304, 400] Møller-Plesset perturbation theory (MP2),[401–403] and different variants of coupled-cluster theory.[134, 404, 405] Each of these methods have different accuracies and computational costs associated with them.

DFT, scaling as $\mathcal{O}(N^3)$ (N is the number of basis functions in the system), can provide a reasonable balance between cost and accuracy of computing H_2 binding energy. However, as the exact density functional remains unknown, different density functional approximations (DFAs), proposed in lieu of the exact density functional, provide varying accuracies for different chemical systems and/or properties computed.[44] In order to address this problem, we adopted a two-pronged approach.[134] (1) We compiled the H2Bind275 dataset that consists of H_2(s) interacting with binding motifs representative of different porous materials known for H_2 adsorption. This dataset consists only of equilibrium geometries, that is, H_2(s) are located at the minimum of the potential energy curve (PEC) with respect to the binding site. We computed highly accurate reference interaction energies using coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) extrapolated to the complete basis set limit for this dataset. (2) We assessed the performance of 55 DFAs and identified the best performing density functionals for this dataset. In addition, we also identified inexpensive semi-local density functionals which give very good performance for low computational cost and are suitable for in silico high-throughput screening purposes.

The H2Bind275 dataset, consisting of 275 data points, provides a balanced representation of different H_2 binding mechanisms like polarization, charge transfer interaction, and dispersion.[304, 406, 407] It also captures the chemical diversity of binding motifs that H_2 interacts with in porous frameworks. This dataset assesses the ability of density functionals to reproduce H_2 binding energies at the minima of the PEC. However, as DFAs are routinely used for geometry optimizations and molecular dynamics simulations either directly or indirectly (by generating reference data for training force fields), they should also be able to reproduce the entire PEC which would ensure accurate nuclear gradients as required for geometry optimization and molecular dynamics simulations. A strategy of assessing the performance of DFAs for PECs has been previously employed for other non-covalent interaction energy datasets like S22, S66, and A24. The S22x5 dataset[390] was created from the S22 dataset[408] by including geometries that are shortened and elongated along a well-defined interaction coordinate. Similarly, the S66x8[391] and A21x12[409] extended datasets were created from the S66 and A24 datasets.[309]

In order to address this issue for H_2 storage, we have extended the H2Bind275 dataset to include geometries that are located at five different points on 78 separate PECs, not just

Table 5.1: Number of geometries and data points by chemical categories for the H2Bind78×7 dataset

	s-block ions	salts	organic ligands	transition metals	total
geometries	19	13	5	41	78
data points at PEC minimum	38	26	10	82	156
data points not at PEC minimum	95	65	25	204[1]	389
H2Bind78×7	133	91	35	286	545

the minimum. This extended dataset, hereafter referred to as the H2Bind78×7 dataset, was generated by shortening and stretching the distance between H_2 and the binding motif. The reference interaction energies were computed using CCSD(T) extrapolated to the complete basis set (CBS) limit using the same strategy outlined in Ref. 134. The performance of 55 DFAs were assessed using regularized relative errors metrics by appropriately weighing the error coming from different points on the PECs. We have analyzed the performance of these DFAs for the extended dataset by comparing and contrasting it with the performance of the original equilibrium H2Bind275 dataset.

This paper is organized as follows. The H2Bind78×7 dataset is introduced and the protocol for computing reference interaction energies is discussed. All the density functionals chosen for assessment in this work are briefly introduced and classified. The distribution of the reference interaction energies at different points on the PEC is outlined. The performance of DFAs for predicting H_2 interaction energies across the PEC is discussed and contrasted with their performance for the previous H2Bind275 dataset. The performance of DFAs for predicting equilibrium geometries and interaction energies at equilibrium geometries is explored. The best DFAs for predicting H_2 binding energies are recommended while considering their computational cost.

5.2 Computational details

H2Bind78×7 dataset

The H2Bind275 dataset consists of 275 H_2 interaction energies but only 86 unique geometries as many of them have multiple H_2s. For example, the geometry of $CaCl_2-(H_2)_4$ has four hydrogen molecules bound to $CaCl_2$ contributing four data points to the H2Bind275 dataset. The H2Bind78×7 dataset was generated by starting from a subset of the original H2Bind275 dataset. This subset was created by choosing only 78 unique geometries and computing their interaction energies adiabatically using the method outlined in Ref. 134. All of these geometries are located on the PEC at their respective minima. The adiabatic interaction energy, which relaxes the geometries of the binding motif and H_2, was chosen as it is closest to

[1]One data point excluded due to convergence issues

experimentally measurable values. For each minimum geometry, five additional geometries were generated by compressing and elongating the distance between the binding motif and the center of mass of H_2 (denoted by r_{eq}). For geometries containing multiple H_2s bound to a single binding moiety, compressed and elongated geometries were generated for only one of the H_2s. This step was necessary in order to maintain low redundancy in the dataset and make its size manageable. In addition to this, the interaction energy at the minimum of the PEC was also computed using vertical interaction energy method. In total, this dataset contains 78 adiabatic and 78 vertical interaction energies (a total of 156 data points) located at the PEC minimum.

In this work, two compressed geometries ($0.75r_{eq}$ and $0.9r_{eq}$) and three elongated geometries ($1.1r_{eq}$, $1.25r_{eq}$, and $1.5r_{eq}$) were considered. These distances were chosen as they are representative of the PEC in both the compressed and elongated regimes. In a porous material, H_2 interacts with not only its primary binding site but also has secondary interactions with other components of the framework. The binding distances of H_2 to its secondary interaction sites of the porous material are often longer than their corresponding equilibrium distances. As a consequence of this, when modeling H_2 in a porous material, the elongated portion of the PEC is sampled more than the compressed part. Additionally, the compressed portion of the PEC is usually significantly higher in energy (repulsive if compressed enough), and is sampled less often in a molecular dynamics or Monte Carlo simulation. Hence, DFAs should be able to reproduce the elongated portion of the PEC more faithfully than the compressed portion. We have included more data points in the elongated regime than the compressed regime in order to underscore its relative importance. As shown in Table 5.1, the number of non-equilibrium data points is roughly 2.5 times the number of data points at equilibrium. In total, counting both the equilibrium and non-equilibrium data points, this H2Bind78×7 dataset consists of 545 H_2 interaction energies with representative binding motifs.

This dataset, like the H2Bind275 dataset, can also be divided into categories based on the chemical nature of the binding motif as shown in Table 5.2: (1) s-block ions: consisting of group 1 and group 2 bare metal cations with unscreened charge binding one or multiple H_2s, (2) salts: consisting of small inorganic salts like AlF_3, $CaCl_2$, and MgF_2 binding one or multiple H_2s, (3) organic ligands: comprising of small aliphatic and aromatic molecules binding one H_2, (4) transition metals: including small transition metal complexes and 3d transition metal cations binding one or multiple H_2s. Each of these categories is also representative of various mechanisms of H_2 binding found in porous materials. For example, H_2 in the organic ligands category is mostly dispersion-bound. The s-block metals category binds H_2 using a combination of electrostatic and forward charge transfer ($H_2 \rightarrow$ metal) interactions.[304] This dataset captures both chemical and mechanistic diversity encountered in H_2 binding to porous materials. For a detailed discussion about different chemical categories in this dataset, we refer readers to Ref. 134.

Table 5.2: All 78 geometries in the H2Bind78×7 dataset categorized by chemical identity of the binding motif.

s-block ions	salts	organic ligands	transition metals
$Li^+-(H_2)_n$, $n = 1,2,3,4,5,6$	AlF_3-H_2	benzene$-H_2$	$MX-H_2$, X=H, F, Cl; M=Cu, Ag, Au
$Na^+-(H_2)_n$, $n = 1,2,3$	$CaF_2-(H_2)_n$, $n = 1,2,3,4$	phenol$-H_2$	CoF_3-H_2
$Mg^{2+}-(H_2)_n$, $n = 1,2,3,4$	$CaCl_2-(H_2)_n$, $n = 1,2,3,4$	pyrrole$-H_2$	$Cu(OMe)-H_2$
$Ca^{2+}-(H_2)_n$, $n = 1,2,3,4,5,6$	$MgF_2-(H_2)_n$, $n = 1,2,3,4$	butene$-H_2$	$CuCN-H_2$
		tetrazole$-H_2$	$Sc^+-(H_2)_n$, $V^+-(H_2)_n$, $n = 3,4$
			$Ti^+-(H_2)_n$, $n = 2,4$
			$Cr^+-(H_2)_n$, $Mn^+-(H_2)_n$, $n = 1,2,3,4$
			$Fe^+-(H_2)_n$, $n = 1,2,3,4$
			$Co^+-(H_2)_n$, $Ni^+-(H_2)_n$, $n = 1,2$
			$Cu^+-(H_2)_n$, $n = 1,2,3$
			$Zn^+-(H_2)_n$, $n = 1,2,3,4$

Reference Binding Energies

Calculation of accurate reference interaction energies is an important task in compiling a dataset. Reference interaction energies were computed using coupled-cluster theory with singles, doubles, and perturbative triples (CCSD(T))[60] extrapolated to the complete basis set limit. Inspired by the success of composite extrapolation methods[323, 326–328] for computing highly accurate reference values, we have developed our own composite extrapolation method using focal point analysis[329, 330] for computing accurate reference H_2 binding energies:

$$E_{\text{ref}} = E_{\text{HF/5Z}} + E_{\text{MP2/QZ}\rightarrow\text{5Z}} + \delta E_{\text{CCSD(T)/TZ}} + \delta E^{\text{core}}_{\text{MP2/TZ}} \tag{5.1}$$

$$\delta E_{\text{CCSD(T)/TZ}} = E_{\text{CCSD(T)/TZ}} - E_{\text{MP2/TZ}} \tag{5.2}$$

$$\delta E^{\text{core}}_{\text{MP2/TZ}} = E^{\text{core=0}}_{\text{MP2/TZ}} - E^{\text{core=n}}_{\text{MP2/TZ}} \tag{5.3}$$

Here, E_{ref} is the reference energy computed using the composite method, $E_{\text{HF/5Z}}$ is the Hartree Fock energy computed using a basis set of quintuple-zeta (5Z) quality, $E_{\text{MP2/QZ}\rightarrow\text{5Z}}$ is the MP2 correlation energy extrapolated to the complete basis set limit with the 2-point extrapolation formula[45] using correlation energies computed with quadruple-zeta (QZ) and quintuple-zeta quality basis sets, and $\delta E_{\text{CCSD(T)/TZ}}$ is the difference between the CCSD(T) and MP2 correlation energies computed with a triple-zeta quality basis set. $\delta E^{\text{core}}_{\text{MP2/TZ}}$ is the core-valence contribution to the correlation energy computed as the difference between MP2 correlation energies with ($E^{\text{core=n}}_{\text{MP2/TZ}}$) and without ($E^{\text{core=0}}_{\text{MP2/TZ}}$) the frozen-core approximation. This composite method for computing reference H_2 interaction energies ensures that the effect of higher-order excitations neglected in CCSD(T) are sufficiently small. It also ensures that the basis set incompleteness errors are small and that both HF and extrapolated correlation

energy components are of complete basis set limit quality. Further details of this scheme can be found in Ref. 134.

The cc-pVnZ[331, 332] (n =T, Q, or 5) family of basis sets was used for all the HF and correlation energy calculations when core electrons were not included in the correlation calculations. cc-pCVnZ[333, 334] (n =T, Q, or 5) family of basis sets were employed when some or all of the core electrons were included in the correlation calculations. For transition metals, the cc-pwCVnZ[335] (n =T, Q, or 5) series of basis sets was used with a neon core excluded in all correlation energy computations.

Density Functional Approximations

55 DFAs, including all the commonly used density functionals, were chosen to perform a thorough assessment. We have also included DFAs that have previously shown very good performance for a range of non-covalent interaction energy prediction problems represented by multiple datasets.[18] Based on the different quantities DFAs depend on, they are categorized into rungs of the metaphorical Jacob's ladder.[20] From the first rung of the Jacob's ladder, in which DFAs depend only on electron density, SVWN5[21, 22] and SPW92[21, 23] DFAs were chosen. From the second rung called Generalized Gradient Approximation (GGA), 12 different DFAs were chosen: the PBE family and its variants (PBE,[24] PBE-D3(0),[37] RPBE,[343] revPBE,[344] and revPBE-D3(op)[345]), BLYP[25, 26] and BLYP-D3(op)[345], dispersion-corrected variants of B97[346] (BLYP-D3(0)[37] and BLYP-D3(BJ)[347]), PW91,[349] and GAM.[348] From the meta-GGA rung, the different variants of TPSS (TPSS,[27] TPSS-D3(BJ),[347] and revTPSS[351]), SCAN[352] and its dispersion-corrected version SCAN-D3(BJ)[353], MS2[359] and MS2-D3(op),[345] the combinatorially-optimized B97M-V[28] and B97M-rV[354, 355] were chosen. In addition, mBEEF[356] and the semi-local Minnesota functionals M06-L[357] and MN15-L[358] were also included in the assessment. Rung four DFAs, containing HF exchange, are generally more accurate than semi-local functionals as they partially alleviate the problem of self-interaction error. In this work, global hybrid density functionals like B3LYP[29] and B3LYP-D3(0),[347] PBE0[361] and PBE0-D3(BJ),[347] TPSSh[362] and TPSSh-D3(BJ),[347] the M06 family of density functionals (M06,[364] M06-2X,[364] M06-2X-D3(0),[37] and revM06[365]), MVSh,[367] and SCAN0[363] which is the hybrid variant of SCAN are included. Range-separated hybrids, which are hybrid functionals containing DFT exchange and some HF exchange in the short-range and only HF exchange in the long range, included in this study are ωB97X-D,[258] ωB97X-D3,[371] ωB97X-V,[372] ωB97M-V,[44] M11[373] and its revised version revM11.[374] Two screened exchange density functionals (HSE-HJS[375, 376] and MN12-SX[377]), which contain DFT exchange in the short range and attenuated HF exchange in the long range are also included. Double hybrid density functionals, which are at the top the Jacob's ladder classification, contain some percentage of correlation energy from wavefunction methods. These DFAs are characterized by their superior accuracy and increased computational cost in comparison to semi-local and hybrid DFAs. We have included seven double

hybrid density functionals in this study: B2PLYP-D3(BJ),[36] XYG3,[33] XYGJ-OS,[380] PBE0-DH,[379], PTPSS-D3(0),[382] DSD-PBEPBE-D3(BJ),[381] and ωB97M(2).[410]

The def2-QZVPPD[173] basis set was used for all DFA calculations with a quadrature grid of 99 Euler-MacLaurin radial points and 590 Lebedev angular points for integrating the exchange-correlation contribution. SG-1[175] integration grid was used for integrating the VV10 component. The choice of core for frozen core approximation and employment of density fitting approximation for computing MP2 correlation energy in double hybrid density functionals is discussed in the Table D.1. All the PECs were interpolated using the one-dimensional Akima interpolator.[411] All computations were performed using Q-Chem 5.[176]

5.3 Results and Discussion

H2Bind78×7 dataset

Typically, the coupled cluster reference H_2 interaction energy with the binding motif is strongest at equilibrium, that is at r_{eq}. This implies that the geometries optimized using ωB97M-V/def2-TZVPD[173] are also close to the CCSD(T)/CBS minima. Fig. 5.1 shows the distribution of interaction energies for the entire H2Bind78×7 dataset consisting of 545 data points. The equilibrium and geometries near equilibrium ($0.9r_{eq}, 1.0r_{eq}$, $1.0r_{eq}^{vert}$, and $1.1r_{eq}$) have attractive interaction energies, with most of them smaller than 100 kJ/mol in magnitude. Geometries that are stretched by 25% ($1.25r_{eq}$) are still attractive in nature, but most interaction energies are smaller than 60 kJ/mol in magnitude. Geometries stretched by 50% of their equilibrium distance are bound only weakly with a median binding energy of −6.9 kJ/mol. At the other extreme of the PEC, geometries that are compressed by 25% ($0.75r_{eq}$) are mostly repulsive with a median interaction energy of +29.7 kJ/mol. This geometry was also chosen in order to sample the repulsive part of the PEC and assess how accurately different density functionals can reproduce it.[412]

The range of interaction energies covered by each PEC is also very large. The coinage metal containing species are the strongest binders, as illustrated by the extreme example of AuF which binds H_2 with an interaction energy of −161.8 kJ/mol at equilibrium and interacts with H_2 with an energy of +87.6 kJ/mol (repulsive) at $0.75r_{eq}$, thus spanning an interaction range of 249.4 kJ/mol. Data points in the organic category have the smallest ranges (average range is 19 kJ/mol). A typical PEC of H_2 interacting with a binding moiety has the shape of a Morse potential. However, there is considerable variation in the well depth, well width, and decay in the long range for different chemical species. This variation can provide some clues into the dominant mechanism of interaction. For example, the AuCl binding motif interacts with one H_2 with an interaction energy of −123.7 kJ/mol at equilibrium which decays to −21.3 kJ/mol at $1.5r_{eq}$ (82.7% decrease). This is a sharp decay in the interaction energy in comparison to the Mg^{2+} case. In the Mg^{2+} interacting with one H_2 case, the interaction energy at equilibrium is −97.8 kJ/mol in comparison to −37.7 kJ/mol at $1.5r_{eq}$

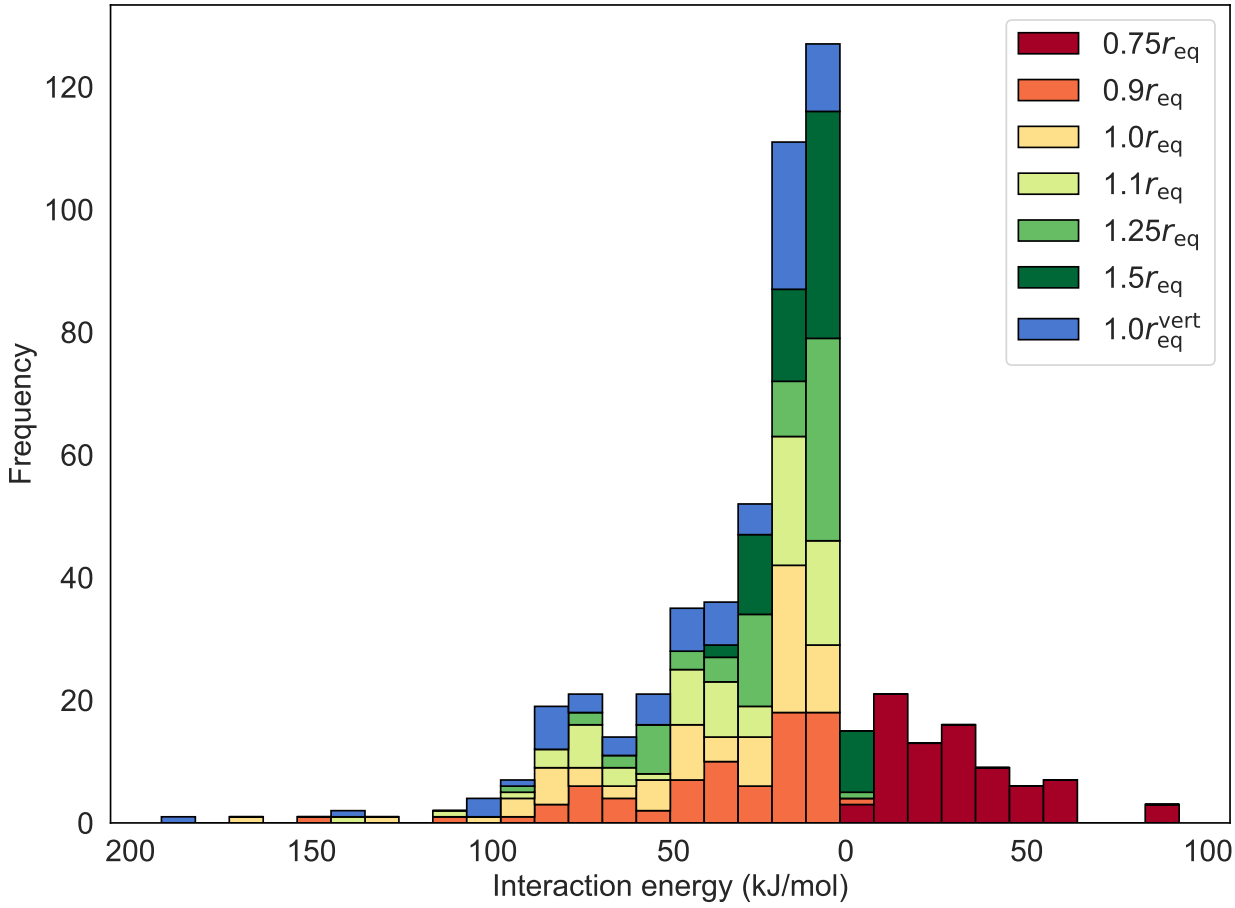


Figure 5.1: Distribution of coupled-cluster reference interaction energies separated by location on the potential energy curve. The reference vertical interaction energy at equilibrium ($1.0r_{eq}^{vert}$) is also shown.

(61.4% decrease). This suggested that the dominant mechanism of interaction in Mg^{2+} case is longer-ranged (like permanent electrostatics) than AuCl which is dominated by orbital controlled short-ranged interactions like charge transfer.

Performance of Density Functional Approximations on PECs

We will discuss the performance of DFAs using multiple error metrics. Each of these metrics gives different weights to different aspects of the dataset. First, we will discuss the performance of DFAs using the root mean square error (RMSE) metric which gives equal importance to all data points in the H2Bind78×7 dataset. The RMSE of all 55 DFAs assessed in this work is shown in Table 5.3. The non-empirical double hybrid functional with just two fixed parameters, PBE0-DH, gives the least RMSE of 2.9 kJ/mol. The second best

Table 5.3: Regularized mean absolute percentage error (RegMAPE) and root mean squared error (RMSE; in kJ/mol) of all DFAs considered in this work for the entire H2Bind78×7 dataset.

Rank	DFA	RMSE (kJ/mol)	DFA	RegMAPE (%)
1	PBE0-DH	2.9	PBE0-DH	5.0
2	DSD-PBEPBE-D3(BJ)	3.7	ωB97X-V	5.4
3	ωB97X-V	4.0	ωB97M-V	6.3
4	ωB97X-D	4.1	DSD-PBEPBE-D3(BJ)	6.3
5	PBE0	4.2	XYGJ-OS	6.9
6	MVSh	4.3	ωB97M(2)	7.4
7	HSE-HJS	4.3	PBE0	7.6
8	ωB97M-V	4.5	HSE-HJS	7.6
9	XYGJ-OS	4.6	B2PLYP-D3(BJ)	8.2
10	ωB97X-D3	4.8	XYG3	8.5
11	XYG3	4.8	ωB97X-D	9.0
12	PTPSS-D3(0)	4.9	B97M-rV	9.0
13	ωB97M(2)	5.1	B97M-V	9.1
14	PBE0-D3(BJ)	5.2	SCAN0	9.1
15	MN15	5.7	PTPSS-D3(0)	9.3
16	B2PLYP-D3(BJ)	5.8	ωB97X-D3	9.8
17	SCAN0	5.8	MVSh	10.2
18	TPSSh	6.0	TPSSh	11.3
19	revM11	6.3	PBE0-D3(BJ)	11.5
20	mBEEF	6.7	M11	12.0
21	revM06	6.9	revTPSS	12.0
22	B3LYP	7.4	revM06	12.4
23	revTPSS	7.5	TPSS	13.6
24	B3LYP-D3(0)	7.5	TPSSh-D3(BJ)	13.8
25	B97M-V	7.6	B3LYP-D3(0)	14.0
26	B97M-rV	7.6	oTPSS-D3(BJ)	14.6
27	TPSS	7.7	TPSS-D3(BJ)	14.7
28	oTPSS-D3(BJ)	7.8	MN15	15.3
29	TPSSh-D3(BJ)	8.0	PBE	15.3
30	MN15-L	8.0	revM11	15.3
31	revPBE-D3(op)	8.5	BLYP-D3(op)	15.4
32	TPSS-D3(BJ)	8.5	B3LYP	15.5
33	revPBE	9.1	SCAN	15.6
34	MN12-SX	9.2	MN12-SX	16.2
35	RPBE	9.3	SCAN-D3(BJ)	16.3
36	M11	9.4	M06	16.9
37	B97-D3(BJ)	9.7	revPBE-D3(op)	17.0
38	BLYP-D3(op)	9.8	PW91	17.4
39	M06	9.9	MS2	17.4
40	BLYP	10.1	mBEEF	17.4
41	PBE	10.3	BP86-D3(BJ)	18.4
42	M06-L	10.5	M06-2X	19.3
43	BP86-D3(BJ)	10.9	M06-2X-D3(0)	19.8
44	B97-D3(0)	11.0	MN15-L	19.8
45	PBE-D3(0)	11.0	PBE-D3(0)	20.1
46	PW91	11.2	MS2-D3(op)	20.3
47	GAM	11.8	M06-L	20.7
48	MS2	12.1	RPBE	20.9
49	MS2-D3(op)	12.4	BLYP	22.1
50	SCAN	12.9	revPBE	23.4
51	SCAN-D3(BJ)	13.2	B97-D3(BJ)	24.2
52	M06-2X	13.4	GAM	24.7
53	M06-2X-D3(0)	13.4	B97-D3(0)	28.3
54	SPW92	32.6	SPW92	63.0
55	SVWN5	32.7	SVWN5	63.0

DFA is another double hybrid DSD-PBEPBE-D3(BJ) with an RMSE of 3.7 kJ/mol. In comparison to the earlier H2Bind275 dataset where DSD-PBEPBE-D3(BJ) was ranked fourth, it performs relatively better for the H2Bind78×7 dataset moving up by two places.[134] This is closely followed by ωB97X-V and ωB97X-D, both of which show a similar RMSEs of 4.0 and 4.1 kJ/mol respectively. Another trend seen in the H2Bind275 dataset that is transferable to the H2Bind78×7 dataset is that the best performing DFA in each rung of the Jacob's ladder performs better than the best performing functional in rung directly below it. The least RMSE DFA in each rung also remains the same: SPW92 for LDA, revPBE-D3(op) for GGAs, mBEEF for meta-GGAs, ωB97X-V in the hybrids rung, and PBE0-DH in the double hybrids rung. The ranking of ωB97M-V deteriorates in the extended dataset in comparison to the previous H2Bind275 dataset. Another interesting observation is the improvement in the ranking of the MN15 density functional which is ranked 15$^{\text{th}}$ in the H2Bind78×7 dataset with an RMSE of 5.7 kJ/mol (MN15 was ranked 25$^{\text{th}}$ with an RMSE of 6.3 kJ/mol in the H2Bind275 dataset). The relative performance of B97M-V and B97M-rV (ranked 25$^{\text{th}}$ and 26$^{\text{th}}$) in the H2Bind78×7 dataset remains comparable to their performance in the H2Bind275 dataset. We also note that commonly used density functionals like M06-2X and M06-2X-D3(0) and the recently developed density functionals like SCAN and SCAN-D3(BJ) show very large RMSEs.

The reference interaction energies in the H2Bind78×7 dataset span a very large range: from −189.0 to 92.1 kJ/mol. However, for H_2 storage applications between 5 and 100 bar, interaction energies in the −15 to −25 kJ/mol range would be ideal.[130, 339, 340] A good error metric should give more weight to data points in this range by considering the following aspects:

1. The H2Bind78×7 dataset contains many model binding motifs, each of them binding H_2 with different interaction energies at their corresponding equilibrium geometry. Binding motifs that bind H_2 with an interaction energy in the range of −15 to −25 kJ/mol should have larger weights.

2. Each binding motif contributes six data points: two compressed data points, three elongated data points, and one data point at equilibrium. Equilibrium geometries should be given larger weight than the non-equilibrium ones.

The regularized mean absolute percentage error (RegMAPE) was formulated in Ref. 134 in order to satisfy requirement (1). RegMAPE uses percentage error in the −15 to −25 kJ/mol range, regularized percentage error (in order to avoid small denominators) for interaction energies weaker than −15 kJ/mol, and absolute error for interaction energies stronger than −25 kJ/mol. The error metrics in neighboring ranges are also smoothly interpolated. For the same amount of error, as percentage error is much larger in magnitude than absolute error, the RegMAPE error metric is able to satisfy criterion (1). For example, an error of 5 kJ/mol for a reference interaction energy of 100 kJ/mol will contribute 5 units to the total error while the same error for a reference interaction energy of 20 kJ/mol will contribute 25 units to the total error. For a given binding motif, the equilibrium geometry should be given more

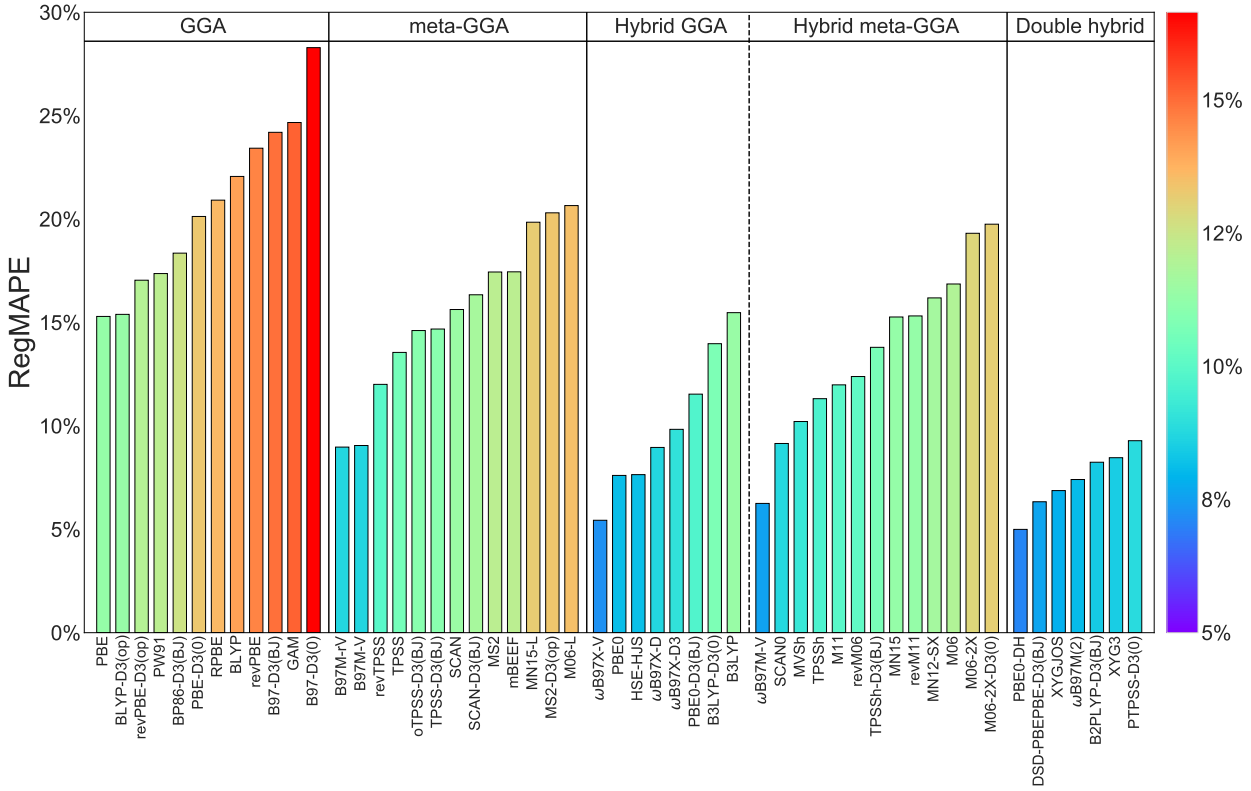


Figure 5.2: Performance of density functional approximations for the H2Bind78×7 dataset assessed using regularized mean absolute percentage error (RegMAPE). The LDA density functionals, SPW92 and SVWN5, are not included in this figure and show a large RegMAPE of 63.0%.

weight as it represents the H_2 interaction with the primary binding site: the main lever to tune while designing binding sites. As non-equilibrium geometries are higher in energy, they would be encountered less frequently in a molecular dynamics or Monte Carlo simulation. Hence, lower weight for non-equilibrium geometries is achieved by using the equilibrium regularization value for non-equilibrium geometries as well. As the equilibrium geometry always has a stronger interaction energy, the regularized value of its reference interaction energy, $\tilde{E}(r_{\text{eq}})$, will be larger in magnitude in comparison to the regularized values of non-equilibrium interaction energies ($\tilde{E}(\alpha r_{\text{eq}}), \alpha \neq 1.0$). The large magnitude of the denominator will give a smaller weight to the errors of non-equilibrium geometries in comparison to the equilibrium one. RegMAPE for equilibrium and non-equilibrium geometries as defined in Eq. (5.4) satisfies requirement (2).

$$\Delta E(\alpha r_{\text{eq}}) = \frac{E^{\text{DFA}}(\alpha r_{\text{eq}}) - E^{\text{ref}}(\alpha r_{\text{eq}})}{\tilde{E}(r_{\text{eq}})}, \quad \alpha \in \{0.75, 0.9, 1.0, 1.1, 1.25, 1.5\} \tag{5.4}$$

where $\Delta E(\alpha r_{\text{eq}})$ is the RegMAPE, $E^{\text{DFA}}(\alpha r_{\text{eq}})$ and $E^{\text{ref}}(\alpha r_{\text{eq}})$ are the DFA and reference interaction energies at αr_{eq} geometry, and $\tilde{E}(r_{\text{eq}})$ is the regularized interaction energy for the equilibrium geometry. As the vertical interaction energy is located at the minimum of the PEC, the error in this data point is regularized using the vertical reference interaction energy.

The performance of DFAs assessed by the RegMAPE error metric is shown in Fig. 5.2 and Table (5.3). While there are some similarities in the relative ordering of density functionals for RegMAPE and RMSE error metrics, there are also noteworthy differences. Again, PBE0-DH shows the best performance with the least RegMAPE of 5.0%. It is followed by the ωB97X-V and ωB97M-V density functionals which have RegMAPEs of 5.4% and 6.3% respectively. The DSD-PBEPBE-D3(BJ) density functional, which was the best performing density functional in the H2Bind275 dataset with RegMAPE of 4.9%, is the fourth best performing DFA for the H2Bind78×7 dataset. The small decline in the performance of DSD-PBEPBE-D3(BJ) can be attributed to its poor performance for the non-equilibrium geometries as shown in the Table D.2. Another noteworthy decline in performance is that of the B97M-V and B97M-rV functionals. These functionals were the best performing semi-local density functionals in the H2Bind275 dataset (ranked 5$^{\text{th}}$ and 6$^{\text{th}}$), and were recommended as inexpensive alternatives to the best performing and expensive density functionals.[134] However, their performance in the current H2Bind78×7 dataset deteriorates with B97M-rV and B97M-V yielding errors of 9.0% and 9.1% (ranked 14$^{\text{th}}$ and 15$^{\text{th}}$) respectively. While this reflects their lacklustre performance for the non-equilibrium geometries, they still remain the best performing semi-local functionals. The next best performing semi-local density functional, revTPSS, is ranked 21$^{\text{st}}$ and shows a RegMAPE of 12.0%. The performance of PBE0 and B2PLYP-D3(BJ) DFAs shows significant improvement relative to their performance in the H2Bind275 dataset with both density functionals entering the top 10 category for the H2Bind78×7 dataset.

Other general trends also hold for DFAs assessed with the RegMAPE error metric. The best DFA of each rung of the Jacob's ladder outperforms the best DFA from the rung below it. The best performing meta-GGA functional is B97M-rV with a RegMAPE of 9.0% and the best GGA is PBE with a RegMAPE of 15.3%. The effect of addition of empirical dispersion correction can also be assessed using the mean signed error (MSE) and RegMAPE metrics. Addition of dispersion correction improves the performance only if the parent density functional has a systematic underbinding problem (characterized by a positive value of MSE). For example, B3LYP has an MSE of 2.7 kJ/mol and a RegMAPE of 15.5% and is systematically underbinding H_2(s). Addition of a dispersion correction to B3LYP leads to the B3LYP-D3(0) functional which overcomes this underbinding problem. B3LYP-D3(0) slightly overbinds with an MSE of -0.8 kJ/mol, but shows an improved RegMAPE of 14.0%. Addition of DFT-D corrections also improves the performance of other underbinding DFAs like revPBE and BLYP. However, addition of these corrections to parent DFAs that are already overbinding exacerbates the overbinding issue leading to poorer performance as exemplified by PBE, PBE0, TPSS, SCAN, and MS2 functionals. Remarkably, PBE without any dispersion correction is the best performing GGA. As empirical dispersion correction is distance

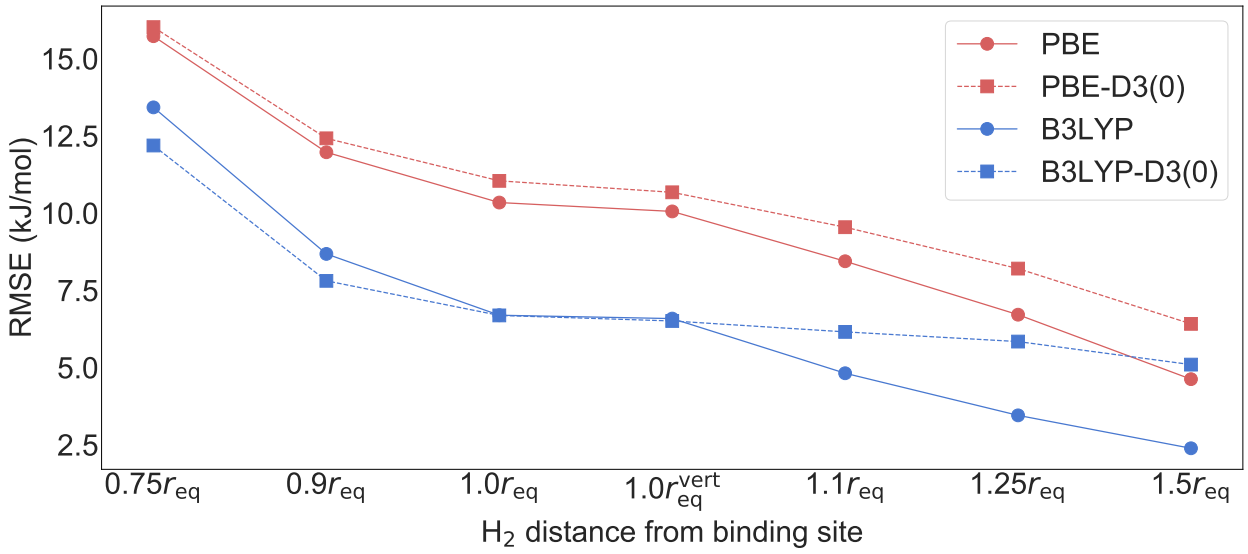


Figure 5.3: Effect of addition of empirical dispersion corrections on the RMSE of overbinding (PBE) and underbinding (B3LYP) density functionals at different points on the potential energy curve.

dependent, it is interesting to see its effect at different points on the PEC. For overbinding functionals, addition of dispersion correction worsens their performance across the PEC as exemplified by the PEC of PBE and PBE-D3(0) in Fig. 5.3. The difference between the RMSEs of PBE and PBE-D3(0) increases with increasing distance of H_2 with the binding site as dispersion corrections are usually damped in the short range. Dispersion corrections improve the performance of underbinding functionals in the short range. However, in the long range, dispersion corrections overestimate its magnitude, causing BLYP, B3LYP, and revPBE to overbind in the elongated regime. MSEs and RMSEs of all the DFAs containing dispersion corrections and their corresponding parent functionals is shown in the Table D.3.

Addition of HF exact exchange is essential to ameliorate the effect of self interaction error in density functionals. Comparing DFAs belonging to the same family, addition of HF exchange improves the performance of semi-local functionals for the H2Bind78×7 dataset. PBE0, which contains 25% HF exchange, is ranked the 7^{th} with a RegMAPE of 7.6%. In contrast, the PBE functional is ranked 29^{th} with a RegMAPE for 15.3%, more than two times that of PBE0. HF exchange is a short-range effect and addition of HF exchange improves the performance of density functionals in the short range as shown in Fig. 5.4. While the hybrid functional performs better than its semi-local counterpart throughout the PEC, its effect is more pronounced in the compressed region than the elongated region. The SCAN and SCAN0 functionals show RMSEs of 21.6 and 9.5 kJ/mol (a difference of 12.2 kJ/mol) at $0.75r_{eq}$ of the PEC. Their RMSEs at $1.5r_{eq}$ is 3.9 and 2.1 kJ/mol, with the hybrid functional improving on the semi-local one by only 1.8 kJ/mol.

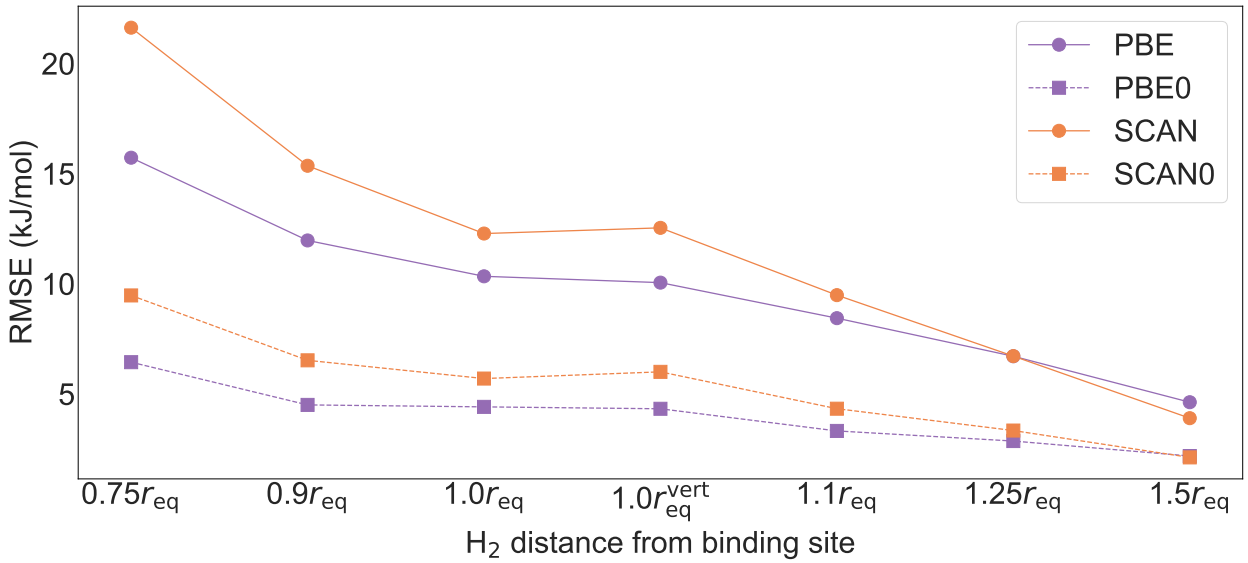


Figure 5.4: Performance of density functionals of the same family with and without Hartree Fock exchange at different points on the potential energy curve.

The RegMAPE error metric gives larger weights to data points whose reference interaction energies are in the interesting range for H_2 storage. Further, it gives more weight to the equilibrium data point than non-equilibrium data points. The relative weights of data points on the PES can be further tuned in order to assess the origin of errors of different DFAs. Elongated geometries are encountered more often than compressed geometries in porous material capable of storing H_2. Compressed geometries are also much higher in energy (geometries compressed by 25% are almost always repulsive) and are encountered less often in simulations. This would suggest retuning the weights of the regularized errors by giving larger weights to equilibrium and elongated regions of the PEC. The weighted RegMAPE (denoted as wRegMAPE or ΔE_w) is defined as:

$$\Delta E_w = \sum_i \frac{w_i \Delta E(\alpha_i r_{eq})}{7} \quad \text{s.t.} \sum_i w_i = 7 \tag{5.5}$$

where $\Delta E(\alpha_i r_{eq})$ is the RegMAPE at the point $\alpha_i r_{eq}$ defined in Eq. (5.4). Ensuring that the weights sum up to 7 would enable an apples-to-apples comparison of wRegMAPE and RegMAPE. In the case of RegMAPE, $w_i = 1$, for all values of i. Reducing the weights of the compressed geometries with the scheme shown in Table 5.4, the wRegMAPE can be computed using Eq. (5.5). This wRegMAPE metric, shown in Table 5.5, gives more weight to the elongated geometries. As the vertical interaction is computed its respective PEC minimum, the $1.0r_{eq}^{vert}$ data point is assigned a weight equal to that of the adiabatic interaction energy at PEC minimum.

Table 5.4: Weights for different points on the adiabatic PEC and vertical interaction energy for calculating the weighted regularized mean absolute percentage error (wRegMAPE) metric.

PEC location	$0.75r_{\mathrm{eq}}$	$0.9r_{\mathrm{eq}}$	$1.0r_{\mathrm{eq}}$	$1.0r_{\mathrm{eq}}^{\mathrm{vert}}$	$1.1r_{\mathrm{eq}}$	$1.25r_{\mathrm{eq}}$	$1.5r_{\mathrm{eq}}$
Weight (w_i)	0.75	0.91	1.07	1.07	1.07	1.07	1.07

Table 5.5: Weighted regularized mean absolute percentage error (wRegMAPE) for selected density functional approximations.

Rank	DFA	wRegMAPE (%)
1	PBE0-DH	4.8
2	ωB97X-V	5.2
3	DSD-PBEPBE-D3(BJ)	5.9
4	ωB97M-V	6.0
5	XYGJ-OS	6.5
6	ωB97M(2)	7.0
7	PBE0	7.4
8	HSE-HJS	7.4
9	XYG3	7.9
10	B2PLYP-D3(BJ)	8.0
11	B97M-rV	8.4
12	B97M-V	8.4
13	ωB97X-D	8.5
14	SCAN0	8.7
15	PTPSS-D3(0)	9.1

Comparing the magnitude of the wRegMAPE (Table 5.5) of different DFAs to their corresponding RegMAPE (Table 5.3), we can notice that the wRegMAPEs are slightly smaller. Smaller wRegMAPEs suggest that density functionals perform better for the equilibrium and elongated geometries in comparison to the compressed ones. However, the relative ordering of density functionals remains more or less the same. The top five best performing DFAs (PBE0-DH, ωB97X-V, ωB97M-V, DSD-PBEPBE-D3(BJ), and XYGJ-OS) according to the RegMAPE metric are also the five best performing functionals according to the wRegMAPE metric. We see that the recently parametrized ωB97M(2) double hybrid density functional, which uses ωB97M-V orbitals, is ranked sixth with wRegMAPE of 7.0%.

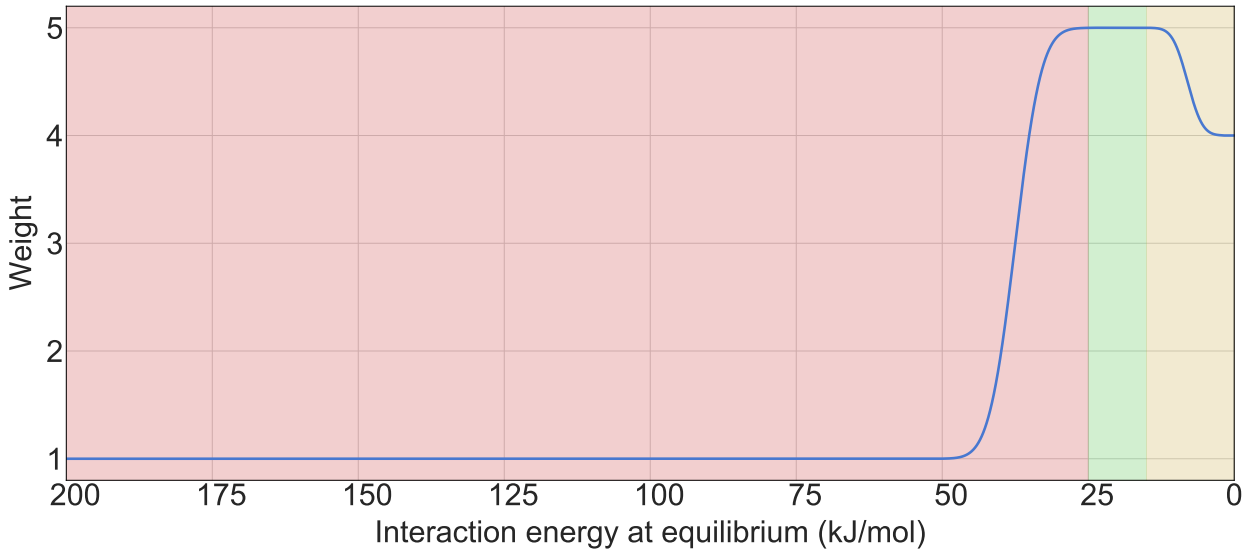


Figure 5.5: Weights of different chemical species as a function of their reference adiabatic interaction energy at equilibrium.

Performance of Density Functional Approximation for Geometries

All the equilibrium geometries for this dataset were obtained by geometry optimization using the ωB97M-V density functional in the def2-TZVPD basis set.[173] With the exception of two chemical systems (AlF_3−H_2 and Ti^+−$(H_2)_2$), the CCSD(T)/CBS equilibrium geometry of all other chemical systems coincides (up to sampling precision) with the ωB97M-V/def2-TZVPD equilibrium geometry. This further validates the choice of equilibrium geometries for the H2Bind78×7 dataset.

In order to assess the error in prediction of equilibrium geometry in a manner that is sensitive for H_2 storage purposes, we have devised a weighting scheme that gives larger weights to more relevant data points. Data points with reference adiabatic interaction energy at equilibrium ($E^{\text{ref}}(1.0r_{\text{eq}})$) in the range of −15 to −25 kJ/mol are given a weight of 5.0. Equilibrium interaction energies stronger than −25 kJ/mol are assigned a weight of 1.0. These weights were chosen to reflect the relative importance of these data points in the RegMAPE metric. In the RegMAPE metric, a density functional yielding an error of 1 kJ/mol in the strong binding regime contributes 1 unit to the total error as absolute error metric is used in this regime. A DFA with an error of the 1 kJ/mol in the middle of the favorable regime for H_2 storage (that is at −20 kJ/mol) contributes 5 units to the total error as percentage error metric is used. The RegMAPE metric assigns a weight that is 5 times larger to the species in the favorable regime in comparison to the strong binders, thus justifying the weights of 5.0 and 1.0 in Fig. (5.5). The weak binders with equilibrium interaction energies weaker than −15 kJ/mol are mostly comprised of the organic ligands.

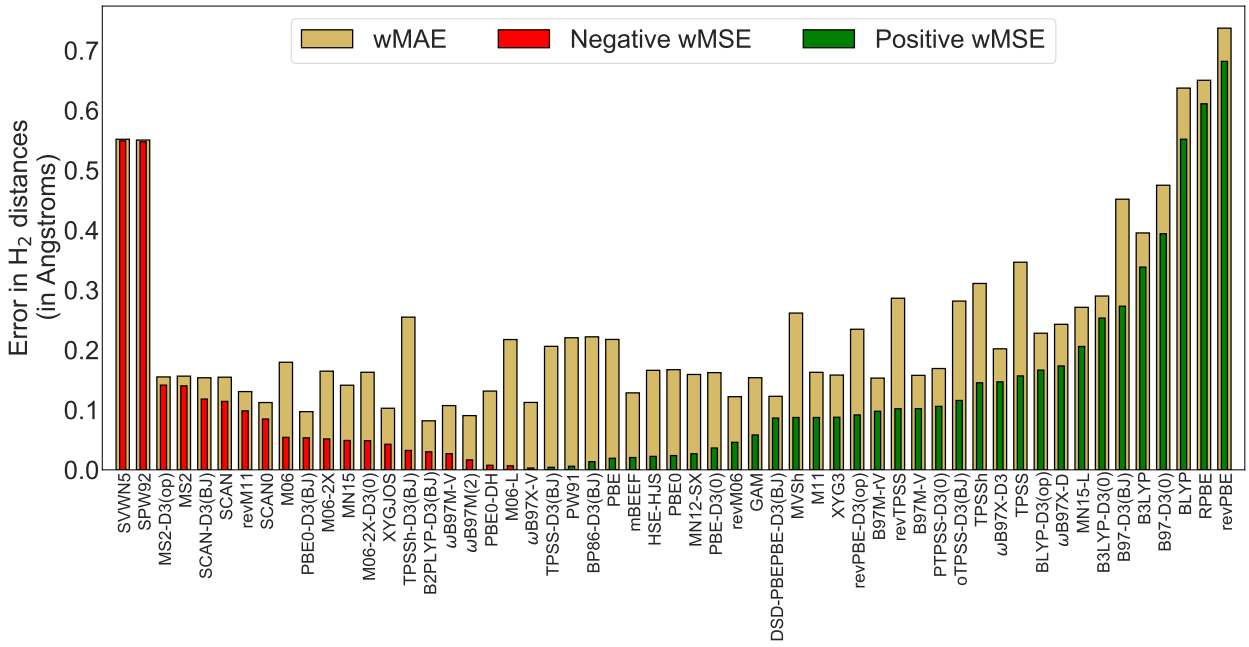


Figure 5.6: Weighted mean absolute error (wMAE) and weighted mean signed error (wMSE) of equilibrium H_2 distances predicted by different DFAs.

These species are ubiquitously found in porous materials capable of adsorbing H_2 (like MOFs) and form secondary binding sites for H_2. As this regime is not as important as the favorable one, it is assigned a weight of 4.0. This weighting scheme is used to form the weighted mean signed error (wMSE) and weighted mean unsigned error (wMAE) metrics in Fig. (5.5). The reference adiabatic interaction energy at equilibrium decides the weight of the corresponding PEC. As the vertical interaction energy does not lie on the adiabatic PEC, those data points were not included in the analyses in this section.

Most DFAs predict longer equilibrium binding motif – H_2 distances which is shown as a positive value of wMSE in Fig. 5.6. With the exception of the LDA density functionals, we see that all other density functionals which predict shorter equilibrium H_2 distances have a very small negative wMSE (> -0.15Å). The double hybrid density functional B2PLYP-D3(BJ) gives the best performance for predicting equilibrium geometry with a wMAE of 0.08Å. It is closely followed by the recently parametrized double hybrid ωB97M(2) density functional with a wMAE of 0.09Å. Both of these DFAs perform much better for equilibrium geometries in comparison to their performance for PECs. It is also rather surprising to note the less good performance of PBE0-DH (wMAE of 0.13Å), which is the best performing DFA for binding energies according to RMSE and RegMAPE. On the other hand, PBE0-D3(BJ) gives good geometries (ranked 3$^{\text{rd}}$ with wMAE of 0.10 Å) but its performance for PECs is mediocre (ranked 19$^{\text{th}}$ with a RegMAPE of 11.5%). However, other top performing DFAs in the energetics category like XYGJ-OS, ωB97M-V, and ωB97X-V also perform well for

geometries giving low wMAEs of 0.10 Å, 0.11 Å, and 0.11Å respectively. In particular, ωB97X-V shows no systematic error with virtually zero wMSE. It is also interesting to note the good performance of some semi-local density functionals like mBEEF and B97M-rV which give very low errors despite having no HF exchange. In light of these observations, and given the enhanced computational cost of double hybrid DFA nuclear gradients,[413] one can use hybrid DFAs like PBE0-D3(BJ) or ωB97M-V to perform a geometry optimization and then use the optimized geometry to perform a single point interaction energy calculation using a hybrid or double hybrid functional.

Another noticeable tread is the performance of DFAs upon the addition of some form of empirical dispersion correction. Addition of empirical dispersion corrections to DFAs reduces their errors for equilibrium H_2 distance prediction when the parent functional overestimates it. For example, the addition of the D3(op) correction to revPBE decreases its wMSE from 0.68Å to 0.09Å (concurrently decreasing wMAE from 0.74Å to 0.23Å). The performance of the commonly used density functional B3LYP and M06-2X is quite poor with large wMSE and wMAEs.

Typically, DFAs are used to optimize geometries of complexes containing an H_2 bound to a binding motif. After the geometry optimization has converged to a minimum on the potential energy surface, the binding energy of H_2 is computed as the difference between the energy of the complex at the minimum of the potential energy surface and energies of the binding motif and H_2 in isolation. Alternatively, DFAs can also be used in molecular dynamics and Monte Carlo simulations either directly[302] or indirectly (as reference energies for parametrizing force fields).[414–417] In these typical use cases, error in equilibrium binding energy can be attributed to two sources: (1) Inaccurate prediction of equilibrium geometry (2) Incorrect prediction of binding energy for the equilibrium geometry. Using DFAs for modeling H_2 binding materials, the DFA equilibrium geometry is typically used for computing the equilibrium binding energy.

We can assess the effect of relaxing the geometry along the PEC (defined by each DFA), the interaction coordinate of H_2 with the binding site. Geometry optimization along this coordinate can either improve or deteriorate the performance of density functionals. The performance of selected density functionals for predicting the minimum energy geometry on its PEC and the H_2 binding energy for this geometry assessed by the RegMAPE error metric is shown in Table 5.6 (the complete table showing the performance of all 55 DFAs is included in the Table D.4.

As both the reference and DFA geometries are relaxed along the potential energy curve, this error metric is adiabatic (ad) in nature as reflected in its subscript ($\text{RegMAPE}_{\text{ad}}$). The top five density functionals by the $\text{RegMAPE}_{\text{ad}}$ error metric are also the top five best performers according to their RegMAPE errors (Table 5.3), further emphasizing the superior performance of these DFAs for computing H_2 interaction energies. While the top five density functionals remain the same, it is interesting to note small changes in their order. ωB97X-V is the best performing DFA with a $\text{RegMAPE}_{\text{ad}}$ of 4.65% which is very closely followed by DSD-PBEPBE-D3(BJ) with a $\text{RegMAPE}_{\text{ad}}$ of 4.68%. Another noteworthy difference is the performance of the B97M-rV and the B97M-V functionals which are ranked sixth and

Table 5.6: Performance of density functional approximations (DFAs) for predicting H_2 binding energy at equilibrium geometry. The adiabatic regularized mean absolute percentage error ($RegMAPE_{ad}$) for 15 best performing DFAs and some commonly used DFAs are shown.

Rank	DFA	$RegMAPE_{ad}$
1	ωB97X-V	4.7
2	DSD-PBEPBE-D3(BJ)	4.7
3	PBE0-DH	5.3
4	ωB97M-V	6.0
5	XYGJ-OS	7.1
6	B97M-rV	7.2
7	B97M-V	7.2
8	ωB97M(2)	7.4
9	ωB97X-D	7.8
10	XYG3	7.9
11	B2PLYP-D3(BJ)	8.3
12	PBE0	8.4
13	HSE-HJS	8.4
14	ωB97X-D3	9.5
15	SCAN0	9.7
28	B3LYP	15.7
30	PBE	16.3
31	SCAN	16.4
33	revPBE-D3(op)	17.3
42	mBEEF	19.9
48	M06-2X	22.3
51	B97-D3(BJ)	24.4
52	GAM	26.5
53	B97-D3(0)	29.4

seventh with $\text{RegMAPE}_{\text{ad}}$ of 7.22% and 7.23% respectively. These DFAs were ranked 12^{th} and 13^{th} with RegMAPE of 9.0% and 9.1%. These functionals show a favorable cancellation of error in prediction of H_2 equilibrium binding energies when the equilibrium geometry is also optimized using the same functional. As these density functionals also do not have any HF exchange, they are computationally less expensive making them well-suited for applications in high-throughput material screening. The rVV10 non-local functional, which is an approximation[355] of the VV10 non-local functional, also allows for efficient evaluation in a plane wave framework and can be useful for modeling periodic systems like MOFs. A thorough assessment of the DFA geometry relaxed on the entire potential energy surface, not just along the one-dimensional PEC, is beyond the scope of this work and we refer interested readers to Ref. 409 for a detailed discussion of this topic.

5.4 Conclusions

The H2Bind275 dataset published recently[134] assesses the performance of density functionals for predicting the interaction energy of H_2 with different model binding motifs at the equilibrium geometry. In this work, we have assessed the ability of DFAs to predict H_2 interaction energies with binding motifs accurately throughout the PEC, not just at equilibrium geometry. To that end, we have extended our previous H2Bind275 dataset by adding two compressed and three elongated geometries along the PEC to form the H2Bind78×7 dataset. The H2Bind78×7 dataset comprises 545 data points at different fixed points along 78 PECs of various model binding motifs with H_2. Reference interaction energies for all data points were computed using CCSD(T) extrapolated to the complete basis set limit. The performance of 55 DFAs was assessed with the CCSD(T) reference interaction energies using multiple error metrics. The RMSE metric is democratic and gives equal important to all the 545 data points. The RegMAPE metric, on the other hand, gives more weight to binding motifs with interaction energies in the range of −15 to −25 kJ/mol at equilibrium geometry. For each binding motif, the RegMAPE metric is designed to give more weight to the equilibrium than non-equilibrium data points as the latter are encountered less often in modeling and simulation. DFAs are also assessed on the basis of their predicted equilibrium geometry and binding energy at predicted equilibrium geometry.

The CCSD(T) reference interaction energies for the H2Bind78×7 dataset span a wide range of attractive and repulsive interaction energies. As repulsive geometries are usually not included in non-covalent interaction energy datasets, the H2Bind78×7 dataset adds considerably to the diversity of the available datasets. The non-empirical double hybrid functional, PBE0-DH, shows the least error (RMSE of 2.9 kJ/mol and RegMAPE of 5.0%) in predicting H_2 binding energy throughout the PEC. The ωB97X-V, ωB97M-V, and DSD-PBEPBE-D3(BJ) density functionals are also top performers. The semi-local density functionals, B97M-V and B97M-rV, show poorer performance for the H2Bind78×7 dataset, in comparison to the previous H2Bind275 dataset using the RegMAPE error metric. For the H2Bind78×7 dataset, B97M-V and B97M-rV are ranked 13^{th} and 12^{th} respectively with

RegMAPEs of 9.1% and 9.0%. Previously in the H2Bind275 dataset, they were ranked 5th and 6th with RegMAPE of 6.8%. In general, the good performance of the top density functionals in the H2Bind275 dataset continues for the H2Bind78×7 dataset. Addition of DFT-D empirical dispersion correction increases the accuracy of underbinding density functionals like revPBE, BLYP, and B3LYP. This addition also decreases the accuracy of overbinding parent density functionals like PBE, PBE0, TPSS, SCAN, and MS2. As DFT-D empirical dispersion corrections are distance dependent, their effect is not felt uniformly across the PEC. The effect of addition of HF exchange, a short-ranged effect, improves the performance of density functionals in the compressed regime more than in the elongated regime of the PEC, thus playing a crucial role in accurately predicting the repulsive wall of the PEC. The weighted RegMAPE metric gives smaller weights to DFA errors in the compressed region of the PEC. This metric shows that, in general, DFAs perform better in the equilibrium and elongated regime than in the compressed region.

Assessment of DFAs for predicting equilibrium geometries reveals that PBE0-DH, which is the best performer for energies, shows less good performance for geometries. However, other hybrid functionals like ωB97M-V and ωB97X-V give good performance for both geometries and energies. Using the adiabatic RegMAPE metric (RegMAPE$_{ad}$) reveals that the semi-local DFAs, B97M-V and B97M-rV, show very small errors. They benefit significantly from cancellation between geometry-driven and energy-driven errors. ωB97M-V and ωB97X-V are the only density functionals that are not double hybrids which consistently show good performance for all the error metrics (energy and geometry-related) defined in this work. As these hybrid functionals have significantly lower computational cost in comparison to double hybrids, we recommend their usage for H_2 binding applications.

The H2Bind78×7 dataset, consisting of highly accurate reference interaction energies, represents a distinctive addition to other non-covalent interaction energy databases. More than half of this dataset consists of transition metal species which are usually underrepresented in non-covalent interaction energy datasets. This dataset is composed of complete PECs, rather than just PEC minimum geometries – only a handful of the non-covalent interaction energy datasets contain this information. Besides, almost all of the PECs in this dataset sample the repulsive wall. For these reasons, using the H2Bind78×7 dataset in training or validating DFAs can improve their performance and transferability. This work further validates the selection of best performing density functionals for the H2Bind275 dataset using a semi-independent dataset that is about two times larger. The definition and generalization of different error metrics (RegMAPE, wRegMAPE, and RegMAPE$_{ad}$) can be used for assessment of other similar datasets with well-defined schemes for weighting different data points. As force field parametrization requires good reference energies throughout the potential energy surface, the top performing density functionals in this work can be used for generating them.

Bibliography

(1) Nagy, B.; Jensen, F. Basis sets in quantum chemistry. *Rev. Comput. Chem.* **2017**, *30*, 93–149.

(2) Tosoni, S.; Tuma, C.; Sauer, J.; Civalleri, B.; Ugliengo, P. A comparison between plane wave and Gaussian-type orbital basis sets for hydrogen bonded systems: Formic acid as a test case. *J. Chem. Phys.* **2007**, *127*, 154102.

(3) Slater, J. C. The theory of complex spectra. *Phys. Rev.* **1929**, *34*, 1293.

(4) Condon, E. The theory of complex spectra. *Phys. Rev.* **1930**, *36*, 1121.

(5) Pulay, P. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(6) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556–560.

(7) Head-Gordon, M.; Pople, J. A. Optimization of wave function and geometry in the finite basis hartree-fock method. *J. Phys. Chem* **1988**, *92*, 3063–3069.

(8) Van Voorhis, T.; Head-gordon, M. A geometric approach to direct minimization. *Mol. Phys.* **2002**, *100*, 1713–1721.

(9) Szabo, A.; Ostlund, N. S., *Modern quantum chemistry: Introduction to advanced electronic structure theory*; Dover Publications: 1996.

(10) Echenique, P.; Alonso, J. L. A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry. *Mol. Phys.* **2007**, *105*, 3057–3098.

(11) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.

(12) Wang, Y. A.; Carter, E. A. In *Theoretical methods in condensed phase chemistry*; Springer: 2002, pp 117–184.

(13) Nagy, Á. Density functional. Theory and application to atoms and molecules. *Phys. Rep.* **1998**, *298*, 1–79.

(14) Capelle, K. A bird's-eye view of density-functional theory. *Braz. J. Phys.* **2006**, *36*, 1318–1343.

(15) Burke, K.; Wagner, L. O. DFT in a nutshell. *Int. J. Quantum Chem.* **2013**, *113*, 96–101.

(16) Parr, R.; Weitao, Y., *Density-Functional Theory of Atoms and Molecules*; International Series of Monographs on Chemistry; Oxford University Press: 1994.

(17) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

(18) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *8976*, 1–58.

(19) Chan, B.; Gill, P. M. W.; Kimura, M. Assessment of DFT Methods for Transition Metals with the TMC151 Compilation of Data Sets and Comparison with Accuracies for Main-Group Chemistry. *J. Chem. Theory Comput.* **2019**, *15*, 3622.

(20) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *J. Chem. Phys.* **2005**, *123*, 62201.

(21) Dirac, P. A. In *Math. Proc. Cambridge Philos. Soc.* 1930; Vol. 26, pp 376–385.

(22) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(23) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244.

(24) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(25) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

(26) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(27) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(28) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **2015**, *142*, 074111.

(29) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(30) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

(31) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618.

(32) Eshuis, H.; Bates, J. E.; Furche, F. Electron correlation methods based on the random phase approximation. *Theor. Chem. Acc.* **2012**, *131*, 1084.

(33) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4963–4968.

(34) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the ωB97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.

(35) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(36) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 6158.

(37) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(38) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.

(39) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

(40) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals density functional for general geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(41) Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B* **2010**, *82*, 081101.

(42) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional made simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(43) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103.

(44) Mardirossian, N.; Head-Gordon, M. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.

(45) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(46) Okoshi, M.; Atsumi, T.; Nakai, H. Revisiting the extrapolation of correlation energies to complete basis set limit. *J. Comput. Chem.* **2015**, *36*, 1075–1082.

(47) Grimme, S. Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel-and antiparallel-spin pair correlation energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(48) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order Møller–Plesset correlation energy: an economical electronic structure method. *J. Chem. Phys.* **2004**, *121*, 9793–9802.

(49) Lochan, R. C.; Head-Gordon, M. Orbital-optimized opposite-spin scaled second-order correlation: An economical method to improve the description of open-shell molecules. *J. Chem. Phys.* **2007**, *126*, 24110.

(50) Stück, D.; Head-Gordon, M. Regularized orbital-optimized second-order perturbation theory. *J. Chem. Phys.* **2013**, *139*, 244109.

(51) Lee, J.; Head-Gordon, M. Regularized Orbital-Optimized Second-Order Møller-Plesset Perturbation Theory: A Reliable Fifth-Order-Scaling Electron Correlation Model with Orbital Energy Dependent Regularizers. *J. Chem. Theory Comput.* **2018**, *14*, 5203–5219.

(52) Coester, F. Bound states of a many-particle system. *Nuclear Physics* **1958**, *7*, 421–424.

(53) Coester, F.; Kümmel, H. Short-range correlations in nuclear wave functions. *Nuclear Physics* **1960**, *17*, 477–485.

(54) Čızek, J. On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(55) Paldus, J.; Čızek, J.; Shavitt, I. Correlation Problems in Atomic and Molecular Systems. IV. Extended Coupled-Pair Many-Electron Theory and Its Application to the B H 3 Molecule. *Phys. Rev. A* **1972**, *5*, 50.

(56) Bartlett, R. J.; Purvis III, G. D. Molecular applications of coupled cluster and many-body perturbation methods. *Phys. Scr.* **1980**, *21*, 255.

(57) Purvis III, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

(58) Scuseria, G. E.; Janssen, C. L.; Schaefer Iii, H. F. An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *J. Chem. Phys.* **1988**, *89*, 7382–7387.

(59) Noga, J.; Bartlett, R. J. The full CCSDT model for molecular electronic structure. *J. Chem. Phys.* **1987**, *86*, 7041–7050.

(60) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth order comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(61) Crawford, T. D.; Schaefer, H. F. An introduction to coupled cluster theory for computational chemists. *Rev. Comput. Chem.* **2000**, *14*, 33–136.

(62) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.

(63) Shavitt, I.; Bartlett, R. J., *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*; Cambridge university press: 2009.

(64) Jensen, F., *Introduction to computational chemistry*; John wiley & sons: 2017.

(65) Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104–111.

(66) White, C. A.; Johnson, B. G.; Gill, P. M.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8–16.

(67) White, C. A.; Head-Gordon, M. A J matrix engine for density functional theory calculations. *J. Chem. Phys.* **1996**, *104*, 2620–2629.

(68) Füsti-Molnár, L.; Pulay, P. The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. *J. Chem. Phys.* **2002**, *117*, 7827–7835.

(69) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(70) Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M. Efficient implementation of the pair atomic resolution of the identity approximation for exact exchange for hybrid and range-separated density functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518–527.

(71) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 24113.

(72) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(73) Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38*, 869–882.

(74) Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis-Sets for Localized Molecular-Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169–178.

(75) Gianinetti, E.; Raimondi, M.; Tornaghi, E. Modification of the Roothaan equations to exclude BSSE from molecular interaction calculations. *Int. J. Quantum Chem.* **1996**, *60*, 157–166.

(76) Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism. *J. Chem. Phys.* **2001**, *115*, 3553–3560.

(77) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.

(78) Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism. *J. Chem. Phys* **2001**, *115*, 3553.

(79) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(80) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discuss.* **1997**, *106*, 79–92.

(81) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185.

(82) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

(83) Hu, L.; Söderhjelm, P.; Ryde, U. On the convergence of QM/MM energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

(84) Vreven, T.; Morokuma, K. Hybrid methods: Oniom (qm: mm) and qm/mm. *Annu. Rep. Comput. Chem.* **2006**, *2*, 35–51.

(85) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1–21.

(86) Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

(87) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(88) Fedorov, D. G.; Kitaura, K. The importance of three-body terms in the fragment molecular orbital method. *J. Chem. Phys.* **2004**, *120*, 6832–6840.

(89) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem A* **2007**, *111*, 6904–6914.

(90) Govind, N.; Wang, Y. A.; Carter, E. A. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *J. Chem. Phys.* **1999**, *110*, 7677–7688.

(91) Gomes, A. S. P.; Jacob, C. R.; Visscher, L. Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353–5362.

(92) Wesołowski, T. A. Embedding a multideterminantal wave function in an orbital-free environment. *Phys. Rev. A* **2008**, *77*, 012504.

(93) Khait, Y. G.; Hoffmann, M. R. Embedding theory for excited states. *J. Chem. Phys.* **2010**, *133*, 044107.

(94) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *J. Chem. Phys.* **2012**, *137*, 224113.

(95) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

(96) Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller, T. F. Embedded mean-field theory. *J. Chem. Theory Comput.* **2015**, *11*, 568–580.

(97) Ding, F.; Manby, F. R.; Miller, T. F. Embedded Mean-Field Theory with Block-Orthogonalized Partitioning. *J. Chem. Theory Comput.* **2017**, *13*, 1605–1615.

(98) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller-Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.

(99) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.

(100) Gao, J.; Wang, Y. Communication: Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method. *J. Chem. Phys.* **2012**, *136*, 71101.

(101) Mahmudov, K. T.; Arimitsu, S.; Saá, J. M.; Sarazin, Y.; Frontera, A.; Yamada, S.; Caminade, A.-M.; Chikkali, S. H.; Mal, P.; Momiyama, N., et al., *Noncovalent Interactions in Catalysis*; Royal Society of Chemistry: 2019.

(102) Proctor, R. S.; Colgan, A. C.; Phipps, R. J. Exploiting attractive non-covalent interactions for the enantioselective catalysis of reactions involving radical intermediates. *Nat. Chem.* **2020**, 1–15.

(103) Ma, S.; Zhou, H.-C. Gas storage in porous metal–organic frameworks for clean energy applications. *Chem. Commun.* **2010**, *46*, 44–53.

(104) Zhao, Z.; Ma, X.; Kasik, A.; Li, Z.; Lin, Y. Gas separation properties of metal organic framework (MOF-5) membranes. *Ind. Eng. Chem. Res.* **2013**, *52*, 1102–1108.

(105) Li, B.; Wen, H.-M.; Zhou, W.; Chen, B. Porous metal–organic frameworks for gas storage and separation: what, how, and why? *J. Phys. Chem. Lett.* **2014**, *5*, 3468–3479.

(106) Vogiatzis, K. D.; Klopper, W.; Friedrich, J. Non-covalent Interactions of CO2 with Functional Groups of Metal–Organic Frameworks from a CCSD(T) Scheme Applicable to Large Systems. *J. Chem. Theory Comput.* **2015**, *11*, 1574–1584.

(107) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. *J. Chem. Phys.* **2008**, *128*, 184112.

(108) McDaniel, J. G.; Schmidt, J. Next-generation force fields from symmetry-adapted perturbation theory. *Annu. Rev. Phys. Chem.* **2016**, *67*, 467–488.

(109) Das, A. K.; Urban, L.; Leven, I.; Loipersberger, M.; Aldossary, A.; Head-Gordon, M.; Head-Gordon, T. Development of an advanced force field for water using variational energy decomposition analysis. *J. Chem. Theory Comput.* **2019**, *15*, 5001–5013.

(110) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.

(111) Morokuma, K. Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity. *Acc. Chem. Res.* **1977**, *10*, 294–300.

(112) Ziegler, T.; Rauk, A. On the calculation of bonding energies by the Hartree Fock Slater method. *Theor. Chim. Acta* **1977**, *46*, 1–10.

(113) Ziegler, T.; Rauk, A. A theoretical study of the ethylene-metal bond in complexes between copper (1+), silver (1+), gold (1+), platinum (0) or platinum (2+) and ethylene, based on the Hartree-Fock-Slater transition-state method. *Inorg. Chem.* **1979**, *18*, 1558–1565.

(114) Bickelhaupt, F. M.; Baerends, E. J. Kohn-Sham density functional theory: predicting and understanding chemistry. *Rev. Comput. Chem.* **2000**, *15*, 1–86.

(115) Mitoraj, M. P.; Michalak, A.; Ziegler, T. A combined charge and energy decomposition scheme for bond analysis. *J. Chem. Theory Comput.* **2009**, *5*, 962–975.

(116) Glendening, E. D.; Streitwieser, A. Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding, strong ionic, and moderate donor–acceptor interactions. *J. Chem. Phys.* **1994**, *100*, 2900–2909.

(117) Glendening, E. D. Natural energy decomposition analysis: Extension to density functional methods and analysis of cooperative effects in water clusters. *J. Phys. Chem. A* **2005**, *109*, 11936–11940.

(118) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.

(119) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 254–272.

(120) Hopffgarten, M. v.; Frenking, G. Energy decomposition analysis. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 43–62.

(121) Phipps, M. J.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.

(122) Pastorczak, E.; Corminboeuf, C. Perspective: Found in translation: Quantum chemical tools for grasping non-covalent interactions. *J. Chem. Phys.* **2017**, *146*, 120901.

(123) Mo, Y.; Gao, J.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530–5538.

(124) Mo, Y.; Song, L.; Lin, Y. Block-localized wavefunction (BLW) method at the density functional theory (DFT) level. *J. Phys. Chem A* **2007**, *111*, 8291–8301.

(125) Mo, Y.; Bao, P.; Gao, J. Energy decomposition analysis based on a block-localized wavefunction and multistate density functional theory. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760–6775.

(126) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753–8765.

(127) Horn, P. R.; Head-Gordon, M. Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions. *J. Chem. Phys.* **2015**, *143*, 114111.

(128) Chen, Z.; Li, P.; Anderson, R.; Wang, X.; Zhang, X.; Robison, L.; Redfern, L. R.; Moribe, S.; Islamoglu, T.; Gómez-Gualdrón, D. A., et al. Balancing volumetric and gravimetric uptake in highly porous materials for clean energy. *Science* **2020**, *368*, 297–303.

(129) Ghoufi, A.; Malfreyt, P. Entropy and enthalpy calculations from perturbation and integration thermodynamics methods using molecular dynamics simulations: applications to the calculation of hydration and association thermodynamic properties. *Mol. Phys.* **2006**, *104*, 2929–2943.

(130) Garrone, E.; Bonelli, B.; Otero Areán, C. Enthalpy-entropy correlation for hydrogen adsorption on zeolites. *Chem. Phys. Lett.* **2008**, *456*, 68–70.

(131) Palomino, G. T.; Cabello, C. P.; Areán, C. O. Enthalpy-entropy correlation for hydrogen adsorption on MOFs: Variable-temperature FTIR study of hydrogen adsorption on MIL-100(Cr) and MIL-101(Cr). *Eur. J. Inorg. Chem.* **2011**, *100*, 1703–1708.

(132) Veccham, S. P.; Lee, J.; Head-Gordon, M. Making many-body interactions nearly pairwise additive: The polarized many-body expansion approach. *J. Chem. Phys.* **2019**, *151*, 194101.

(133) Veccham, S. P.; Lee, J.; Mao, Y.; Horn, P. R.; Head-Gordon, M. A Non-Perturbative Pairwise-Additive Analysis of Charge Transfer Contributions to Intermolecular Interaction Energies. *arXiv preprint arXiv:2011.04918* **2020**.

(134) Veccham, S. P.; Head-Gordon, M. Density Functionals for Hydrogen Storage: Defining the H2Bind275 Test Set with Ab Initio Benchmarks and Assessment of 55 Functionals. *J. Chem. Theory Comput.* **2020**, *16*, 4963–4982.

(135) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. K. L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, *345*, 640–643.

(136) Beran, G. J. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.

(137) Boese, A. D.; Sauer, J. Embedded and DFT Calculations on the Crystal Structures of Small Alkanes, Notably Propane. *Cryst. Growth Des.* **2017**, *17*, 1636–1646.

(138) Demerdash, O.; Yap, E.-H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149–174.

(139) Zgid, D.; Chan, G. K. L. Dynamical mean-field theory from a quantum chemical perspective. *J. Chem. Phys.* **2011**, *134*, 94115.

(140) Lan, T. N.; Zgid, D. Generalized Self-Energy Embedding Theory. *J. Phys. Chem. Lett.* **2017**, *8*, 2200–2205.

(141) Ishida, H. Surface-embedded Green-function method: A formulation using a linearized-augmented-plane-wave basis set. *Phys. Rev. B* **2001**, *63*, 165409.

(142) Chibani, W.; Ren, X.; Scheffler, M.; Rinke, P. Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept. *Phys. Rev. B* **2016**, *93*, 165106.

(143) Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 24501.

(144) Huang, C. Extending the density functional embedding theory to finite temperature and an efficient iterative method for solving for embedding potentials. *J. Chem. Phys.* **2016**, *144*, 124106.

(145) Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.

(146) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

(147) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891–5928.

(148) Knizia, G.; Chan, G. K. L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.

(149) Wouters, S.; Jiménez-Hoyos, C. A.; Sun, Q.; Chan, G. K. A Practical Guide to Density Matrix Embedding Theory in Quantum Chemistry. *J. Chem. Theory Comput.* **2016**, *12*, 2706–2719.

(150) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11635–11638.

(151) Vreven, T.; Morokuma, K. In *Annu. Rep. Comput. Chem.* C, 2006; Vol. 2; Chapter 3, pp 35–51.

(152) Guo, W.; Wu, A.; Xu, X. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203–208.

(153) Chung, L. W.; Sameera, W. M.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(154) Dahlke, E.; Truhlar, D. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(155) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.

(156) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for Benchmark Accuracy with the Many-Body Expansion. *Acc. Chem. Res.* **2014**, *47*, 2828–2836.

(157) Bygrave, P. J.; Allan, N. L.; Manby, F. R. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.* **2012**, *137*, 164102.

(158) Song, L.; Han, J.; Lin, Y. L.; Xie, W.; Gao, J. Explicit polarization (X-pol) potential using ab initio molecular orbital theory and density functional theory. *J. Phys. Chem. A* **2009**, *113*, 11656–11664.

(159) Gao, J. Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations. *J. Phys. Chem. B* **1997**, *101*, 657–663.

(160) Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346–2354.

(161) Xie, W.; Gao, J. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.

(162) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877–3886.

(163) Lao, K. U.; Herbert, J. M. Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.

(164) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. On the Interfragment Exchange in the X-Pol Method. *J. Chem. Theory Comput.* **2010**, *6*, 2469–2476.

(165) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(166) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.

(167) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.

(168) Fedorov, D. G.; Kitaura, K. In *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier Science: Amsterdam, 2006; Chapter 1, 3–38.

(169) Cullen, J. M. An examination of the effects of basis set and charge transfer in hydrogen-bonded dimers with a constrained Hartree–Fock method. *Int. J. Quantum Chem.* **1991**, *40*, 193–207.

(170) Horn, P. R.; Sundstrom, E. J.; Baker, T. A.; Head-Gordon, M. Unrestricted absolutely localized molecular orbitals for energy decomposition analysis: Theory and applications to intermolecular interactions involving radicals. *J. Chem. Phys.* **2013**, *138*, 30–32.

(171) Thirman, J.; Head-Gordon, M. An energy decomposition analysis for second-order Møller–Plesset perturbation theory based on absolutely localized molecular orbitals. *J. Chem. Phys.* **2015**, *143*, 084124.

(172) Gao, J.; Cembran, A.; Mo, Y. Generalized X-Pol theory and charge delocalization states. *J. Chem. Theory Comput.* **2010**, *6*, 2402–2410.

(173) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(174) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(175) Gill, P. M.; Johnson, B. G.; Pople, J. A. A standard grid for density functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.

(176) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.

(177) Lee, J.; Head-Gordon, M. Distinguishing artificial and essential symmetry breaking in a single determinant: Approach and application to the C60 , C36 , and C20 fullerenes. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4763–4778.

(178) Lee, J.; Head-Gordon, M. Two single-reference approaches to singlet biradicaloid problems: Complex, restricted orbitals and approximate spin-projection combined with regularized orbital-optimized Møller-Plesset perturbation theory. *J. Chem. Phys.* **2019**, *150*, 244106.

(179) Bertels, L. W.; Lee, J.; Head-Gordon, M. Third-Order Møller–Plesset Perturbation Theory Made Useful? Choice of Orbitals and Scaling Greatly Improves Accuracy for Thermochemistry, Kinetics, and Intermolecular Interactions. *J. Phys. Chem. Lett.* **2019**, *10*, 4170–4176.

(180) Mata, R. A.; Werner, H. J. Calculation of smooth potential energy surfaces using local electron correlation methods. *J. Chem. Phys.* **2006**, *125*, 184110.

(181) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. The limits of local correlation theory: Electronic delocalization and chemically smooth potential energy surfaces. *J. Chem. Phys.* **2008**, *128*, 34103.

(182) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.

(183) Sum, A.; Sandler, S. Ab Initio Calculations of Cooperativity Effects on Clusters of Methanol, Ethanol, 1-Propanol, and Methanethiol. *J. Phys. Chem. A* **2000**, *104*, 1121–1129.

(184) Rincón, L.; Almeida, R.; García-Aldea, D.; Diez Y Riega, H. Hydrogen bond cooperativity and electron delocalization in hydrogen fluoride clusters. *J. Chem. Phys.* **2001**, *114*, 5552–5561.

(185) Luck, W. A. The importance of cooperativity for the properties of liquid water. *J. Mol. Struct.* **1998**, *448*, 131–142.

(186) Yuan, D.; Li, Y.; Ni, Z.; Pulay, P.; Li, W.; Li, S. Benchmark Relative Energies for Large Water Clusters with the Generalized Energy-Based Fragmentation Method. *J. Chem. Theory Comput.* **2017**, *13*, 2696–2704.

(187) Temelso, B.; Renner, C. R.; Shields, G. C. Importance and reliability of small basis set CCSD(T) corrections to MP2 binding and relative energies of water clusters. *J. Chem. Theory Comput.* **2015**, *11*, 1439–1448.

(188) Friedrich, J.; Fiedler, B. Accurate calculation of binding energies for molecular clusters - Assessment of different models. *Chem. Phys.* **2016**, *472*, 72–80.

(189) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

(190) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(191) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(192) Tainter, C. J.; Pieniazek, P. A.; Lin, Y. .-S.; Skinner, J. L. Robust three-body water simulation model. *J. Chem. Phys.* **2011**, *134*, 234505.

(193) Babin, V.; Leforestier, C.; Paesani, F. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.

(194) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501–7528.

(195) Medders, G. R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through "First-Principles" Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145–1154.

(196) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.

(197) Dang, L. X.; Chang, T.-M. Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. *J. Chem. Phys.* **1997**, *106*, 8149–8159.

(198) Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n=2-21) and the phonon spectrum of ice Ih. *J. Chem. Phys.* **1999**, *110*, 4566–4581.

(199) Wang, C. C.; Zielke, P.; Sigurbjörnsson, Ó. F.; Viteri, C. R.; Signorell, R. Infrared Spectra of C2H6 , C2H4 , C2H2 , and CO2 Aerosols Potentially Formed in Titan's Atmosphere. *J. Phys. Chem. A* **2009**, *113*, 11129–11137.

(200) DeSimone, J. M. Practical approaches to green solvents. *Science* **2002**, *297*, 799–803.

(201) Bukowski, R.; Sadlej, J.; Jeziorski, B.; Jankowski, P.; Szalewicz, K.; Kucharski, S. A.; Williams, H. L.; Rice, B. M. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *J. Chem. Phys.* **1999**, *110*, 3785–3803.

(202) Horn, P. R.; Mao, Y.; Head-Gordon, M. Defining the contributions of permanent electrostatics, Pauli repulsion, and dispersion in density functional theory calculations of intermolecular interaction energies. *J. Chem. Phys.* **2016**, *144*, 114107.

(203) The CCSD(T) interaction energy reported in this work have been extrapolated to the Complete Basis Set (CBS) limit using the focal point approximation analysis. $E_{CCSD(T)/CBS} \approx E^{SCF}_{HF/aug-cc-pV5Z} + E^{corr}_{MP2/aug-cc-pVQZ \rightarrow aug-cc-pV5Z} + (E^{corr}_{CCSD(T)/aug-cc-pVTZ} - E^{corr}_{MP2/aug-cc-pVTZ})$. The MP2 correlation energy was extrapolated to uses the 2-point extrapolation formula [45] using the the aug-cc-pVQZ and the aug-cc-pV5Z basis sets.

(204) Lemke, K. H.; Seward, T. M. Thermodynamic properties of carbon dioxide clusters by M06-2X and dispersion-corrected B2PLYP-D theory. *Chem. Phys. Lett.* **2013**, *573*, 19–23.

(205) Jordan, K. D.; Sen, K. In *Chemical Modelling: Volume 13*; The Royal Society of Chemistry: 2017; Vol. 13, pp 105–131.

(206) Ludwig, R. Water: From Clusters to the Bulk. *Angew. Chem., Int. Ed.* **2001**, *40*, 1808–1827.

(207) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. The water dipole moment in water clusters. *Science* **1997**, *275*, 814–817.

(208) Badyal, Y. S.; Saboungi, M.-L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. Electron Distribution in Water. *J. Chem. Phys.* **2000**, *112*, 9206.

(209) Batista, E. R.; Xantheas, S. S.; Jonsson, H. Molecular multipole moments of water molecules in ice Ih. *J. Chem. Phys.* **1998**, *109*, 4546–4551.

(210) Silvestrelli, P. L.; Parrinello, M. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **1999**, *111*, 3572–3580.

(211) Ghanty, T. K.; Ghosh, S. K. Polarizability of water clusters: An ab initio investigation. *J. Chem. Phys.* **2003**, *118*, 8547–8550.

(212) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Electron donation in the water-water hydrogen bond. *Chem. - Eur. J.* **2009**, *15*, 851–855.

(213) Cobar, E. A.; Horn, P. R.; Bergman, R. G.; Head-Gordon, M. Examination of the hydrogen-bonding networks in small water clusters (n = 2-5, 13, 17) using absolutely localized molecular orbital energy decomposition analysis. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15328–15339.

(214) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. Structure and stability of water clusters (H2O)n, n = 8-20: An ab initio investigation. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.

(215) González, B. S.; Hernández-Rojas, J.; Wales, D. J. Global minima and energetics of Li+(H2O)n and Ca2+(H2O)n clusters for n 20. *Chem. Phys. Lett.* **2005**, *412*, 23–28.

(216) Egan, C. K.; Paesani, F. Assessing Many-Body Effects of Water Self-Ions. I: OH – (H 2 O) n Clusters. *J. Chem. Theory Comput.* **2018**, *14*, 1982–1997.

(217) Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.* **2014**, *141*, 014108.

(218) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(219) Becke, A. Density Functional Thermochemistry III The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(220) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

(221) ErgoSCF, http://www.ergoscf.org/ [Accessed: 13 May 2019], 2019.

(222) Almlof, J.; Faegri, K.; Korsell, K. Principles for a direct SCF approach to LCAO-MO ab-initio calculations. *J. Comput. Chem.* **1982**, *3*, 385–399.

(223) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663–1669.

(224) Hu, W.; Lin, L.; Yang, C. DGDFT: A massively parallel method for large scale density functional theory calculations. *J. Chem. Phys.* **2015**, *143*, 124110.

(225) Chen, W.; Gordon, M. S. Energy decomposition analyses for many-body interaction and applications to water complexes. *J. Phys. Chem.* **1996**, *100*, 14316–14328.

(226) Su, P.; Li, H. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.* **2009**, *131*, 014102.

(227) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899–926.

(228) Mao, Y.; Ge, Q.; Horn, P. R.; Head-Gordon, M. On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2401–2417.

(229) Dewar, J. A review of the pi-complex theory. *Bull. Soc. Chim. Fr.* **1951**, *18*, C71–C79.

(230) Chatt, J.; Duncanson, L. Olefin co-ordination compounds. Part III. Infra-red spectra and structure: attempted preparation of acetylene complexes. *J. Chem. Soc.* **1953**, 2939–2947.

(231) Van der Lubbe, S. C.; Fonseca Guerra, C. The nature of hydrogen bonds: A delineation of the role of different energy components on hydrogen bond strengths and lengths. *Chem. - Asian J.* **2019**, *14*, 2760–2769.

(232) Horn, P. R.; Mao, Y.; Head-Gordon, M. Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067–23079.

(233) Mao, Y.; Levine, D. S.; Loipersberger, M.; Horn, P. R.; Head-Gordon, M. Probing radical–molecule interactions with a second generation energy decomposition analysis of DFT calculations using absolutely localized molecular orbitals. *Phys. Chem. Chem. Phys.* **2020**, *22*, 12867–12885.

(234) Thirman, J.; Head-Gordon, M. Efficient Implementation of Energy Decomposition Analysis for Second-Order Møller–Plesset Perturbation Theory and Application to Anion–π Interactions. *J. Phys. Chem. A* **2017**, *121*, 717–728.

(235) Loipersberger, M.; Lee, J.; Mao, Y.; Das, A. K.; Ikeda, K.; Thirman, J.; Head-Gordon, T.; Head-Gordon, M. Energy Decomposition Analysis for Interactions of Radicals: Theory and Implementation at the MP2 Level with Application to Hydration of Halogenated Benzene Cations and Complexes between $CO_2^{-\bullet}$ and Pyridine and Imidazole. *J. Phys. Chem. A* **2019**, *123*, 9621–9633.

(236) Dapprich, S.; Frenking, G. Investigation of donor-acceptor interactions: a charge decomposition analysis using fragment molecular orbitals. *J. Phys. Chem.* **1995**, *99*, 9352–9362.

(237) Stevens, W. J.; Fink, W. H. Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem. Phys. Lett.* **1987**, *139*, 15–22.

(238) Bagus, P. S.; Hermann, K.; Bauschlicher Jr, C. W. A new analysis of charge transfer and polarization for ligand–metal bonding: Model studies of Al4CO and Al4NH3. *J. Chem. Phys.* **1984**, *80*, 4378–4386.

(239) Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy decomposition analysis. *WIREs: Comput. Mol. Sci.* **2018**, *8*, e1345.

(240) Stone, A. J.; Misquitta, A. J. Charge-transfer in symmetry-adapted perturbation theory. *Chem. Phys. Lett.* **2009**, *473*, 201–205.

(241) Misquitta, A. J. Charge transfer from regularized symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2013**, *9*, 5313–5326.

(242) Reed, A. E.; Weinhold, F. Natural bond orbital analysis of near-Hartree–Fock water dimer. *J. Chem. Phys.* **1983**, *78*, 4066–4073.

(243) Stone, A. J. Natural bond orbitals and the nature of the hydrogen bond. *J. Phys. Chem. A* **2017**, *121*, 1531–1534.

(244) Wu, Q.; Van Voorhis, T. Direct optimization method to study constrained systems within density-functional theory. *Phys. Rev. A* **2005**, *72*, 024502.

(245) Wu, Q.; Van Voorhis, T. Constrained density functional theory and its application in long-range electron transfer. *J. Chem. Theory Comput.* **2006**, *2*, 765–774.

(246) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained density functional theory. *Chem. Rev.* **2012**, *112*, 321–370.

(247) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.

(248) Lao, K. U.; Herbert, J. M. Energy Decomposition Analysis with a Stable Charge-Transfer Term for Interpreting Intermolecular Interactions. *J. Chem. Theory Comput.* **2016**, *12*, 2569–2582.

(249) Azar, R. J.; Horn, P. R.; Sundstrom, E. J.; Head-Gordon, M. Useful lower limits to polarization contributions to intermolecular interactions using a minimal basis of localized orthogonal orbitals: Theory and analysis of the water dimer. *J. Chem. Phys.* **2013**, *138*, 084102.

(250) Mao, Y.; Horn, P. R.; Head-Gordon, M. Energy decomposition analysis in an adiabatic picture. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5944–5958.

(251) Mao, Y.; Head-Gordon, M. Probing Blue-Shifting Hydrogen Bonds with Adiabatic Energy Decomposition Analysis. *J. Phys. Chem. Lett.* **2019**, *10*, 3899–3905.

(252) Loipersberger, M.; Zee, D. Z.; Panetier, J. A.; Chang, C. J.; Long, J. R.; Head-Gordon, M. Computational Study of an Iron(II) Polypyridine Electrocatalyst for CO2 Reduction: Key Roles for Intramolecular Interactions in CO2 Binding and Proton Transfer. *Inorg. Chem.* **2020**, *59*, 8146–8160.

(253) Loipersberger, M.; Mao, Y.; Head-Gordon, M. Variational Forward–Backward Charge Transfer Analysis Based on Absolutely Localized Molecular Orbitals: Energetics and Molecular Properties. *J. Chem. Theory Comput.* **2020**, *16*, 1073–1089.

(254) Liang, W.; Head-Gordon, M. An exact reformulation of the diagonalization step in electronic structure calculations as a set of second order nonlinear equations. *J. Chem. Phys.* **2004**, *120*, 10379–10384.

(255) Liang, W.; Head-Gordon, M. Approaching the basis set limit in density functional theory calculations using dual basis sets without diagonalization. *J. Phys. Chem. A* **2004**, *108*, 3206–3210.

(256) Abramowitz, M., *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,* Dover Publications, Inc.: USA, 1974.

(257) Head-Gordon, M.; Maslen, P. E.; White, C. A. A tensor formulation of many-electron theory in a nonorthogonal single-particle basis. *J. Chem. Phys.* **1998**, *108*, 616–625.

(258) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(259) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(260) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(261) Mao, Y.; Horn, P. R.; Mardirossian, N.; Head-Gordon, T.; Skylaris, C.-K.; Head-Gordon, M. Approaching the basis set limit for DFT calculations using an environment-adapted minimal basis with perturbation theory: Formulation, proof of concept, and a pilot implementation. *J. Chem. Phys.* **2016**, *145*, 044109.

(262) Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **2008**, *47*, 5336–5353.

(263) Lippard, S. J.; Berg, J. M., *Principles of bioinorganic chemistry*; University Science Books Mill Valley, CA: 1994; Vol. 70.

(264) Kaim, W.; Schwederski, B.; Klein, A., *Bioinorganic Chemistry–Inorganic Elements in the Chemistry of Life: An Introduction and Guide*; John Wiley & Sons: 2013.

(265) Sigel, H. Interactions of metal ions with nucleotides and nucleic acids and their constituents. *Chem. Soc. Rev.* **1993**, *22*, 255–267.

(266) Lippert, B. Multiplicity of metal ion binding patterns to nucleobases. *Coord. Chem. Rev.* **2000**, *200*, 487–516.

(267) Valls, N.; Usón, I.; Gouyette, C.; Subirana, J. A. A cubic arrangement of DNA double helices based on nickel- guanine interactions. *J. Am. Chem. Soc.* **2004**, *126*, 7812–7816.

(268) Stasyuk, O. A.; Solà, M.; Swart, M.; Fonseca Guerra, C.; Krygowski, T. M.; Szatylowicz, H. Effect of alkali metal cations on length and strength of hydrogen bonds in DNA base pairs. *ChemPhysChem* **2020**, *21*, 2112–2126.

(269) Poater, J.; Sodupe, M.; Bertran, J.; Solà, M. Hydrogen bonding and aromaticity in the guanine–cytosine base pair interacting with metal cations (M= Cu+, Ca2+ and Cu2+). *Mol. Phys.* **2005**, *103*, 163–173.

(270) Marder, T. B. Will we soon be fueling our automobiles with ammonia–borane? *Angew. Chem., Int. Ed.* **2007**, *46*, 8116–8118.

(271) Peng, B.; Chen, J. Ammonia borane as an efficient and lightweight hydrogen storage medium. *Energy Environ. Sci.* **2008**, *1*, 479–483.

(272) Hirao, H.; Omoto, K.; Fujimoto, H. Lewis acidity of boron trihalides. *J. Phys. Chem. A* **1999**, *103*, 5807–5811.

(273) Plumley, J. A.; Evanseck, J. D. Periodic trends and index of boron Lewis acidity. *J. Phys. Chem. A* **2009**, *113*, 5985–5992.

(274) Branchadell, V.; Oliva, A. The Lewis acidity scale of boron trihalides: an ab initio study. *J. Mol. Struct.: THEOCHEM* **1991**, *236*, 75–84.

(275) Bessac, F.; Frenking, G. Why is BCl3 a stronger Lewis acid with respect to strong bases than BF3? *Inorg. Chem.* **2003**, *42*, 7990–7994.

(276) Staubitz, A.; Robertson, A. P.; Sloan, M. E.; Manners, I. Amine- and phosphine-borane adducts: new interest in old molecules. *Chem. Rev.* **2010**, *110*, 4023–4078.

(277) Xu, K.; Peng, H.; Lam, J. W. Y.; Poon, T. W. H.; Dong, Y.; Xu, H.; Sun, Q.; Cheuk, K. K. L.; Salhi, F.; Lee, P. P. S., et al. Transition metal carbonyl catalysts for polymerizations of substituted acetylenes. *Macromolecules* **2000**, *33*, 6918–6924.

(278) Sunley, G. J.; Watson, D. J. High productivity methanol carbonylation catalysis using iridium: the Cativa™ process for the manufacture of acetic acid. *Catal. Today* **2000**, *58*, 293–307.

(279) Rossomme, E.; Lininger, C. N.; Bell, A. T.; Head-Gordon, T.; Head-Gordon, M. Electronic structure calculations permit identification of the driving forces behind frequency shifts in transition metal monocarbonyls. *Phys. Chem. Chem. Phys.* **2020**, *22*, 781–798.

(280) Huber, K.-P., *Molecular spectra and molecular structure: IV. Constants of diatomic molecules*; Springer Science & Business Media: 2013.

(281) Abel, E.; McLean, R.; Tyfield, S.; Braterman, P.; Walker, A.; Hendra, P. Vibrational and electronic spectra and bonding in ionic transition metal hexacarbonyls. *J. Mol. Spectrosc.* **1969**, *30*, 29–50.

(282) Reed, Z. D.; Duncan, M. A. Infrared spectroscopy and structures of manganese carbonyl cations, Mn (CO) n+(n= 1- 9). *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 739–749.

(283) Allendorf, M. D. et al. An assessment of strategies for the development of solid-state adsorbents for vehicular hydrogen storage. *Energy Environ. Sci.* **2018**, *11*, 2784–2812.

(284) Chen, B.; Eddaoudi, M.; Reineke, T. M.; Kampf, J. W.; O'Keeffe, M.; Yaghi, O. M. Cu2(ATC)·6H2O: Design of Open Metal Sites in Porous MetalOrganic Crystals (ATC: 1,3,5,7-Adamantane Tetracarboxylate). *J. Am. Chem. Soc.* **2000**, *122*, 11559–11560.

(285) Kapelewski, M. T.; Geier, S. J.; Hudson, M. R.; Stück, D.; Mason, J. A.; Nelson, J. N.; Xiao, D. J.; Hulvey, Z.; Gilmour, E.; FitzGerald, S. A.; Head-Gordon, M.; Brown, C. M.; Long, J. R. M2(m-dobdc) (M = Mg, Mn, Fe, Co, Ni) Metal–Organic Frameworks Exhibiting Increased Charge Density and Enhanced H2 Binding at the Open Metal Sites. *J. Am. Chem. Soc.* **2014**, *136*, 12119–12129.

(286) Dincă, M.; Han, W. S.; Liu, Y.; Dailly, A.; Brown, C. M.; Long, J. R. Observation of Cu2+–H2 Interactions in a Fully Desolvated Sodalite-Type Metal–Organic Framework. *Angew. Chem., Int. Ed.* **2007**, *46*, 1419–1422.

(287) Jackson, K. T.; Reich, T. E.; El-Kaderi, H. M. Targeted synthesis of a porous borazine-linked covalent organic framework. *Chem. Commun.* **2012**, *48*, 8823–8825.

(288) Zhu, S.; Li, T. Hydrogenation-assisted graphene origami and its application in programmable molecular mass uptake, storage, and release. *ACS Nano* **2014**, *8*, 2864–2872.

(289) Kim, T. H.; Bae, J.; Lee, T. H.; Hwang, J.; Jung, J. H.; Kim, D. K.; Lee, J. S.; Kim, D. O.; Lee, Y. H.; Ihm, J. Room-temperature hydrogen storage via two-dimensional potential well in mesoporous graphene oxide. *Nano Energy* **2016**, *27*, 402–411.

(290) Burress, J. W.; Gadipelli, S.; Ford, J.; Simmons, J. M.; Zhou, W.; Yildirim, T. Graphene oxide framework materials: theoretical predictions and experimental results. *Angew. Chem., Int. Ed.* **2010**, *49*, 8902–8904.

(291) Teichmann, D.; Arlt, W.; Wasserscheid, P.; Freymann, R. A future energy supply based on Liquid Organic Hydrogen Carriers (LOHC). *Energy Environ. Sci.* **2011**, *4*, 2767–2773.

(292) Areán, C. O.; Chavan, S.; Cabello, C. P.; Garrone, E.; Palomino, G. T. Thermodynamics of Hydrogen Adsorption on Metal-Organic Frameworks. *ChemPhysChem* **2010**, *11*, 3237–3242.

(293) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038–5071.

(294) Tsivion, E.; Mason, J. A.; Gonzalez, M. I.; Long, J. R.; Head-Gordon, M. A computational study of CH4 storage in porous framework materials with metalated linkers: connecting the atomistic character of CH4 binding sites to usable capacity. *Chem. Sci.* **2016**, *7*, 4503–4518.

(295) Tsivion, E.; Veccham, S. P.; Head-Gordon, M. High-Temperature Hydrogen Storage of Multiple Molecules: Theoretical Insights from Metalated Catechols. *ChemPhysChem* **2017**, *18*, 184–188.

(296) Colón, Y. J.; Snurr, R. Q. High-throughput computational screening of metal–organic frameworks. *Chem. Soc. Rev.* **2014**, *43*, 5735–5749.

(297) Zong, S.; Zhang, Y.; Lu, N.; Ma, P.; Wang, J.; Shi, X.-R. A DFT Screening of M-HKUST-1 MOFs for Nitrogen-Containing Compounds Adsorption. *Nanomaterials* **2018**, *8*, 958.

(298) Campbell, C.; Ferreiro-Rangel, C. A.; Fischer, M.; Gomes, J. R.; Jorge, M. A transferable model for adsorption in MOFs with unsaturated metal sites. *J. Phys. Chem. C* **2017**, *121*, 441–458.

(299) Lee, K.; Howe, J. D.; Lin, L.-C.; Smit, B.; Neaton, J. B. Small-molecule adsorption in open-site metal–organic frameworks: a systematic density functional theory study for rational design. *Chem. Mater.* **2015**, *27*, 668–678.

(300) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.

(301) Hait, D.; Head-Gordon, M. How accurate are static polarizability predictions from density functional theory? An assessment over 132 species at equilibrium geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800–19810.

(302) Fetisov, E. O.; Shah, M. S.; Long, J. R.; Tsapatsis, M.; Siepmann, J. I. First principles Monte Carlo simulations of unary and binary adsorption: CO 2, N 2, and H 2 O in Mg-MOF-74. *Chem. Commun.* **2018**, *54*, 10816–10819.

(303) Sumida, K.; Stück, D.; Mino, L.; Chai, J.-D.; Bloch, E. D.; Zavorotynska, O.; Murray, L. J.; Dincă, M.; Chavan, S.; Bordiga, S.; Head-Gordon, M.; Long, J. R. Impact of Metal and Anion Substitutions on the Hydrogen Storage Properties of M-BTT Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2013**, *135*, 1083–1091.

(304) Tsivion, E.; Long, J. R.; Head-Gordon, M. Hydrogen physisorption on metal-organic framework linkers and metalated linkers: A computational study of the factors that control binding strength. *J. Am. Chem. Soc.* **2014**, *136*, 17827–17835.

(305) Kállay, M.; Surján, P. R. Higher excitations in coupled-cluster theory. *J. Chem. Phys.* **2001**, *115*, 2945–2954.

(306) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.

(307) Kállay, M.; Rolik, Z.; Csontos, J.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Samu, G.; Petrov, K.; Farkas, M.; Nagy, P., et al. MRCC, a quantum chemical program suite. *URL: http://www. mrcc. hu, accessed August 26th* **2016**.

(308) Hobza, P.; Šponer, J. Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chem. Rev.* **1999**, *99*, 3247–3276.

(309) Řezáč, J.; Hobza, P. Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard," CCSD(T) at the complete basis set limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.

(310) Kubas, G. J. Metal-dihydrogen and σ-bond coordination: The consummate extension of the Dewar-Chatt-Duncanson model for metal-olefin π bonding. *J. Organomet. Chem.* **2001**, *635*, 37–68.

(311) Kubas, G. J. Fundamentals of H 2 binding and reactivity on transition metals underlying hydrogenase function and H 2 production and storage. *Chem. Rev.* **2007**, *107*, 4152–4205.

(312) Saillard, J. Y.; Hoffmann, R. Carbon-hydrogen and hydrogen-hydrogen activation in transition metal complexes and on surfaces. *J. Am. Chem. Soc* **1984**, *106*, 2006–2026.

(313) Runčevski, T.; Kapelewski, M. T.; Torres-Gavosto, R. M.; Tarver, J. D.; Brown, C. M.; Long, J. R. Adsorption of two gas molecules at a single metal site in a metal–organic framework. *Chem. Commun.* **2016**, *2*, 8251–8254.

(314) Getman, R. B.; Miller, J. H.; Wang, K.; Snurr, R. Q. Metal alkoxide functionalization in metal-organic frameworks for enhanced ambient-temperature hydrogen storage. *J. Phys. Chem. C* **2011**, *115*, 2066–2075.

(315) Getman, R. B.; Miller, J. H.; Wang, K.; Snurr, R. Q. Metal Alkoxide Functionalization in Metal-Organic Frameworks for Enhanced Ambient-Temperature Hydrogen Storage. *J. Phys. Chem. C* **2011**, *115*, 2066–2075.

(316) Kemper, P. R.; Weis, P.; Bowers, M. T.; Maître, P. Origin of bonding interactions in Cu+(H2)(n) clusters: An experimental and theoretical investigation. *J. Am. Chem. Soc.* **1998**, *120*, 13494–13502.

(317) Plitt, H. S.; Bär, M. R.; Ahlrichs, R.; Schnöckel, H. [Cu(2-H2)Cl], a Model Compound for H2 Complexes. Ab Initio Calculations and Identification by IR Spectroscopy. *Angew. Chem., Int. Ed.* **1991**, *30*, 832–834.

(318) Frohman, D. J.; Grubbs, G. S.; Yu, Z.; Novick, S. E. Probing the chemical nature of dihydrogen complexation to transition metals, a gas phase case study: H2-CuF. *Inorg. Chem.* **2013**, *52*, 816–822.

(319) Grubbs, G. S.; Obenchain, D. A.; Pickett, H. M.; Novick, S. E. H2—AgCl: A spectroscopic study of a dihydrogen complex. *J. Chem. Phys.* **2014**, *141*, 114306.

(320) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.

(321) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.

(322) Curtiss, L. A.; Raghavach Ari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* **1998**, *109*, 7764–7776.

(323) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108.

(324) Martin, J. M.; De Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry - W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.

(325) Daniel Boese, A.; Oren, M.; Atasoylu, O.; Martin, J. M.; Kállay, M.; Gauss, J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. *J. Chem. Phys.* **2004**, *120*, 4129–4141.

(326) Karton, A.; Rabinovich, E.; Martin, J. M.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 11599.

(327) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599–11613.

(328) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 84108.

(329) East, A. L. L.; Allen, W. D. The heat of formation of NCO. *J. Chem. Phys.* **1993**, *99*, 4638–4650.

(330) Császár, A. G.; Allen, W. D.; Schaefer, H. F. In pursuit of the ab initio limit for conformational energy prototypes. *J. Chem. Phys.* **1998**, *108*, 9751–9764.

(331) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(332) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(333) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

(334) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.

(335) Balabanov, N. B.; Peterson, K. A. Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3d elements Sc-Zn. *J. Chem. Phys.* **2005**, *123*, 64107.

(336) Jensen, F. Atomic orbital basis sets. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 273–295.

(337) Pulay, P.; Hamilton, T. P. UHF natural orbitals for defining and starting MC-SCF calculations. *J. Chem. Phys.* **1988**, *88*, 4926–4933.

(338) Bofill, J. M.; Pulay, P. The unrestricted natural orbital-complete active space (UNO-CAS) method: An inexpensive alternative to the complete active space-self-consistent-field (CAS-SCF) method. *J. Chem. Phys.* **1989**, *90*, 3637–3646.

(339) Bhatia, S. K.; Myers, A. L. Optimum conditions for adsorptive storage. *Langmuir* **2006**, *22*, 1688–1700.

(340) Bae, Y. S.; Snurr, R. Q. Optimal isosteric heat of adsorption for hydrogen storage and delivery using metal-organic frameworks. *Microporous Mesoporous Mater.* **2010**, *132*, 300–303.

(341) Cybulski, S. M.; Lytle, M. L. The origin of deficiency of the supermolecule second-order Møller-Plesset approach for evaluating interaction energies. *J. Chem. Phys.* **2007**, *127*, 141102.

(342) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(343) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **1999**, *59*, 7413.

(344) Zhang, Y.; Yang, W. Comment on "Generalized gradient approximation made simple". *Phys. Rev. Lett.* **1998**, *80*, 890.

(345) Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 Damping Functions Across Widely Used Density Functionals: Can We Do Better? *J. Chem. Theory Comput.* **2017**, *13*, 2043–2052.

(346) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.

(347) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(348) Yu, H. S.; Zhang, W.; Verma, P.; He, X.; Truhlar, D. G. Nonseparable exchange–correlation functional for molecules, including homogeneous catalysis involving transition metals. *Phys. Chem. Chem. Phys.* **2015**, *17*.

(349) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **1992**, *46*, 6671–6687.

(350) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822.

(351) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.* **2009**, *103*, 026403.

(352) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(353) Brandenburg, J.; Bates, J.; Sun, J.; Perdew, J. Benchmark tests of a strongly constrained semilocal functional with a long-range dispersion correction. *Phys. Rev. B* **2016**, *94*, 115144.

(354) Mardirossian, N.; Ruiz Pestana, L.; Womack, J. C.; Skylaris, C.-K.; Head-Gordon, T.; Head-Gordon, M. Use of the rVV10 nonlocal correlation functional in the B97M-V density functional: defining B97M-rV and related functionals. *J. Phys. Chem. Lett.* **2016**, *8*, 35–40.

(355) Sabatini, R.; Gorni, T.; De Gironcoli, S. Nonlocal van der Waals density functional made simple and efficient. *Phys. Rev. B* **2013**, *87*, 4–7.

(356) Wellendorff, J.; Lundgaard, K. T.; Jacobsen, K. W.; Bligaard, T. mBEEF: An accurate semi-local Bayesian error estimation density functional. *J. Chem. Phys.* **2014**, *140*, 144107.

(357) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

(358) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A new local exchange-correlation functional for Kohn–Sham density functional theory with broad accuracy for atoms, molecules, and solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280–1293.

(359) Sun, J.; Haunschild, R.; Xiao, B.; Bulik, I. W.; Scuseria, G. E.; Perdew, J. P. Semilocal and hybrid meta-generalized gradient approximations based on the understanding of the kinetic-energy-density dependence. *J. Chem. Phys.* **2013**, *138*, 044113.

(360) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(361) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(362) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

(363) Hui, K.; Chai, J.-D. SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters. *J. Chem. Phys.* **2016**, *144*, 044114.

(364) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(365) Wang, Y.; Verma, P.; Jin, X.; Truhlar, D. G.; He, X. Revised M06 density functional for main-group and transition-metal chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10257–10262.

(366) Haoyu, S. Y.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn–Sham global-hybrid exchange–correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem. Sci.* **2016**, *7*, 5032–5051.

(367) Sun, J.; Perdew, J. P.; Ruzsinszky, A. Semilocal density functional obeying a strongly tightened bound for exchange. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 685–689.

(368) Gill, P. M.; Adamson, R. D.; Pople, J. A. Coulomb-attenuated exchange energy density functionals. *Mol. Phys.* **1996**, *88*, 1005–1009.

(369) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **1997**, *275*, 151–160.

(370) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

(371) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-range corrected hybrid density functionals with improved dispersion corrections. *J. Chem. Theory Comput.* **2012**, *9*, 263–272.

(372) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

(373) Peverati, R.; Truhlar, D. G. Improving the accuracy of hybrid meta-GGA density functionals by range separation. *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.

(374) Verma, P.; Wang, Y.; Ghosh, S.; He, X.; Truhlar, D. G. Revised M11 Exchange-Correlation Functional for Electronic Excitation Energies and Ground-State Properties. *J. Phys. Chem. A* **2019**, *123*, 2966–2990.

(375) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106.

(376) Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. Generalized gradient approximation model exchange holes for range-separated hybrids. *J. Chem. Phys.* **2008**, *128*, 194105.

(377) Peverati, R.; Truhlar, D. G. Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187–16191.

(378) Seidl, A.; Görling, A.; Vogl, P.; Majewski, J. A.; Levy, M. Generalized Kohn-Sham schemes and the band-gap problem. *Phys. Rev. B* **1996**, *53*, 3764–3774.

(379) Brémond, E.; Adamo, C. Seeking for parameter-free double-hybrid functionals: The PBE0-DH model. *J. Chem. Phys.* **2011**, *135*, 024106.

(380) Zhang, I. Y.; Xu, X.; Jung, Y.; Goddard, W. A. A fast doubly hybrid density functional method close to chemical accuracy using a local opposite spin ansatz. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 19896–19900.

(381) Kozuch, S.; Martin, J. M. L. Spin-component-scaled double hybrids: An extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *J. Comput. Chem.* **2013**, *34*, 2327–2344.

(382) Goerigk, L.; Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals- evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

(383) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(384) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical accuracy for the van der Waals density functional. *J. Phys.: Condens. Matter* **2009**, *22*, 022201.

(385) Carrete, J.; Longo, R.; Gallego, L.; Vega, A.; Balbás, L. Al enhances the H 2 storage capacity of graphene at nanoribbon borders but not at central sites: a study using nonlocal van der Waals density functionals. *Phys. Rev. B* **2012**, *85*, 125435.

(386) Xu, L.; Ge, Q. Effect of defects and dopants in graphene on hydrogen interaction in graphene-supported NaAlH4. *Int. J. Hydrogen Energy* **2013**, *38*, 3670–3680.

(387) Lebon, A.; Carrete, J.; Gallego, L.; Vega, A. Ti-decorated zigzag graphene nanoribbons for hydrogen storage. A van der Waals-corrected density-functional study. *Int. J. Hydrogen Energy* **2015**, *40*, 4960–4968.

(388) Kocman, M.; Jurečka, P.; Dubeck, M.; Otyepka, M.; Cho, Y.; Kim, K. S. Choosing a density functional for modeling adsorptive hydrogen storage: reference quantum mechanical calculations and a comparison of dispersion-corrected density functionals. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6423–6432.

(389) Boys, S. F.; Bernardi, F. d. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(390) Gráfová, L.; Pitonak, M.; Rezac, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory and Comput.* **2010**, *6*, 2365–2376.

(391) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(392) Takagi, H.; Hatori, H.; Soneda, Y.; Yoshizawa, N.; Yamada, Y. Adsorptive hydrogen storage in carbon and porous materials. *Mater. Sci. Eng., B* **2004**, *108*, 143–147.

(393) Thomas, K. M. Hydrogen adsorption and storage on porous materials. *Catal. Today* **2007**, *120*, 389–398.

(394) Murray, L. J.; Dincă, M.; Long, J. R. Hydrogen storage in metal–organic frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1294–1314.

(395) Bastos-Neto, M.; Patzschke, C.; Lange, M.; Möllmer, J.; Möller, A.; Fichtner, S.; Schrage, C.; Lässig, D.; Lincke, J.; Staudt, R., et al. Assessment of hydrogen storage by physisorption in porous materials. *Energy Environ. Sci.* **2012**, *5*, 8294–8303.

(396) Park, N.; Choi, K.; Hwang, J.; Kim, D. W.; Kim, D. O.; Ihm, J. Progress on first-principles-based materials design for hydrogen storage. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19893–19899.

(397) Hirscher, M.; Yartys, V. A.; Baricco, M.; von Colbe, J. B.; Blanchard, D.; Bowman Jr, R. C.; Broom, D. P.; Buckley, C. E.; Chang, F.; Chen, P., et al. Materials for hydrogen-based energy storage–past, recent progress and future outlook. *J. Alloys Compd.* **2020**, *827*, 153548.

(398) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 2844–2854.

(399) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. *Nat. Commun.* **2019**, *10*, 1–9.

(400) Mueller, T.; Ceder, G. A density functional theory study of hydrogen adsorption in MOF-5. *J. Phys. Chem. B* **2005**, *109*, 17974–17983.

(401) Cabria, I.; López, M.; Alonso, J. Hydrogen storage capacities of nanoporous carbon calculated by density functional and Møller-Plesset methods. *Phys. Rev. B* **2008**, *78*, 075415.

(402) Cabria, I.; López, M.; Alonso, J. Simulation of the hydrogen storage in nanoporous carbons with different pore shapes. *Int. J. Hydrogen Energy* **2011**, *36*, 10748–10759.

(403) Niaz, S.; Manzoor, T.; Islam, N.; Pandith, A. H. Theoretical investigations on C2H4Nb complex as a potential hydrogen storage system, using moller–plesset (MP2) and density functional theory. *Int. J. Quantum Chem.* **2014**, *114*, 449–457.

(404) Kocman, M.; Jurečka, P.; Dubeck, M.; Otyepka, M.; Cho, Y.; Kim, K. S. Choosing a density functional for modeling adsorptive hydrogen storage: reference quantum mechanical calculations and a comparison of dispersion-corrected density functionals. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6423–6432.

(405) Ma, L.-J.; Jia, J.; Wu, H.-S. Computational investigation of hydrogen storage on scandium–acetylene system. *Int. J. Hydrogen Energy* **2015**, *40*, 420–428.

(406) Sillar, K.; Hofmann, A.; Sauer, J. Ab Initio Study of Hydrogen Adsorption in MOF-5. *J. Am. Chem. Soc.* **2009**, *131*, 4143–4150.

(407) Koizumi, K.; Nobusada, K.; Boero, M. Hydrogen storage mechanism and diffusion in metal–organic frameworks. *Phys. Chem. Chem. Phys.* **2019**, *21*, 7756–7764.

(408) Jurečka, P.; Šponer, J.; Černy, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(409) Witte, J.; Goldey, M.; Neaton, J. B.; Head-Gordon, M. Beyond energies: Geometries of nonbonded molecular complexes as metrics for assessing electronic structure approaches. *J. Chem. Theory Comput.* **2015**, *11*, 1481–1492.

(410) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the ωB97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.

(411) Akima, H. A new method of interpolation and smooth curve fitting based on local procedures. *J. Assoc. Comput. Mach.* **1970**, *17*, 589–602.

(412) Taylor, D. E. et al. Blind test of density-functional-based methods on intermolecular interaction energies. *J. Chem. Phys.* **2016**, *145*, 124105.

(413) Neese, F.; Schwabe, T.; Grimme, S. Analytic derivatives for perturbatively corrected "double hybrid" density functionals: theory, implementation, and applications. *J. Chem. Phys.* **2007**, *126*, 124115.

(414) Fang, H.; Kamakoti, P.; Ravikovitch, P. I.; Aronson, M.; Paur, C.; Sholl, D. S. First principles derived, transferable force fields for CO2 adsorption in Na-exchanged cationic zeolites. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12882–12894.

(415) Fang, H.; Demir, H.; Kamakoti, P.; Sholl, D. S. Recent developments in first-principles force fields for molecules in nanoporous materials. *J. Mater. Chem. A* **2014**, *2*, 274–291.

(416) Becker, T. M.; Heinen, J.; Dubbeldam, D.; Lin, L.-C.; Vlugt, T. J. H. Polarizable Force Fields for CO2 and CH4 Adsorption in M-MOF-74. *J. Phys. Chem. C* **2017**, *121*, 4659–4673.

(417) Dubbeldam, D.; Walton, K. S.; Vlugt, T. J.; Calero, S. Design, parameterization, and implementation of atomic force fields for adsorption in nanoporous materials. *Adv. Theory Simul.* **2019**, *2*, 1900135.

Appendix A

Supporting Information: The Polarized Many-Body Expansion Approach

Table A.1: Energy decomposition analysis for all the systems studied in this work averaged over for all clusters of a given type. All the energy decomposition analysis were performed at ωB97M-V in def2-TZVPPD and are reported in kJ/mol.

	$(CO_2)_n$	$(H_2O)_n$	$Ca^{2+}(H_2O)_n$	$OH^-(H_2O)_n$
Frozen	-8.34	-34.70	-880.15	-84.48
Electrostatics	-194.49	-222.86	-1240.41	-772.44
Pauli repulsion	252.01	216.97	419.04	736.74
Dispersion	-65.86	-28.81	-58.78	-48.77
Polarization	-11.39	-23.88	-427.27	-144.85
Charge transfer	-7.84	-23.49	-67.35	-109.33

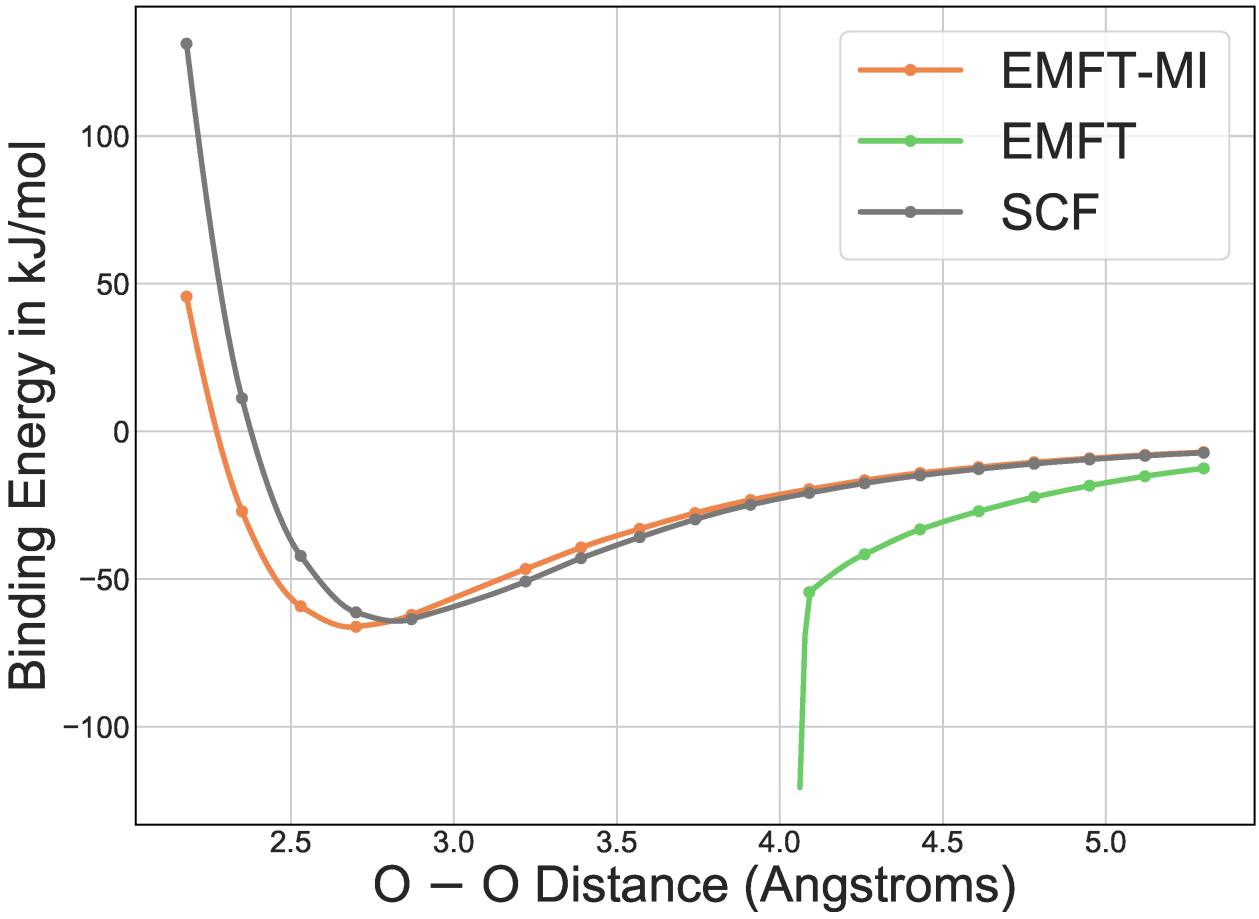


Figure A.1: A potential energy surface showing the binding energy of water trimer as a function of distance between the water molecules. The SCF curves indicates that the trimer energy is calculated using full SCF, and serves as the reference. The EMFT-MI curve shows one of the 1-body terms in PolBE, ie., monomer 1 is treated using XC1/BS1 and the other two monomers treated using XC2/BS2 and SCF-MI is used for selectively updating the density of monomer 1. The EMFT curves shows monomer 1 treated at XC1/BS1 and the remaining monomers treated at XC2/BS2 and using a full diagonalization to update the density. This causes an unnatural shift in the density to monomer 1 causing severe numerical instability at interacting distances.

Table A.2: Binding energies as predicted by full SCF, SCFMI, vacuum MBE, and PolBE and associated errors for CO_2 clusters of different sizes. All values are in kJ/mol.

	Binding energy				Error			Error per fragment		
n	SCF	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE
3	-15.43	-12.39	-14.91	-14.82	3.04	0.52	0.61	1.01	0.17	0.20
4	-27.66	-22.09	-24.80	-26.08	5.57	2.85	1.58	1.39	0.71	0.39
5	-40.51	-31.27	-36.06	-38.75	9.24	4.45	1.76	1.85	0.89	0.35
6	-54.12	-42.72	-48.06	-51.66	11.40	6.06	2.46	1.90	1.01	0.41
7	-70.98	-54.92	-61.64	-68.48	16.06	9.34	2.50	2.29	1.33	0.36
8	-85.91	-66.80	-74.34	-83.78	19.11	11.57	2.12	2.39	1.45	0.27
9	-102.39	-80.81	-88.98	-100.53	21.58	13.41	1.85	2.40	1.49	0.21
10	-116.96	-91.53	-101.92	-114.69	25.43	15.05	2.27	2.54	1.50	0.23
11	-133.95	-105.99	-115.29	-130.39	27.96	18.67	3.56	2.54	1.70	0.32
12	-153.15	-121.97	-131.71	-150.16	31.17	21.44	2.99	2.60	1.79	0.25
13	-175.10	-140.73	-150.31	-172.41	34.37	24.79	2.69	2.64	1.91	0.21
14	-189.86	-151.61	-162.64	-188.51	38.26	27.22	1.36	2.73	1.94	0.10
15	-207.72	-166.08	-178.20	-205.78	41.64	29.52	1.94	2.78	1.97	0.13
16	-225.29	-179.86	-193.46	-223.01	45.43	31.83	2.28	2.84	1.99	0.14

Table A.3: Binding energies as predicted by full SCF, SCFMI, vacuum MBE, and PolBE and associated errors for hydrated water clusters upto 20 water molecules. All values are in kJ/mol.

	Binding Energy				Error			Error per fragment		
n	SCF	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE
3	-63.60	-45.81	-55.80	-64.49	17.8	7.80	-0.89	5.93	2.60	-0.30
4	-111.19	-73.74	-91.27	-113.14	37.4	19.92	-1.95	9.36	4.98	-0.49
5	-145.72	-95.69	-115.67	-147.92	50.0	30.05	-2.20	10.01	6.01	-0.44
6	-186.88	-129.99	-158.72	-191.50	56.9	28.16	-4.61	9.48	4.69	-0.77
7	-234.20	-162.46	-196.61	-241.02	71.7	37.59	-6.82	10.25	5.37	-0.97
8	-296.07	-201.17	-248.71	-306.22	94.9	47.36	-10.15	11.86	5.92	-1.27
9	-333.15	-225.18	-276.58	-342.43	108.0	56.57	-9.27	12.00	6.29	-1.03
10	-379.13	-256.78	-314.75	-391.99	122.3	64.37	-12.87	12.23	6.44	-1.29
11	-406.12	-284.80	-344.62	-421.71	121.3	61.50	-15.59	11.03	5.59	-1.42
12	-477.49	-332.77	-402.74	-495.80	144.7	74.75	-18.31	12.06	6.23	-1.53
13	-499.75	-349.00	-422.15	-520.06	150.7	77.60	-20.31	11.60	5.97	-1.56
14	-508.23	-358.38	-432.34	-522.78	149.9	75.89	-14.55	10.70	5.42	-1.04
15	-607.89	-419.37	-504.44	-633.34	188.5	103.45	-25.44	12.57	6.90	-1.70
17	-687.70	-488.70	-584.89	-715.91	199.0	102.81	-28.21	11.71	6.05	-1.66
18	-719.58	-516.84	-614.42	-748.28	202.7	105.16	-28.70	11.26	5.84	-1.59
19	-784.02	-547.63	-652.85	-816.84	236.4	131.17	-32.82	12.44	6.90	-1.73
20	-845.22	-598.38	-712.87	-881.37	246.8	132.35	-36.15	12.34	6.62	-1.81

Table A.5: SCF, SCFMI, vacuum MBE, and PolBE root mean square binding energies and their associated root mean square errors for series of hydrated OH^- molecules. All values are in kJ/mol.

	Binding energy				Error			Error per fragment		
n	SCF	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE
3	231.00	134.17	246.66	238.50	96.83	15.66	7.50	32.28	5.22	2.50
4	305.25	204.09	338.13	316.21	102.31	33.88	11.30	25.58	8.47	2.83
5	372.43	237.35	397.97	389.29	137.07	30.54	16.93	27.41	6.11	3.39
6	435.60	295.51	475.09	456.20	141.39	42.43	20.69	23.57	7.07	3.45

Table A.4: Binding energies as predicted by full SCF, vacuum MBE, and PolBE and associated errors for hydrated Ca^{2+} clusters upto 20 water molecules. All values are in kJ/mol.

	Binding energy				Error			Error per fragment		
n	SCF	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE	SCFMI	MBE	PolBE
3	-433.12	-400.07	-456.06	-435.47	33.06	-22.94	-2.34	11.02	-7.65	-0.78
4	-616.77	-571.49	-675.63	-621.67	45.28	-58.86	-4.91	11.32	-14.72	-1.23
5	-777.36	-723.14	-889.08	-785.71	54.21	-111.72	-8.35	10.84	-22.34	-1.67
6	-910.32	-851.21	-1095.29	-921.39	59.11	-184.98	-11.07	9.85	-30.83	-1.85
7	-1032.27	-969.05	-1297.24	-1045.21	63.22	-264.97	-12.94	9.03	-37.85	-1.85
8	-1111.41	-1047.89	-1461.24	-1120.28	63.52	-349.84	-8.88	7.94	-43.73	-1.11
9	-1187.63	-1124.37	-1621.10	-1194.02	63.26	-433.47	-6.39	7.03	-48.16	-0.71
10	-1277.82	-1172.66	-1622.47	-1295.77	105.16	-344.64	-17.94	10.52	-34.46	-1.79
11	-1349.21	-1224.79	-1698.89	-1371.52	124.42	-349.68	-22.31	11.31	-31.79	-2.03
12	-1402.88	-1281.70	-1840.61	-1421.30	121.18	-437.73	-18.42	10.10	-36.48	-1.53
13	-1468.02	-1329.18	-1917.01	-1492.30	138.84	-448.99	-24.28	10.68	-34.54	-1.87
14	-1530.72	-1378.24	-1986.40	-1557.72	152.48	-455.69	-27.01	10.89	-32.55	-1.93
15	-1588.21	-1407.93	-2036.41	-1622.30	180.28	-448.21	-34.09	12.02	-29.88	-2.27
16	-1652.42	-1448.50	-2094.27	-1689.43	203.92	-441.85	-37.01	12.75	-27.62	-2.31
17	-1706.86	-1490.15	-2166.43	-1746.91	216.70	-459.57	-40.05	12.75	-27.03	-2.36
18	-1769.34	-1530.27	-2226.06	-1812.74	239.07	-456.73	-43.41	13.28	-25.37	-2.41
19	-1825.79	-1576.23	-2296.31	-1872.15	249.56	-470.52	-46.36	13.13	-24.76	-2.44
20	-1887.46	-1597.59	-2331.39	-1942.47	289.87	-443.93	-55.01	14.49	-22.20	-2.75
21	-1935.60	-1626.01	-2395.60	-1998.50	309.60	-459.99	-62.89	14.74	-21.90	-2.99

Table A.7: A table showing the CHELPG charges used for embedding in EE-MBE

System	Atom	ChelPG charge
CO_2	C	0.72
	O	-0.36
H_2O	H	0.34
	O	-0.69
Ca^{2+}	Ca	2.00
OH^-	H	0.20
	O	-1.20

Table A.6: PolBE errors for different representations of the environment. XC1=ωB97M-V, XC2=PBE, BS1=def2-TZVPPD, BS2=def2-SV(P). All values are in kJ/mol.

nfrgm	XC1/BS1	XC1/BS2	XC2/BS1	XC2/BS2
3	-0.45	-0.65	-0.13	-0.89
4	-1.11	-1.48	-1.92	-1.95
5	-1.46	-1.88	-2.75	-2.20
6	-2.95	-4.12	-4.34	-4.61
6	-3.09	-4.36	-5.53	-5.26
7	-4.43	-6.05	-7.45	-6.82
8	-7.13	-9.69	-10.49	-10.15
9	-7.69	-10.31	-11.15	-9.27
10	-9.48	-12.79	-14.73	-12.87
11	-10.68	-14.33	-17.23	-15.59
12	-13.51	-18.35	-21.75	-18.31
13	-14.09	-18.98	-24.49	-20.31
14	-13.44	-17.89	-18.22	-14.55
15	-17.82	-23.88	-30.78	-25.44
17	-20.59	-27.78	-35.73	-28.21
18	-21.26	-28.72	-36.45	-28.70
19	-23.57	-31.49	-40.36	-32.82
20	-25.98	-35.11	-46.24	-36.15

Table A.8: A table showing the dependence of EE-MBE binding energy errors (in kJ/mol) on the embedding charges used for two different types of embedding charges for water clusters. The errors are also heavily dependent on the basis set used for the calculations.

	Mulliken	CHELPG
ωB97M-V/def2-TZVPPD	2.85	0.53
ωB97M-V/cc-pVDZ	1.96	1.51

A note about computational timing analysis

All the timing calculations reported in this paper are wall clock times for 1 iteration ie., 1 iteration in case of SCF and total wall clock time for running 1 iteration of the 0-body term and each of the 1-body and 2-body terms in the case of PolBE. The timings reported in the main paper are the wall clock times taken to run the calculations using 16 threads. For 256 and 512 water molecule clusters, the PolBE 1-body and 2-body timings are estimated by averaging over 100 randomly chosen 1-body terms and 1000 randomly chosen 2-body terms

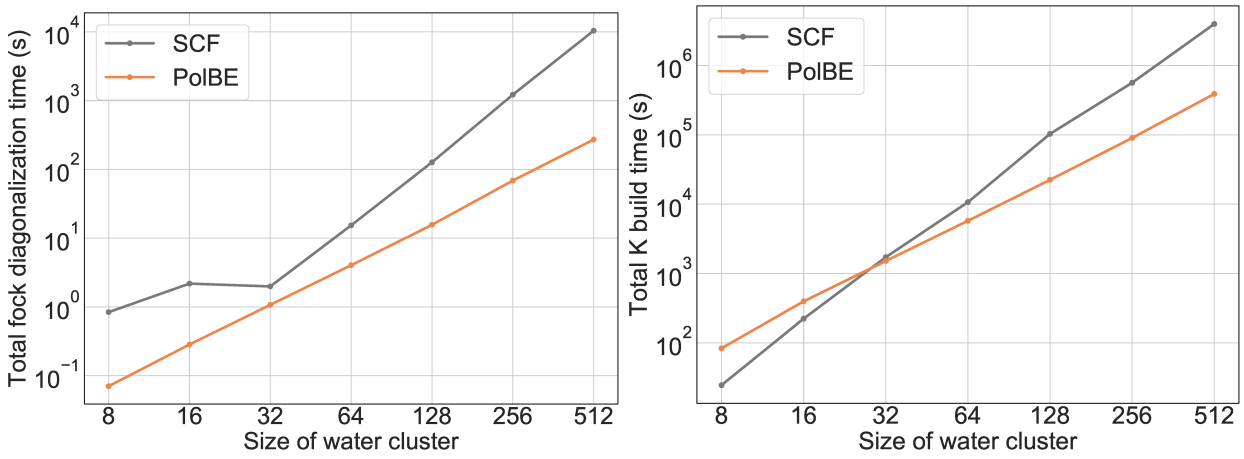


Figure A.2: (a) CPU time taken for Fock matrix diagonalization(s) in SCF and PolBE as a function of cluster size (b) CPU time taken for construction of exact exchange matrix(ces) for SCF and PolBE as a function of cluster size. F is the number of monomers in the system which indicates system size.

and extrapolating to get the total timings for all the ${}^{F}C_1$ 1-body terms and ${}^{F}C_2$ 2-body terms.

We have also performed the same analysis using CPU times instead of wall clock times and the corresponding figures are shown in Fig. A.2 and A.3. The performance analysis of SCF and PolBE does not depend on the type of timings (wall clock or CPU timings) used to perform the analysis.

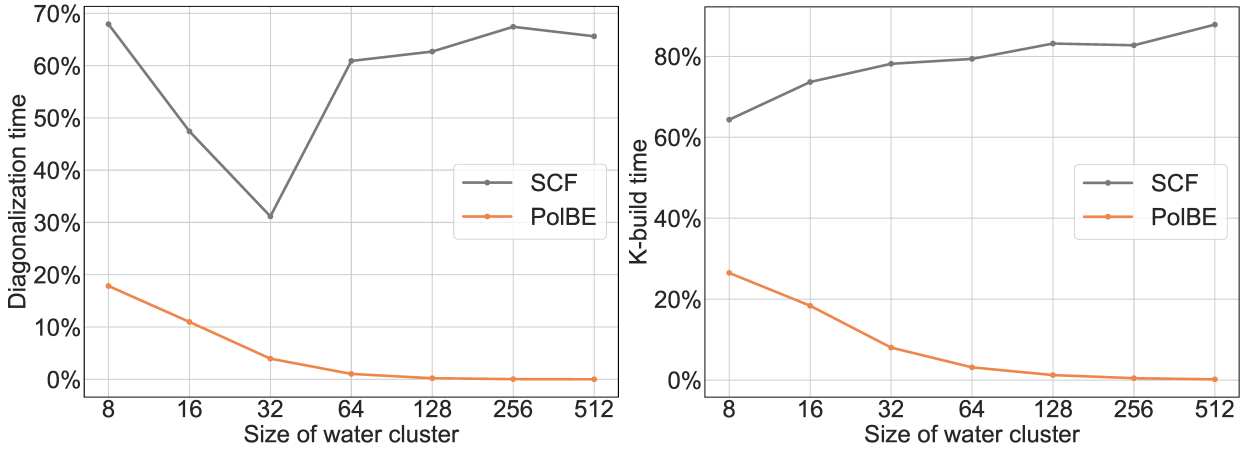


Figure A.3: (a) CPU time taken for diagonalization of Fock matrix(ces) as a percentage of total time taken for all linear algebra manipulations for SCF and PolBE as a function of cluster size (b) CPU time taken for construction of the exact exchange bit as a percentage of total time taken for construction of the Fock matrix(ces) for SCF and PolBE as a function of cluster size.

Appendix B

Supporting Information: A Non-Perturbative Pairwise-Additive Analysis of Charge Transfer Contributions

B.1 Working basis

In this section, we describe a method to construct an orthogonal working molecular orbital (MO) basis. We will work with the occupied and unoccupied space defined by the polarized wavefunction throughout this section. The polarized wavefunction (Φ_{POL}) gives a set of non-orthogonal occupied orbitals spanning the occupied space. In order to convert this set to an orthogonal basis, we symmetrically orthogonalize it as shown in Eq. (B.1) where the columns of the $\mathbf{T}^{\text{ortho}}$ and $\mathbf{T}$ contain the orthogonal and non-orthogonal occupied orbitals. Symmetric orthogonalization is sufficient as there is no linear dependency in the occupied space.

$$\mathbf{T}^{\text{ortho}} = \mathbf{T}\boldsymbol{\sigma}_{OO}^{-1/2} \tag{B.1}$$

$$\boldsymbol{\sigma}_{OO} = \mathbf{T}^{T}\mathbf{S}\mathbf{T} \tag{B.2}$$

Φ_{POL} also gives a set of non-orthogonal virtual orbitals defined by FERFs. As these orbitals might have some over overlap with the occupied space, we will project them out first as shown in Eq. (B.3) where $\mathbf{P} = \mathbf{T}\boldsymbol{\sigma}_{OO}^{-1}\mathbf{T}^{T}$ is the projector onto the occupied space. These orbitals are orthogonalized using standard canonical orthogonalization in order to account for linear dependence in the virtual space. $\mathbf{U}$ is matrix of eigenvectors with eigenvalues greater than threshold and $\mathbf{s}$ is a diagonal matrix of eigenvalues greater than threshold.

$$\mathbf{V}^{\text{proj}} = (\mathbf{I} - \mathbf{P}\mathbf{S})\mathbf{V} \tag{B.3}$$

$$\mathbf{V}^{\text{ortho}} = \mathbf{V}^{\text{proj}}\mathbf{U}\mathbf{s}^{-1/2} \tag{B.4}$$

The space not spanned by these the occupied and the FERF virtual space is the void space. An orthogonal basis for this space is formed by the eigenvectors of the projector ($\mathbf{R}_{\mathrm{POL}}$) for this space (Eq. (B.6)). and

$$\mathbf{P}_{\mathrm{POL}} = \mathbf{C}_{\mathrm{POL}}\boldsymbol{\sigma}\mathbf{C}_{\mathrm{POL}}^{T} \tag{B.5}$$
$$\mathbf{R}_{\mathrm{POL}} = \mathbf{I} - \mathbf{P}_{\mathrm{POL}} \tag{B.6}$$
$$\mathbf{R}^{\mathrm{ortho}} \leftarrow eigenvectors(\mathbf{R}_{\mathrm{POL}}) \tag{B.7}$$

An orthogonal basis for the entire Hilbert space is formed by the columns of $\mathbf{T}^{\mathrm{ortho}}$, $\mathbf{V}^{\mathrm{ortho}}$, and $\mathbf{R}^{\mathrm{ortho}}$.

B.2 Cost function and gradient

In this section, we describe the cost function to be minimized to solve for $\mathbf{X}_{OV}^{\mathrm{CT}}$ and derive its corresponding gradient. Let $\mathbf{U}_{XY}$ be the cumulative unitary transformation generated by $\boldsymbol{\theta}_X$ and $\boldsymbol{\theta}_Y$ respectively where $\mathbf{U}_X$ (referred to as $\mathbf{U}_{\mathrm{curr}}^{\mathrm{CT}}$ in Eq. (26)) denotes the current step taken and $\mathbf{U}_Y$ denotes the next step.

$$\mathbf{U}_X = e^{\boldsymbol{\theta}_X} \tag{B.8}$$
$$\mathbf{U}_Y = e^{\boldsymbol{\theta}_Y} \tag{B.9}$$
$$\mathbf{U}_{XY} = \mathbf{U}_X\mathbf{U}_Y \tag{B.10}$$

As we are trying to find the generator of unitary transformation that rotates the polarized density matrix ($\mathbf{P}^{\mathrm{POL}}$) to the fully-relaxed density matrix ($\mathbf{P}^{\mathrm{CT}}$), we are trying to find the $\boldsymbol{\theta}$ that satisfies Eq. (B.11)

$$\mathbf{P}^{\mathrm{CT}} = e^{\boldsymbol{\theta}}\mathbf{P}^{\mathrm{POL}}e^{-\boldsymbol{\theta}} \tag{B.11}$$

Hence, we define the cost function C as shown in Eq. (B.12)

$$C = ||\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY}\mathbf{P}^{\mathrm{POL}}\mathbf{U}_{XY}^{\dagger}||_F^2 \tag{B.12}$$

Before minimizing the cost function C, let us form a working orthogonal basis. Let H denote the Hilbert space of the whole complex. For the polarized wavefunction Φ_{POL}, H can be partitioned into occupied space (T), virtual space spanned by FERFs (Q), and the void space which is the unoccupied space not spanned by FERFs (R). Q and R can be clubbed together to form the unoccupied space V. These spaces can also be made strongly orthogonal to each other without loss of generality.

$$\begin{aligned} H &= T \oplus Q \oplus R \\ H &= T \oplus V \qquad\qquad \text{where } V = Q \oplus R \end{aligned} \tag{B.13}$$

Let the columns of $\mathbf{T}$ and $\mathbf{V}$ form a orthonormal basis for the occupied and unoccupied subspaces respectively. Let N_T and N_V be the dimensions of these subspaces respectively. In this orthonormal projectors, the density matrix which is the projector on the occupied space can be written as

$$\mathbf{P}^{\mathrm{POL}} = \begin{bmatrix} \mathbf{I}_{N_T \times N_T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{B.14}$$

Expressing the cost function in this orthonormal basis and taking its derivative with respect to the elements of the generator $\boldsymbol{\theta}_Y$,

$$\begin{aligned} \frac{\partial C}{\partial (\boldsymbol{\theta}_Y)_{ia}} &= \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \sum_{p,q} \left(\left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq}^2 \right) \\ &= 2 \sum_{p,q} \left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq} \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq} \end{aligned} \tag{B.15}$$

Consider,

$$\begin{aligned} \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq} &= \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \left(\mathbf{P}^{\mathrm{CT}}_{pq} - \sum_{k,l,m,n} (\mathbf{U}_X)_{pk} (\mathbf{U}_Y)_{kl} P^{\mathrm{POL}}_{lm} (\mathbf{U}^{\dagger}{}_Y)_{mn} (\mathbf{U}_X^{\dagger})_{nq} \right) \\ &= - \sum_{k,l,m,n} (\mathbf{U}_X)_{pk} \frac{\partial (U_Y)_{kl}}{\partial (\boldsymbol{\theta}_Y)_{ia}} P^{\mathrm{POL}}_{lm} (\mathbf{U}^{\dagger}{}_Y)_{mn} (\mathbf{U}_X^{\dagger})_{nq} \\ &\quad - \sum_{k,l,m,n} (\mathbf{U}_X)_{pk} (\mathbf{U}_Y)_{kl} P^{\mathrm{POL}}_{lm} \frac{\partial (\mathbf{U}^{\dagger}{}_Y)_{mn}}{\partial (\boldsymbol{\theta}_Y)_{ia}} (\mathbf{U}_X^{\dagger})_{nq} \end{aligned} \tag{B.16}$$

We know that $\frac{\partial (U_Y)_{kl}}{\partial (\boldsymbol{\theta}_Y)_{ia}} = \delta_{ki}\delta_{la} - \delta_{ka}\delta_{li}$. Substituting this above,

$$\begin{aligned} \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq} &= - \sum_{m,n} (\mathbf{U}_X)_{pi} P^{\mathrm{POL}}_{am} (\mathbf{U}_Y^{\dagger})_{mn} (\mathbf{U}_X^{\dagger})_{nq} + \sum_{m,n} (\mathbf{U}_X)_{pa} P^{\mathrm{POL}}_{im} (\mathbf{U}_Y^{\dagger})_{mn} (\mathbf{U}_X^{\dagger})_{nq} \\ &\quad - \sum_{k,l} (\mathbf{U}_X)_{pk} (\mathbf{U}_Y)_{kl} P^{\mathrm{POL}}_{la} (\mathbf{U}_X^{\dagger})_{iq} + \sum_{k,l} (\mathbf{U}_X)_{pk} (\mathbf{U}_Y)_{kl} P^{\mathrm{POL}}_{li} (\mathbf{U}_X^{\dagger})_{aq} \end{aligned}$$

Evaluating the above expression at $\boldsymbol{\theta}_Y = \mathbf{0}$, we have $\mathbf{U}_Y = \mathbf{I}$ and $\mathbf{U}_{XY} = \mathbf{U}_X$.

$$\begin{aligned} \frac{\partial}{\partial (\boldsymbol{\theta}_Y)_{ia}} \left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY} \mathbf{P}^{\mathrm{POL}} \mathbf{U}_{XY}^{\dagger} \right)_{pq} \Big|_{\boldsymbol{\theta}_Y = \mathbf{0}} &= - \sum_{m} (\mathbf{U}_X)_{pi} P^{\mathrm{POL}}_{am} (\mathbf{U}_X^{\dagger})_{mq} + \sum_{m} (\mathbf{U}_X)_{pa} P^{\mathrm{POL}}_{im} (\mathbf{U}_X^{\dagger})_{mq} \\ &\quad - \sum_{l} (\mathbf{U}_X)_{pl} P^{\mathrm{POL}}_{la} (\mathbf{U}_X^{\dagger})_{iq} + \sum_{l} (\mathbf{U}_X)_{pl} P^{\mathrm{POL}}_{li} (\mathbf{U}_X^{\dagger})_{aq} \end{aligned}$$

From Eq. (B.14), we know that $P^{\mathrm{POL}}_{am} = P^{\mathrm{POL}}_{la} = 0$ as a is an index for the unoccupied space. Thus we have,

$$\begin{aligned}\frac{\partial}{\partial(\boldsymbol{\theta}_Y)_{ia}}\left(\mathbf{P}^{\mathrm{CT}} - \mathbf{U}_{XY}\mathbf{P}^{\mathrm{POL}}\mathbf{U}^{\dagger}_{XY}\right)_{pq}\Big|_{\boldsymbol{\theta}_Y=\mathbf{0}} &= \sum_m (\mathbf{U}_X)_{pa} P^{\mathrm{POL}}_{im} (\mathbf{U}^{\dagger}_X)_{mq} + \sum_l (\mathbf{U}_X)_{pl} P^{\mathrm{POL}}_{li} (\mathbf{U}^{\dagger}_X)_{aq} \\ &= (\mathbf{U}_X)_{pa}(\mathbf{U}^{\dagger}_X)_{iq} + (\mathbf{U}_X)_{pi}(\mathbf{U}^{\dagger}_X)_{aq} \end{aligned} \tag{B.17}$$

Plugging Eq. (B.17) into Eq. (B.15),

$$\begin{aligned}\frac{\partial C}{\partial(\boldsymbol{\theta}_Y)_{ia}}\Big|_{\boldsymbol{\theta}_Y=\mathbf{0}} &= 2\sum_{p,q}\left(P^{\mathrm{CT}}_{pq} - \sum_{j,k}^{N_T}(\mathbf{U}_X)_{pj} P^{\mathrm{POL}}_{jk}(\mathbf{U}^{\dagger}_{kq})\right)\left((\mathbf{U}_X)_{pa}(\mathbf{U}^{\dagger}_X)_{iq} + (\mathbf{U}_X)_{pi}(\mathbf{U}^{\dagger}_X)_{aq}\right) \\ &= 2\sum_{p,q}\Bigg((\mathbf{U}^{\dagger}_X)_{iq} P^{\mathrm{CT}}_{qp}(\mathbf{U}_X)_{pa} + (\mathbf{U}^{\dagger}_X)_{ip} P^{\mathrm{CT}}_{pq}(\mathbf{U}_X)_{qa} \\ &\quad - \sum_k^{N_T}(\mathbf{U}^{\dagger}_X)_{jq}(\mathbf{U}^{\dagger}_X)_{jp}(\mathbf{U}_X)_{pa}(\mathbf{U}^{\dagger}_X)_{iq} - \sum_k^{N_T}(\mathbf{U}_X)_{pj}(\mathbf{U}_X)_{pi}(\mathbf{U}^{\dagger}_X)_{jq}(\mathbf{U}_X)_{qa}\Bigg) \end{aligned} \tag{B.18}$$

From Eq. (B.18), we can see that the first two terms of the RHS are equal and the third and fourth terms are 0 as $\sum_p (\mathbf{U}^{\dagger}_X)_{jp}(\mathbf{U}_X)_{pa} = 0$ and $\sum_p (\mathbf{U}^{\dagger}_X)_{jq}(\mathbf{U}_X)_{qa} = 0$ as basis vectors spanning T and V are orthonormal to each other. Finally, we have

$$\frac{\partial C}{\partial(\boldsymbol{\theta}_Y)_{ia}}\Big|_{\boldsymbol{\theta}_Y=\mathbf{0}} = 4\sum_{p,q}(\mathbf{U}^{\dagger}_X)_{iq} P^{\mathrm{CT}}_{qp}(\mathbf{U}_X)_{pa} \tag{B.19}$$

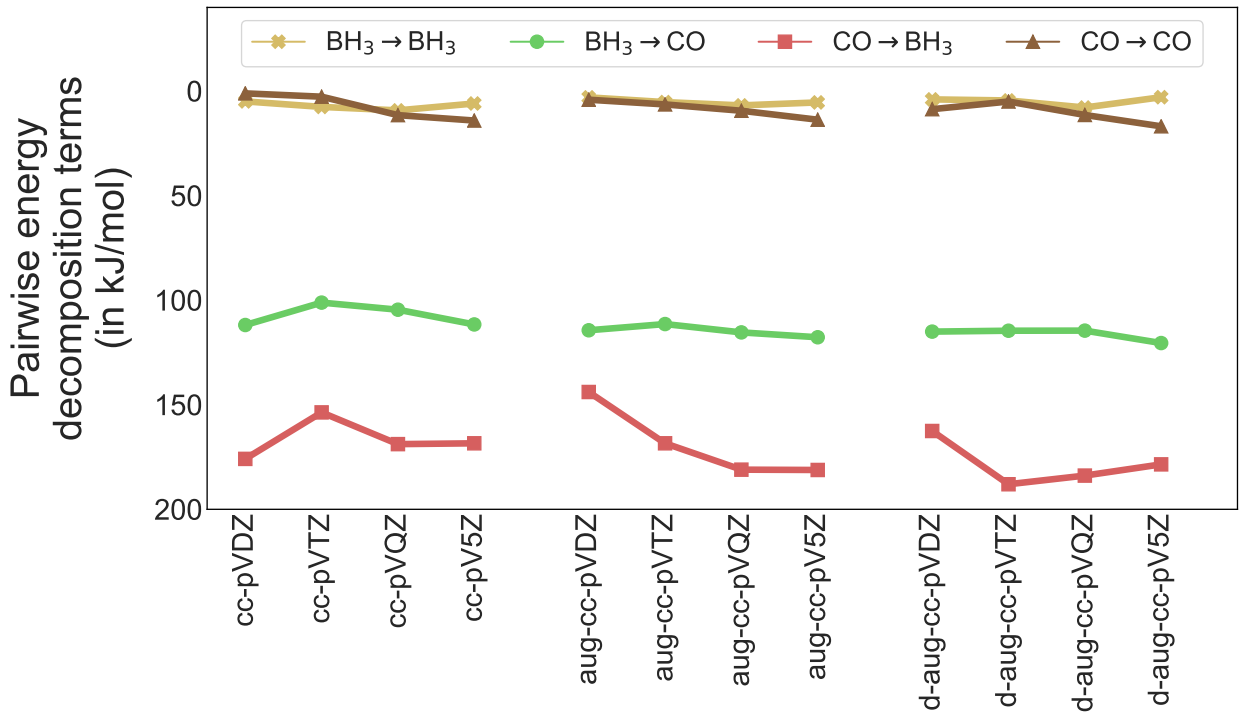


Figure B.1: Convergence properties of the non-perturbative pairwise energy decomposition components of the total charge transfer energy with respect to increasing the highest angular momentum of the basis set for the Dunning basis set series: cc-pVXZ, aug-ccpVXZ, and d-aug-cc-pVXZ (X=D, T, Q, and 5) for the BH_3–CO system at equilibrium geometry using ωB97X-D.

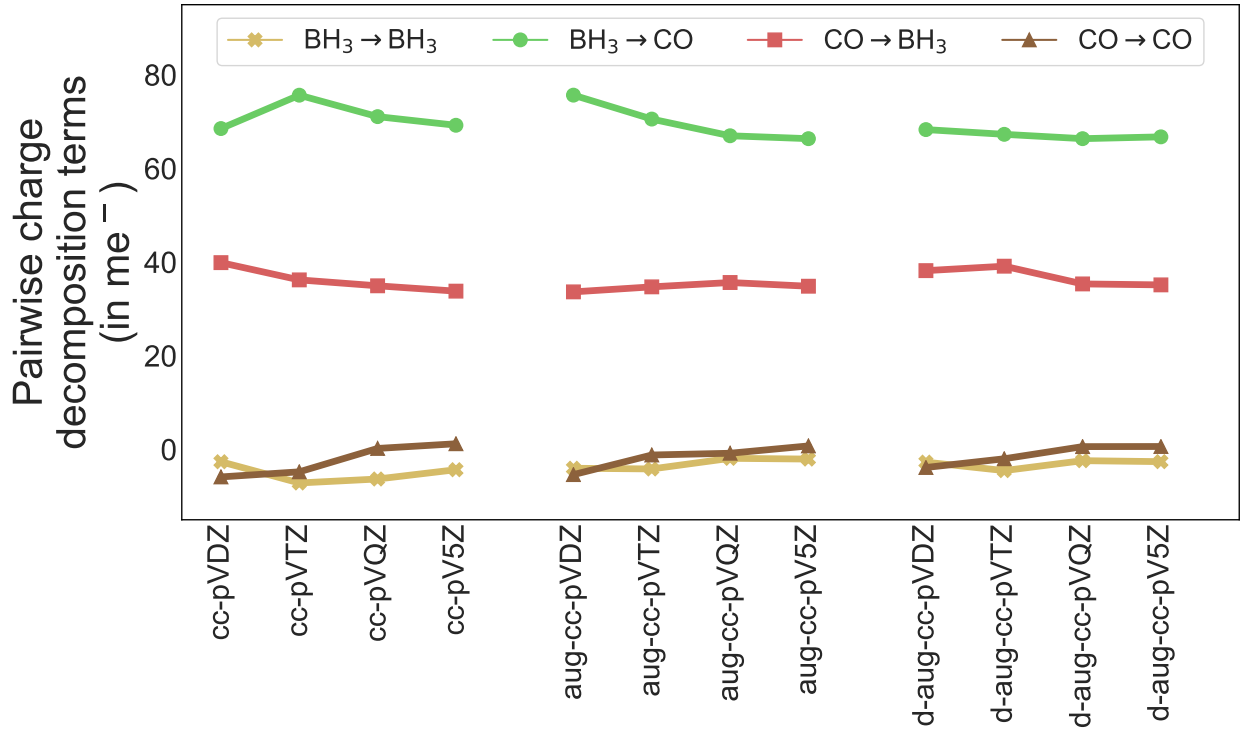


Figure B.2: Convergence properties of the non-perturbative pairwise charge decomposition components of total charge transfer with respect to increasing the highest angular momentum of the basis set for the Dunning basis set series: cc-pVXZ, aug-ccpVXZ, and d-aug-cc-pVXZ (X=D, T, Q, and 5) for the BH_3–CO system at equilibrium geometry using ωB97X-D.

Table B.1: Adiabatic energy decomposition analysis for the DNA base pairs thymine(T)-adenine(A), guanine(G)-cytosine(C), and their corresponding metallated versions. All values are in kJ/mol.

System	ΔE_{INT}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}
T(1):A(2)	-69.4	-31.8	-11.5	-20.6
G(1):C(2)	-134.3	-64.8	-27.5	-28.5
Na^+ G(1):C(2)	-149.3	-77.1	-31.7	-29.4
Mg^{2+} G(1):C(2)	-214.2	-120.9	-53.7	-21.2
Ca^{2+} G(1):C(2)	-195.2	-114.1	-48.3	-16.5

Table B.2: Energy decomposition analysis and non-perturbative charge decomposition analysis for the adduct BH_3–NCl_pH_q ($p+q=3$). All values are in kJ/mol.

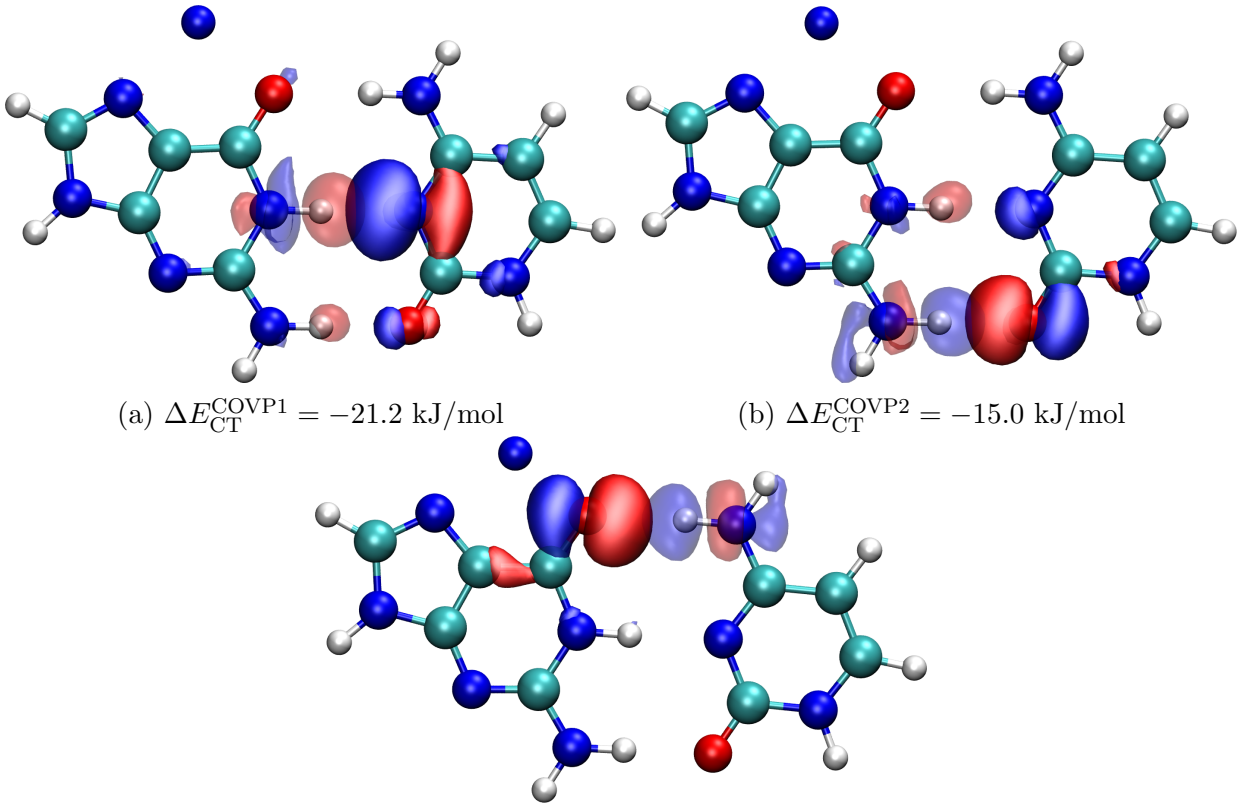
	Energy decomposition analysis					non-perturbative charge decomposition			
	ΔE_{INT}	ΔE_{GD}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}	$BH_3 \rightarrow BH_3$	$BH_3 \rightarrow NCl_mH_n$	$NCl_mH_n \rightarrow BH_3$	$NCl_mH_n \rightarrow NCl_pH_q$
BH_3-NH_3	-133.1	56.3	116.5	-148.1	-157.7	-2.7	-16.4	-138.1	-0.5
BH_3-$NClH_2$	-113.7	49.3	141.3	-156.5	-147.9	-2.6	-28.3	-117.9	0.9
BH_3-NCl_2H	-87.3	42.6	152.4	-138.7	-143.6	-2.3	-32.7	-109.2	0.6
BH_3-NCl_3	-58.2	34.1	145.5	-111.1	-126.7	-1.7	-29.8	-95.6	0.5

Table B.3: Adiabatic energy decomposition analysis for the adduct BX_3–NH_3 (X = F, Cl, or, Br). All values are in kJ/mol.

System	ΔE_{INT}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}
BF_3–NH_3	-90.8	-20.4	-6.6	-63.8
BCl_3–NH_3	-107.7	-10.6	-3.2	-93.8
BBr_3–NH_3	-119.2	-9.9	-3.0	-106.3

Table B.4: Adiabatic energy decomposition analysis for the adduct BH_3–NMe_pH_q ($p+q=3$). All values are in kJ/mol.

System	ΔE_{INT}	ΔE_{FRZ}	ΔE_{POL}	ΔE_{CT}
BH_3–NH_3	-133.1	-16.7	-14.5	-101.9
BH_3–$NMeH_2$	-153.2	-18.3	-28.2	-106.8
BH_3–NMe_2H	-163.1	-19.7	-37.4	-106.0
BH_3–NMe_3	-163.8	-21.1	-40.4	-102.3

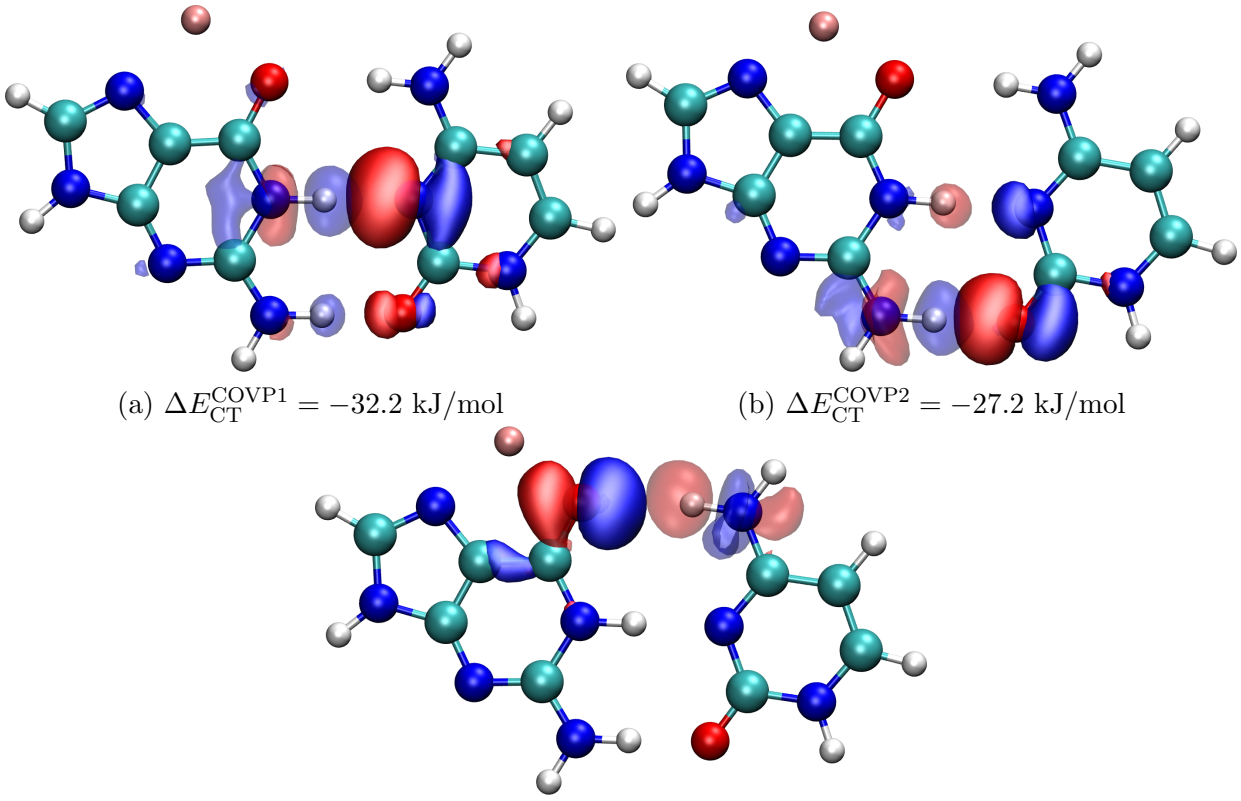


(a) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP1}} = -21.2$ kJ/mol

(b) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP2}} = -15.0$ kJ/mol

(c) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP3}} = -8.1$ kJ/mol

Figure B.3: (a) Most significant COVP for Na^+ guanine:cytosine complex (b) Second most significant COVP (c) Third most significant COVP

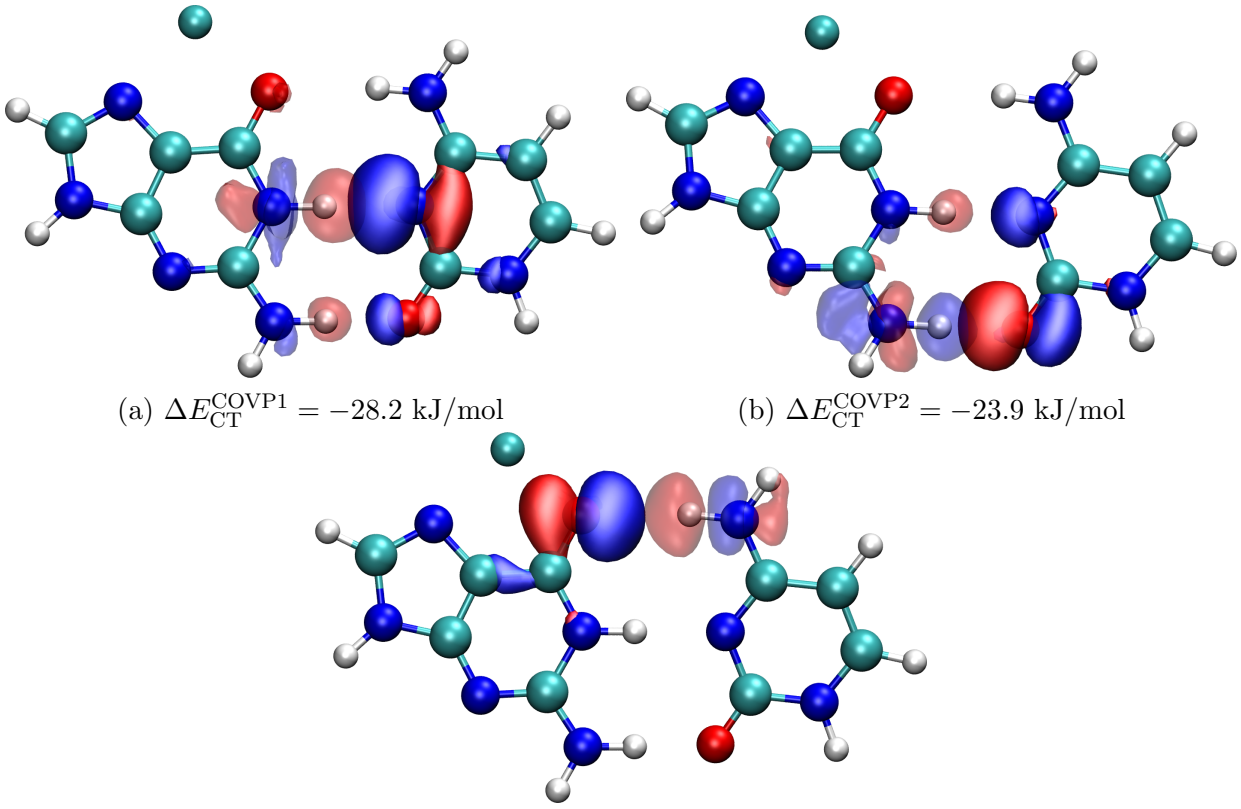


(a) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP1}} = -32.2$ kJ/mol

(b) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP2}} = -27.2$ kJ/mol

(c) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP3}} = -2.7$ kJ/mol

Figure B.4: (a) Most significant COVP for Mg^{2+} guanine:cytosine complex (b) Second most significant COVP (c) Third most significant COVP.



(a) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP1}} = -28.2$ kJ/mol

(b) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP2}} = -23.9$ kJ/mol

(c) $\Delta E_{\mathrm{CT}}^{\mathrm{COVP3}} = -2.9$ kJ/mol

Figure B.5: (a) Most significant COVP for Ca^{2+} guanine:cytosine complex (b) Second most significant COVP (c) Third most significant COVP.

Table B.5: Variational forward-backward charge transfer analysis for a series of 3d transition metal hexacarbonyls. All values are in kJ/mol.

System	$M(CO)_5 \rightarrow CO$	$CO \rightarrow M(CO)_5$
$V(CO)_6^-$	-76.2	-18.1
$Cr(CO)_6$	-45.9	-28.2
$Mn(CO)_6^+$	-22.6	-39.7

Table B.6: Residual energy not decomposed into pairwise additive terms using the non-perturbative CT analysis for some representative systems considered in this work using ωB97X-D density functional and def2-TZVPD basis set. Residual energy is shown in both kJ/mol and as a percentage of total variational CT energy.

System	Residual CT energy (kJ/mol)	Residual CT energy
$BF_3{-}NH_3$	6.86E-04	4.40E-04%
$BH_3{-}NH_3$	9.03E-04	5.78E-04%
$BH_3{-}CO$	7.95E-04	2.70E-04%
Mg^{2+} guanine : cytosine	7.95E-04	1.18E-03%
$Cr(CO)_6$	1.13E-06	5.33E-07%

B.3 Geometries

$BH_3{-}CO$
Charge=0, Multiplicity=1

B	-1.39961	0.70844	0.18735
H	-0.23322	0.93106	-0.03460
H	-1.99575	0.22613	-0.74595
H	-1.98125	1.61908	0.72730
C	-1.37842	-0.39863	1.23375
O	-1.36206	-1.21576	2.00602

Adenine − thymine
Charge = 0, Multiplicity=1

N	0.93525	0.00327	-0.37856
C	1.67122	0.00447	0.74242
C	3.06945	0.00430	0.59464
C	3.55409	0.00299	-0.70204
N	2.83563	0.00192	-1.82718
C	1.54067	0.00208	-1.57294
N	4.09798	0.00504	1.51249
C	5.17274	0.00449	0.78245
N	4.91459	0.00351	-0.56354
N	1.06445	0.00576	1.93073
H	0.86880	0.00126	-2.42521
H	6.18223	0.00493	1.16396
H	5.58348	0.00302	-1.31239
H	0.04879	0.00524	1.99236
H	1.62641	0.00609	2.76146
N	-3.90108	0.00699	-1.51169
C	-4.59567	0.00885	-0.33183
C	-3.98779	0.00849	0.86365
C	-2.52733	0.00612	0.87694
N	-1.90988	0.00427	-0.35524
C	-2.52226	0.00443	-1.58215
C	-4.71384	0.01056	2.16875
O	-1.86472	0.00574	1.90468
O	-1.92589	0.00347	-2.63504
H	-4.37932	0.00746	-2.39553
H	-0.86725	0.00300	-0.36549
H	-5.67377	0.01071	-0.42975
H	-4.44606	-0.86505	2.76243
H	-4.44328	0.88616	2.76118
H	-5.79328	0.01219	2.01585

guanine − cytosine

Charge=0, Multiplicity=1

N	-2.71342	-0.03847	0.00000
O	1.20960	2.30520	0.00000
N	-0.79313	1.23201	0.00000
C	-2.15789	1.18326	0.00000
N	-2.91822	2.25468	0.00000
C	-2.20356	3.39259	0.00000
C	-0.82822	3.56058	0.00000
C	-0.01516	2.39072	0.00000
N	-2.71003	4.65671	0.00000
C	-1.63428	5.51622	0.00000
N	-0.49613	4.89892	0.00000
H	-2.16332	-0.89459	0.00000
H	-3.71436	-0.08261	0.00000
H	-0.26632	0.34822	0.00000
H	-3.68517	4.89551	0.00000
H	-1.76809	6.58698	0.00000
O	-1.11321	-2.47931	0.00000
N	2.72139	-0.02967	0.00000
N	0.78997	-1.23187	0.00000
C	2.89188	-2.42337	0.00000
C	0.10811	-2.39792	0.00000
N	0.85582	-3.57753	0.00000
C	2.20860	-3.58117	0.00000
C	2.11745	-1.21115	0.00000
H	3.97064	-2.40651	0.00000
H	0.32769	-4.43351	0.00000
H	2.69059	-4.54922	0.00000
H	2.16866	0.84189	0.00000
H	3.72260	0.02130	0.00000

Na^+ guanine − cytosine

Charge=+1, Multiplicity=1

N	-2.65427	-0.03416	0.03627
O	1.23315	2.41597	-0.06545
N	-0.73630	1.24166	-0.01668
C	-2.11519	1.17712	0.00788
N	-2.89187	2.24861	0.00465
C	-2.21123	3.38907	-0.02376
C	-0.84074	3.55313	-0.04865
C	-0.01845	2.40865	-0.04524
N	-2.70715	4.66311	-0.03474
C	-1.64204	5.51877	-0.06468
N	-0.49939	4.89434	-0.07407
H	-2.09506	-0.89372	0.03818
H	-3.65671	-0.08599	0.05442
H	-0.18946	0.35722	-0.01308
H	-3.68093	4.91639	-0.02263
H	-1.77192	6.58931	-0.07877
Na	1.84056	4.62888	-0.10276
O	-1.11024	-2.36217	0.03702
N	2.82195	-0.07269	-0.05055
N	0.84004	-1.19096	-0.00644
C	2.88894	-2.46889	-0.02441
C	0.11296	-2.33555	0.01748
N	0.80910	-3.53840	0.01934
C	2.15829	-3.59976	-0.00106
C	2.16524	-1.23419	-0.02692
H	3.96712	-2.49681	-0.04085
H	0.24667	-4.37374	0.03659
H	2.60010	-4.58655	0.00255
H	2.30412	0.80227	-0.05347
H	3.82405	-0.05981	-0.06808

Mg^{2+} guanine − cytosine

Charge=+2, Multiplicity=1

N	-2.61052	-0.03711	0.02913
O	1.23629	2.56983	-0.07040
N	-0.68532	1.23801	-0.02048
C	-2.08767	1.15804	0.00365
N	-2.88581	2.23779	0.00206
C	-2.24699	3.37849	-0.02531
C	-0.88101	3.52903	-0.05089
C	-0.05249	2.41972	-0.04797
N	-2.71709	4.67819	-0.03546
C	-1.66054	5.52027	-0.06499
N	-0.51309	4.87276	-0.07549
H	-2.03211	-0.90577	0.02630
H	-3.61605	-0.09801	0.04688
H	-0.10135	0.34587	-0.01622
H	-3.68792	4.95986	-0.02287
H	-1.77273	6.59371	-0.07830
Mg	1.45076	4.48589	-0.10116
O	-1.10289	-2.22319	0.01710
N	2.94599	-0.12510	-0.03382
N	0.89565	-1.13327	-0.00904
C	2.88131	-2.51154	-0.01200
C	0.12312	-2.25302	0.00700
N	0.75563	-3.48006	0.01205
C	2.09895	-3.60922	0.00302
C	2.22167	-1.24973	-0.01835
H	3.95711	-2.58991	-0.01938
H	0.15265	-4.28890	0.02365
H	2.49260	-4.61628	0.00852
H	2.48475	0.76840	-0.04205
H	3.94843	-0.16363	-0.04308

Ca^{2+} guanine−cytosine

Charge=+2, Multiplicity=1

N	-2.58298	-0.02635	0.12588
O	1.24827	2.53708	-0.29794
N	-0.67352	1.25761	-0.06985
C	-2.06212	1.17282	0.06706
N	-2.85085	2.24999	0.14043
C	-2.21180	3.39024	0.04367
C	-0.85156	3.55722	-0.11985
C	-0.02982	2.43714	-0.16788
N	-2.70333	4.67550	0.07637
C	-1.66509	5.53215	-0.06648
N	-0.51361	4.91001	-0.18933
H	-2.01933	-0.89045	0.00183
H	-3.58350	-0.08686	0.22211
H	-0.09705	0.37090	-0.08356
Ca	1.76233	4.63674	-0.55969
H	-3.67270	4.93583	0.18876
H	-1.79983	6.60293	-0.07610
O	-1.09576	-2.22619	-0.24098
N	2.93860	-0.14267	0.13480
N	0.89972	-1.14774	-0.05726
C	2.87318	-2.53252	0.10355
C	0.12503	-2.26268	-0.14235
N	0.75440	-3.49334	-0.11493
C	2.09120	-3.62701	0.00868
C	2.21776	-1.26905	0.05960
H	3.94410	-2.61578	0.20225
H	0.15213	-4.29975	-0.18306
H	2.48143	-4.63531	0.02510
H	2.47403	0.74792	0.07500
H	3.93527	-0.18024	0.24002

BF_3−NH_3

Charge=0, Multiplicity=1

B	-1.59518	0.54532	0.32606
F	-0.23695	0.73407	0.37665
F	-2.04606	0.02288	-0.85972
F	-2.32812	1.61046	0.78543
N	-1.88362	-0.66855	1.44789
H	-1.37027	-1.50567	1.19306
H	-1.58449	-0.36790	2.36960
H	-2.87397	-0.88713	1.47898

BCl_3−NH_3

Charge=0, Multiplicity=1

B	-1.59730	0.47668	0.37371
Cl	0.21507	0.75918	0.43351
Cl	-2.17485	-0.19770	-1.23203
Cl	-2.57784	1.91478	0.95444
N	-1.87918	-0.69508	1.45460
H	-1.36490	-1.53304	1.19591
H	-1.57836	-0.39161	2.37738
H	-2.87293	-0.90943	1.48346

BBr_3−NH_3

Charge=0, Multiplicity=1

B	-1.62062	0.49314	0.42622
Br	0.36052	0.77141	0.45700
Br	-2.28711	-0.20549	-1.32786
Br	-2.65735	2.06939	1.09335
N	-1.90097	-0.68153	1.48934
H	-1.41915	-1.53020	1.20189
H	-1.56429	-0.40332	2.40867
H	-2.89964	-0.86928	1.54658

BH_3−NH_3

Charge=0, Multiplicity=1

B	-1.58381	0.55528	0.34974
H	-0.38328	0.72225	0.38092
H	-1.98319	0.12157	-0.70965
H	-2.21409	1.51121	0.74819
N	-1.87902	-0.64048	1.44259
H	-1.38608	-1.48859	1.18744
H	-1.57317	-0.36299	2.36835
H	-2.86964	-0.85142	1.48661

BH_3−$NMeH_2$

Charge=0, Multiplicity=1

B	-1.59029	0.54147	0.39945
H	-0.39079	0.72711	0.43604
H	-1.97430	0.09320	-0.66147
H	-2.23445	1.49911	0.77674
N	-1.87937	-0.63349	1.49587
H	-1.38938	-1.46918	1.19558
C	-1.48404	-0.28508	2.87007
H	-2.86906	-0.85414	1.46563
H	-0.42001	-0.05929	2.87101
H	-1.69613	-1.09524	3.56844
H	-2.02592	0.61129	3.16307

BH_3−NMe_2H

Charge=0, Multiplicity=1

B	-1.65861	0.49965	0.45298
H	-0.45988	0.69270	0.47355
H	-2.05038	0.05252	-0.60590
H	-2.30362	1.45632	0.83556
N	-1.93468	-0.67196	1.55441
H	-1.39656	-1.47217	1.23838
C	-1.44483	-0.29306	2.88935
C	-3.35278	-1.06310	1.59141
H	-0.38434	-0.06317	2.82100
H	-1.61396	-1.09433	3.61145
H	-1.97644	0.60373	3.20305
H	-3.65413	-1.38649	0.59816
H	-3.93824	-0.18679	1.86471
H	-3.52118	-1.86102	2.31737

BH_3−NMe_3

Charge=0, Multiplicity=1

B	-1.64959	0.48063	0.45236
H	-0.45143	0.67820	0.46539
H	-2.05248	0.07545	-0.61921
H	-2.29029	1.42127	0.87602
N	-1.90758	-0.74278	1.51306
C	-1.16084	-1.93826	1.07956
C	-1.44677	-0.33716	2.85395
C	-3.34795	-1.05555	1.56332
H	-0.38655	-0.10000	2.80202
H	-1.61644	-1.14081	3.57503
H	-1.99170	0.55415	3.15628
H	-3.67753	-1.34798	0.56899
H	-3.88925	-0.16164	1.86463
H	-3.53785	-1.86403	2.27386
H	-1.49956	-2.22155	0.08562
H	-1.32323	-2.76194	1.77935
H	-0.10253	-1.69175	1.03217

BH_3−$NClH_2$

Charge=0, Multiplicity=1

B	-1.53814	0.56757	0.34596
H	-0.32860	0.71229	0.34544
H	-1.97360	0.10722	-0.68119
H	-2.14127	1.51869	0.77989
N	-1.68261	-0.60847	1.46811
H	-1.16707	-1.43357	1.17646
H	-1.29749	-0.29016	2.35227
Cl	-3.29323	-1.16311	1.82886

BH_3−NCl_2H

Charge=0, Multiplicity=1

B	-1.61591	0.60936	0.30585
H	-0.43193	0.81038	0.22872
H	-2.12018	0.10704	-0.67266
H	-2.26226	1.50726	0.79531
N	-1.80504	-0.58507	1.42578
Cl	-1.07122	-2.10160	0.98489
Cl	-1.31931	-0.13703	3.03652
H	-2.79617	-0.79988	1.51139

BH_3−NCl_3

Charge=0, Multiplicity=1

B	-1.57622	0.64506	0.27283
H	-0.38144	0.79375	0.24696
H	-2.06443	0.13474	-0.70278
H	-2.20824	1.54234	0.76857
N	-1.75093	-0.58737	1.43594
Cl	-0.89748	-2.03312	0.95071
Cl	-1.11081	-0.08456	2.98367
Cl	-3.43245	-1.00039	1.65990

$V(CO)_5$−CO

Charge=-1, Multiplicity=1

V	-1.05210	1.59272	-0.03670
C	0.90067	1.57110	-0.03405
C	-3.00482	1.61663	-0.04089
C	-1.06987	-0.36021	-0.05396
C	-1.05834	1.58186	1.91658
C	-1.04625	1.60436	-1.98915
O	-1.06337	1.57707	3.06784
O	-1.04170	1.61032	-3.14041
O	2.05187	1.55644	-0.03428
O	-4.15598	1.63404	-0.04307
O	-1.07863	-1.51140	-0.06736
C	-1.03419	3.54531	-0.02414
O	-1.02535	4.69655	-0.01610

$Cr(CO)_5$−CO

Charge=0, Multiplicity=1

Cr	-1.05211	1.59284	-0.03692
C	0.85796	1.57173	-0.03412
C	-2.96212	1.61590	-0.04103
C	-1.06970	-0.31747	-0.05341
C	-1.05809	1.58204	1.87381
C	-1.04646	1.60430	-1.94662
O	-1.06308	1.57722	3.00875
O	-1.04195	1.61002	-3.08156
O	1.99283	1.55718	-0.03416
O	-4.09695	1.63306	-0.04290
O	-1.07824	-1.45232	-0.06656
C	-1.03442	3.50269	-0.02454
O	-1.02574	4.63759	-0.01643

$Mn(CO)_5$−CO

Charge=+1, Multiplicity=1

Mn	-1.05217	1.59359	-0.03842
C	0.85778	1.57339	-0.03352
C	-2.96213	1.61388	-0.04247
C	-1.07189	-0.31631	-0.05159
C	-1.05710	1.58062	1.87154
C	-1.04733	1.60611	-1.94849
O	-1.05998	1.57267	2.99262
O	-1.04449	1.61318	-3.06958
O	1.97881	1.56117	-0.02988
O	-4.08317	1.62573	-0.04420
O	-1.08351	-1.43732	-0.05906
C	-1.03233	3.50355	-0.02536
O	-1.02056	4.62455	-0.01727

Appendix C

Supporting Information: Benchmarking Density Functionals for Hydrogen Storage: The H2Bind275 Dataset

Addition files referenced in this chapter can be found online.

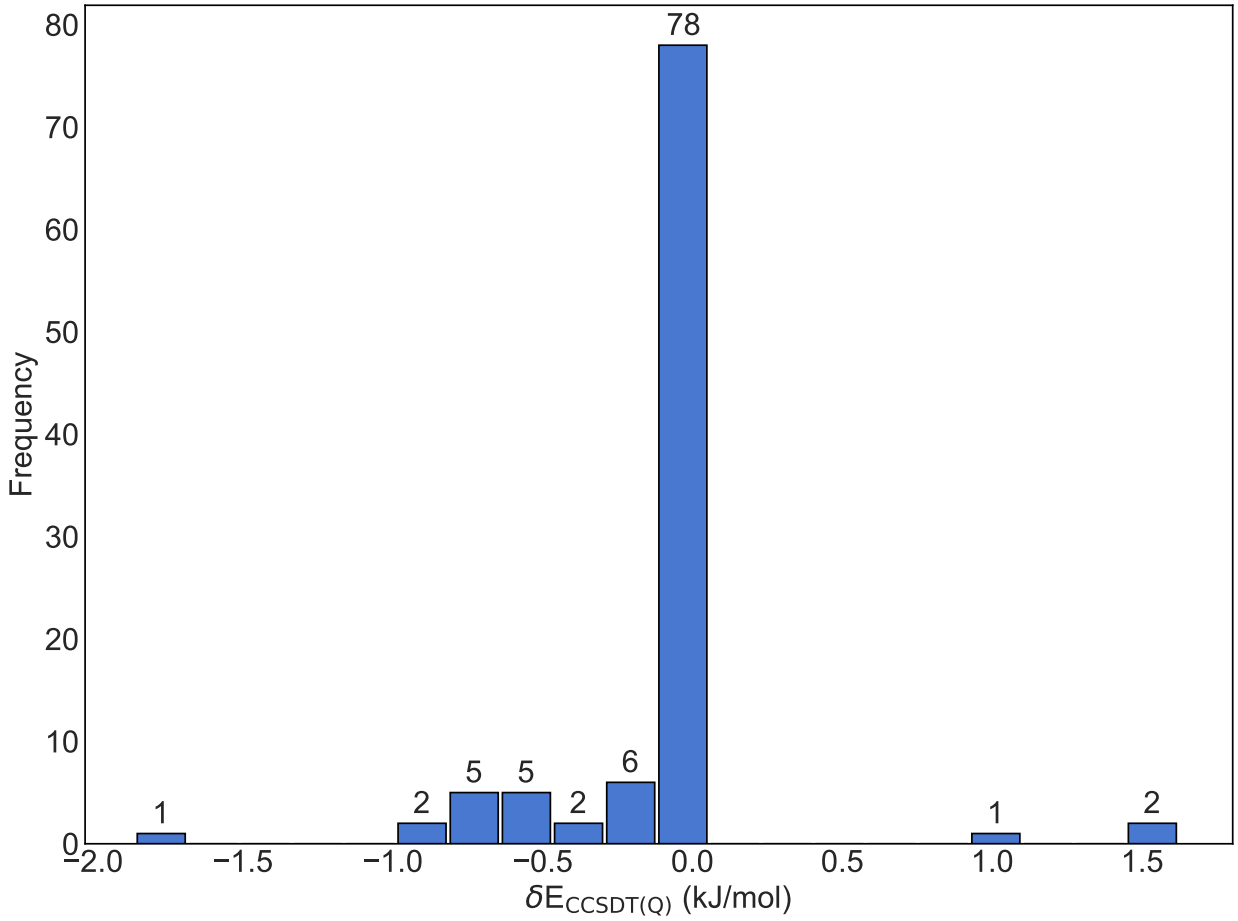


Figure C.1: Distribution of post-CCSD(T) corrections to the interaction energy ($\delta E_{\mathrm{CCSDT(Q)}}$) in the 6-31G**(mod) basis set.

Table C.1: Parameters for treatment of MP2 correlation energy for each of the double hybrid functionals considered in this work.

Density functional	Density fitting employed?	Level of frozen core approximation
ωB97M(2)	yes	all core electrons
DSD-PBEPBE-D3(BJ)	yes	all core electrons
XYG3	no	all core electrons
B2PLYP-D3(BJ)	yes	no core electrons were frozen
PTPSS-D3(0)	yes	all core electrons
PBE0-DH	no	no core electrons were frozen
XYGJOS	yes	same as the coupled cluster reference

Table C.2: The ground state spin multiplicity predicted by CCSD(T) and used in the H2Bind275 dataset. In all cases where the corresponding experimental references were available, the CCSD(T) spin ground state agrees with experiments.

Species	CCSD(T)	Experiment	Reference for experiment
Sc^+	3	3	J. Sugar and C. Corliss, J. Phys. Chem. Ref. Data 14, Suppl. 2 (1985)
Ti^+	4	4	J. E. Sohl, Y. Zhu, and R. D. Knight, J. Opt. Soc. Am. B 7, 9 (1990).
V^+	5	5	A. M. James, P. Kowalczyk, E. Langlois, M. D. Campbell, A. Ogawa, and B. Simard, J. Chem. Phys. 101, 4485 (1994).
Cr^+	6	6	J. Sugar and C. Corliss, J. Phys. Chem. Ref. Data 14, Suppl. 2 (1985).
Mn^+	7	7	J. Sugar and C. Corliss, J. Phys. Chem. Ref. Data 14, Suppl. 2 (1985).
Fe^+	6	6	J. Sugar and C. Corliss, J. Phys. Chem. Ref. Data 14, Suppl. 2 (1985).
Co^+	3	3	R. H. Page and C. S. Gudeman, J. Opt. Soc. Am. B 7, 1761 (1990).
Ni^+	2	2	R. H. Page and C. S. Gudeman, J. Opt. Soc. Am. B 7, 1761 (1990).
Cu^+	1	1	J. Sugar and A. Musgrove, J. Phys. Chem. Ref. Data 19, 527 (1990)
Zn^+	2	2	J. Sugar and A. Musgrove, J. Phys. Chem. Ref. Data 24, 1803 (1995).
CoF_3	5	n/a	
$CrCl_2$	5	5	B. Vest, Z. Varga, M. Hargittai, A. Hermann, and P. Schwerdtfeger Chem. Eur. J. 14, 5130 (2008)
$Ni(CN)_2$	3	n/a	
$Ni(OH)_2$	1	n/a	
CuCN	1	n/a	
CuCl	1	n/a	
CuOMe	1	n/a	
CuX (X=H,F,Cl)	1	n/a	
AuX (X=H,F,Cl)	1	n/a	
AgX (X=H,F,Cl)	1	n/a	

Table C.3: Performance of density functional approximations with and without exact exchange by three error metrics: RegMAPE, RMSE in kJ/mol, and rank of the DFA determined by RegMAPE.

Local				Hybrid			
Functional	RegMAPE	RMSE	Rank	Functional	RegMAPE	RMSE	Rank
TPSS	14.0	6.7	26	TPSSh	11.8	5.2	21
BLYP	21.7	8.2	49	B3LYP	15.1	6.0	30
PBE	15.2	9.3	31	PBE0	8.0	4.1	11
SCAN	14.7	10.9	29	SCAN0	9.2	5.3	15
M06-L	22.9	10.2	50	M06	19.4	9.1	42
MN15-L	19.4	6.7	41	MN15	17.9	6.3	38

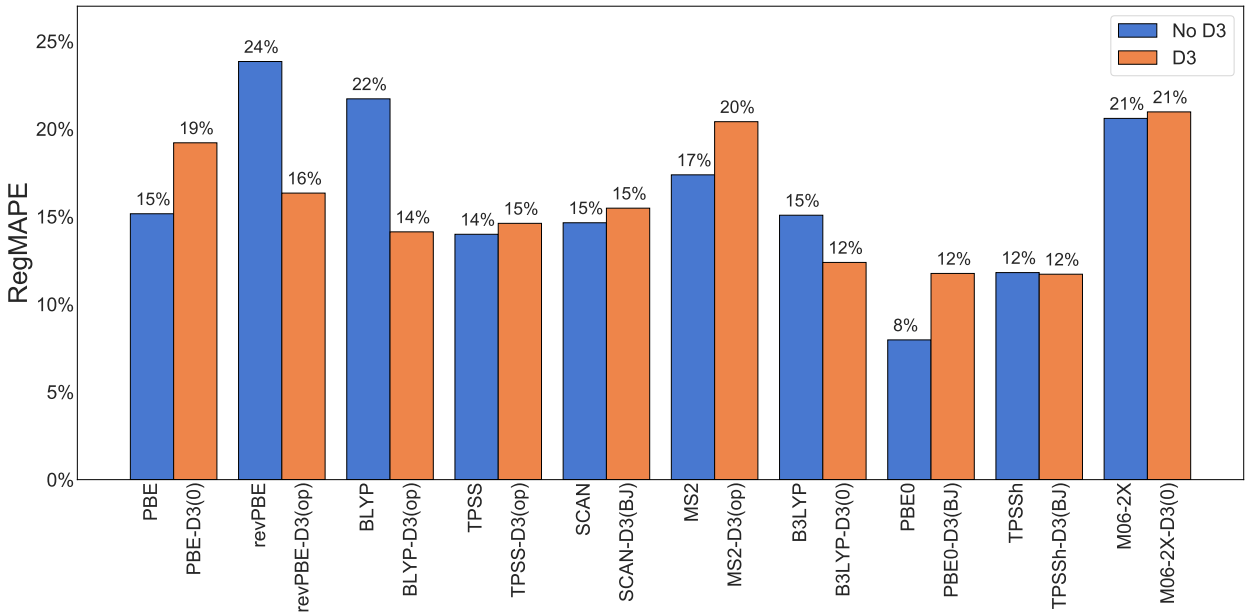


Figure C.2: The performance of density functional approximations with and without empirical D3 dispersion corrections.

Table C.4: RMSE (in kJ/mol) for seven selected density functionals on a smaller dataset for ωB97M-V optimized geometries (H2Bind7) and displaced geometries (H2Bind7x5).

Functional	H2Bind7	H2Bind7x5
DSD-PBEPBE-D3(BJ)	1.6	2.2
ωB97M-V	1.9	2.5
ωB97X-V	2.0	2.1
ωB97X-D	3.3	3.2
B97M-V	4.9	3.9
BLYP-D3(op)	8.6	7.8
revPBE-D3op	8.6	8.1

In order assess if using ωB97M-V geometries induces bias in the performance on density functionals, we assessed the performance of density functionals on geometries around the ωB97M-V optimized geometries. In order to perform this, we displaced the optimized geometries by $0.9r_e$, $1.2r_e$, $1.5r_e$, and $2.0r_e$. These displacements were previously employed to compile the S22x5 dataset. A small subset consisting of Na$-1\,H_2$, Mg$-1\,H_2$, Li$-1\,H_2$, $MgF_2-1\,H_2$, $CaF_2-1\,H_2$, $AlF_3-1\,H_2$, benzene$-1\,H_2$, and $Cr^+-1\,H_2$ (called H2Bind7) and their corresponding displaced geometries (called H2Bind7x5) were used to perform this analysis. We can see that the relative ordering of the density functionals for the H2Bind7 dataset

and for the H2Bind7x5 dataset is more or less identical (as seen in Table (4.5)). This shows that using the ωB97M-V geometries does not bias the dataset.

Table C.5: Data points in the H2Bind275 dataset along with the charge, spin multiplicity, category, and corresponding filename.

Molecule	# H2	Charge	Multiplicity	Category	Filename
Ca^{2+}−1 H_2(vert)	1	2	1	s-block ion	Ca-1H2.xyz
Ca^{2+}−2 H_2(vert)	2	2	1	s-block ion	Ca-2H2.xyz
Ca^{2+}−3 H_2(vert)	3	2	1	s-block ion	Ca-3H2.xyz
Ca^{2+}−4 H_2(vert)	4	2	1	s-block ion	Ca-4H2.xyz
Ca^{2+}−5 H_2(vert)	5	2	1	s-block ion	Ca-5H2.xyz
Ca^{2+}−6 H_2(vert)	6	2	1	s-block ion	Ca-6H2.xyz
Li^{+}−1 H_2(vert)	1	1	1	s-block ion	Li-1H2.xyz
Li^{+}−2 H_2(vert)	2	1	1	s-block ion	Li-2H2.xyz
Li^{+}−3 H_2(vert)	3	1	1	s-block ion	Li-3H2.xyz
Li^{+}−4 H_2(vert)	4	1	1	s-block ion	Li-4H2.xyz
Li^{+}−5 H_2(vert)	5	1	1	s-block ion	Li-5H2.xyz
Li^{+}−6 H_2(vert)	6	1	1	s-block ion	Li-6H2.xyz
Mg^{2+}−1 H_2(vert)	1	2	1	s-block ion	Mg-1H2.xyz
Mg^{2+}−2 H_2(vert)	2	2	1	s-block ion	Mg-2H2.xyz
Mg^{2+}−3 H_2(vert)	3	2	1	s-block ion	Mg-3H2.xyz
Mg^{2+}−4 H_2(vert)	4	2	1	s-block ion	Mg-4H2.xyz
Na^{+}−1 H_2(vert)	1	1	1	s-block ion	Na-1H2.xyz
Na^{+}−2 H_2(vert)	2	1	1	s-block ion	Na-2H2.xyz
Na^{+}−3 H_2(vert)	3	1	1	s-block ion	Na-3H2.xyz
Ca^{2+}−1 H_2(ad)	1	2	1	s-block ion	Ca-1H2.xyz
Ca^{2+}−2 H_2(ad)	1	2	1	s-block ion	Ca-2H2.xyz
Ca^{2+}−3 H_2(ad)	1	2	1	s-block ion	Ca-3H2.xyz
Ca^{2+}−4 H_2(ad)	1	2	1	s-block ion	Ca-4H2.xyz
Ca^{2+}−5 H_2(ad)	1	2	1	s-block ion	Ca-5H2.xyz
Ca^{2+}−6 H_2(ad)	1	2	1	s-block ion	Ca-6H2.xyz
Li^{+}−1 H_2(ad)	1	1	1	s-block ion	Li-1H2.xyz
Li^{+}−2 H_2(ad)	1	1	1	s-block ion	Li-2H2.xyz
Li^{+}−3 H_2(ad)	1	1	1	s-block ion	Li-3H2.xyz
Li^{+}−4 H_2(ad)	1	1	1	s-block ion	Li-4H2.xyz
Li^{+}−5 H_2(ad)	1	1	1	s-block ion	Li-5H2.xyz
Li^{+}−6 H_2(ad)	1	1	1	s-block ion	Li-6H2.xyz
Mg^{2+}−1 H_2(ad)	1	2	1	s-block ion	Mg-1H2.xyz
Mg^{2+}−2 H_2(ad)	1	2	1	s-block ion	Mg-2H2.xyz
Mg^{2+}−3 H_2(ad)	1	2	1	s-block ion	Mg-3H2.xyz

Table C.5 continued from previous page

Molecule	# H2	Charge	Multiplicity	Category	Filename
Mg^{2+}−4 H_2(ad)	1	2	1	s-block ion	Mg-4H2.xyz
Na^{+}−1 H_2(ad)	1	1	1	s-block ion	Na-1H2.xyz
Na^{+}−2 H_2(ad)	1	1	1	s-block ion	Na-2H2.xyz
Na^{+}−3 H_2(ad)	1	1	1	s-block ion	Na-3H2.xyz
AlF_3−1 H_2(vert)	1	0	1	salt	alf3-1H2.xyz
$CaCl_2$−1 H_2(vert)	1	0	1	salt	cacl-1H2.xyz
$CaCl_2$−2 H_2(vert)	2	0	1	salt	cacl-2H2.xyz
$CaCl_2$−3 H_2(vert)	3	0	1	salt	cacl-3H2.xyz
$CaCl_2$−4 H_2(vert)	4	0	1	salt	cacl-4H2.xyz
CaF_2−1 H_2(vert)	1	0	1	salt	caf-1H2.xyz
CaF_2−2 H_2(vert)	2	0	1	salt	caf-2H2.xyz
CaF_2−3 H_2(vert)	3	0	1	salt	caf-3H2.xyz
CaF_2−4 H_2(vert)	4	0	1	salt	caf-4H2.xyz
MgF_2−1 H_2(vert)	1	0	1	salt	mgf-1H2.xyz
MgF_2−2 H_2(vert)	2	0	1	salt	mgf-2H2.xyz
MgF_2−3 H_2(vert)	3	0	1	salt	mgf-3H2.xyz
MgF_2−4 H_2(vert)	4	0	1	salt	mgf-4H2.xyz
AlF_3−1 H_2(ad)	1	0	1	salt	alf3-1H2.xyz
$CaCl_2$−1 H_2(ad)	1	0	1	salt	cacl-1H2.xyz
$CaCl_2$−2 H_2(ad)	1	0	1	salt	cacl-2H2.xyz
$CaCl_2$−3 H_2(ad)	1	0	1	salt	cacl-3H2.xyz
$CaCl_2$−4 H_2(ad)	1	0	1	salt	cacl-4H2.xyz
CaF_2−1 H_2(ad)	1	0	1	salt	caf-1H2.xyz
CaF_2−2 H_2(ad)	1	0	1	salt	caf-2H2.xyz
CaF_2−3 H_2(ad)	1	0	1	salt	caf-3H2.xyz
CaF_2−4 H_2(ad)	1	0	1	salt	caf-4H2.xyz
MgF_2−1 H_2(ad)	1	0	1	salt	mgf-1H2.xyz
MgF_2−2 H_2(ad)	1	0	1	salt	mgf-2H2.xyz
MgF_2−3 H_2(ad)	1	0	1	salt	mgf-3H2.xyz
MgF_2−4 H_2(ad)	1	0	1	salt	mgf-4H2.xyz
benzene−1 H_2(vert)	1	0	1	organic ligand	benzene-1H2.xyz
phenol−1 H_2(vert)	1	0	1	organic ligand	phenol-1H2.xyz
pyrrole−1 H_2(vert)	1	0	1	organic ligand	pyrrole-1H2.xyz
tetrazole−1 H_2(vert)	1	0	1	organic ligand	tetrazole-1H2.xyz
butene−1 H_2(vert)	1	0	1	organic ligand	butene-1H2.xyz
benzene−1 H_2(ad)	1	0	1	organic ligand	benzene-1H2.xyz
phenol−1 H_2(ad)	1	0	1	organic ligand	phenol-1H2.xyz
pyrrole−1 H_2(ad)	1	0	1	organic ligand	pyrrole-1H2.xyz
tetrazole−1 H_2(ad)	1	0	1	organic ligand	tetrazole-1H2.xyz
butene−1 H_2(ad)	1	0	1	organic ligand	butene-1H2.xyz

Table C.5 continued from previous page

Molecule	# H2	Charge	Multiplicity	Category	Filename
CoF_3−1 H_2(vert)	1	0	5	transition metal	CoF3_M5-1H2.xyz
$CrCl_2$−1 H_2(vert)	1	0	5	transition metal	CrCl2_M5-1H2.xyz
CuCN−1 H_2(vert)	1	0	1	transition metal	CuCN-1H2.xyz
CuOMe−1 H_2(vert)	1	0	1	transition metal	CuOMe-1H2.xyz
$Ni(OH)_2$−1 H_2(vert)	1	0	1	transition metal	NiOH2-1H2.xyz
AuH−1 H_2(vert)	1	0	1	transition metal	AuH-1H2.xyz
AuF−1 H_2(vert)	1	0	1	transition metal	AuF-1H2.xyz
AuCl−1 H_2(vert)	1	0	1	transition metal	AuCl-1H2.xyz
CuH−1 H_2(vert)	1	0	1	transition metal	CuH-1H2.xyz
CuF−1 H_2(vert)	1	0	1	transition metal	CuF-1H2.xyz
CuCl−1 H_2(vert)	1	0	1	transition metal	CuCl-1H2.xyz
AgH−1 H_2(vert)	1	0	1	transition metal	AgH-1H2.xyz
AgF−1 H_2(vert)	1	0	1	transition metal	AgF-1H2.xyz
AgCl−1 H_2(vert)	1	0	1	transition metal	AgCl-1H2.xyz
Sc^+−2 H_2(vert)	2	1	3	transition metal	Sc-2H2.xyz
Sc^+−3 H_2(vert)	3	1	3	transition metal	Sc-3H2.xyz
Sc^+−4 H_2(vert)	4	1	3	transition metal	Sc-4H2.xyz
Ti^+−1 H_2(vert)	1	1	4	transition metal	Ti-1H2.xyz
Ti^+−2 H_2(vert)	2	1	4	transition metal	Ti-2H2.xyz
Ti^+−3 H_2(vert)	3	1	4	transition metal	Ti-3H2.xyz
Ti^+−4 H_2(vert)	4	1	4	transition metal	Ti-4H2.xyz
V^+−2 H_2(vert)	2	1	5	transition metal	V-2H2.xyz
V^+−3 H_2(vert)	3	1	5	transition metal	V-3H2.xyz
V^+−4 H_2(vert)	4	1	5	transition metal	V-4H2.xyz
Cr^+−1 H_2(vert)	1	1	6	transition metal	Cr-1H2.xyz
Cr^+−2 H_2(vert)	2	1	6	transition metal	Cr-2H2.xyz
Cr^+−3 H_2(vert)	3	1	6	transition metal	Cr-3H2.xyz
Cr^+−4 H_2(vert)	4	1	6	transition metal	Cr-4H2.xyz
Mn^+−1 H_2(vert)	1	1	7	transition metal	Mn-1H2.xyz
Mn^+−2 H_2(vert)	2	1	7	transition metal	Mn-2H2.xyz
Mn^+−3 H_2(vert)	3	1	7	transition metal	Mn-3H2.xyz
Mn^+−4 H_2(vert)	4	1	7	transition metal	Mn-4H2.xyz
Fe^+−1 H_2(vert)	1	1	6	transition metal	Fe-1H2.xyz
Fe^+−2 H_2(vert)	2	1	6	transition metal	Fe-2H2.xyz
Fe^+−3 H_2(vert)	3	1	6	transition metal	Fe-3H2.xyz
Fe^+−4 H_2(vert)	4	1	6	transition metal	Fe-4H2.xyz
Co^+−1 H_2(vert)	1	1	3	transition metal	Co-1H2.xyz
Co^+−2 H_2(vert)	2	1	3	transition metal	Co-2H2.xyz
Co^+−4 H_2(vert)	4	1	3	transition metal	Co-4H2.xyz
Ni^+−1 H_2(vert)	1	1	2	transition metal	Ni-1H2.xyz

Table C.5 continued from previous page

Molecule	# H2	Charge	Multiplicity	Category	Filename
Ni^{+}−2 H_2(vert)	2	1	2	transition metal	Ni-2H2.xyz
Ni^{+}−3 H_2(vert)	3	1	2	transition metal	Ni-3H2.xyz
Cu^{+}−1 H_2(vert)	1	1	1	transition metal	Cu-1H2.xyz
Cu^{+}−2 H_2(vert)	2	1	1	transition metal	Cu-2H2.xyz
Cu^{+}−3 H_2(vert)	3	1	1	transition metal	Cu-3H2.xyz
Zn^{+}−1 H_2(vert)	1	1	2	transition metal	Zn-1H2.xyz
Zn^{+}−2 H_2(vert)	2	1	2	transition metal	Zn-2H2.xyz
Zn^{+}−3 H_2(vert)	3	1	2	transition metal	Zn-3H2.xyz
Zn^{+}−4 H_2(vert)	4	1	2	transition metal	Zn-4H2.xyz
CoF_3−1 H_2(ad)	1	0	5	transition metal	CoF3_M5-1H2.xyz
CuCN−1 H_2(ad)	1	0	1	transition metal	CuCN-1H2.xyz
CuOMe−1 H_2(ad)	1	0	1	transition metal	CuOMe-1H2.xyz
AuH−1 H_2(ad)	1	0	1	transition metal	AuH-1H2.xyz
AuF−1 H_2(ad)	1	0	1	transition metal	AuF-1H2.xyz
AuCl−1 H_2(ad)	1	0	1	transition metal	AuCl-1H2.xyz
CuH−1 H_2(ad)	1	0	1	transition metal	CuH-1H2.xyz
CuF−1 H_2(ad)	1	0	1	transition metal	CuF-1H2.xyz
CuCl−1 H_2(ad)	1	0	1	transition metal	CuCl-1H2.xyz
AgH−1 H_2(ad)	1	0	1	transition metal	AgH-1H2.xyz
AgF−1 H_2(ad)	1	0	1	transition metal	AgF-1H2.xyz
AgCl−1 H_2(ad)	1	0	1	transition metal	AgCl-1H2.xyz
Sc^{+}−3 H_2(ad)	1	1	3	transition metal	Sc-3H2.xyz
Sc^{+}−4 H_2(ad)	1	1	3	transition metal	Sc-4H2.xyz
Ti^{+}−1 H_2(ad)	1	1	4	transition metal	Ti-1H2.xyz
Ti^{+}−2 H_2(ad)	1	1	4	transition metal	Ti-2H2.xyz
Ti^{+}−3 H_2(ad)	1	1	4	transition metal	Ti-3H2.xyz
Ti^{+}−4 H_2(ad)	1	1	4	transition metal	Ti-4H2.xyz
V^{+}−3 H_2(ad)	1	1	5	transition metal	V-3H2.xyz
V^{+}−4 H_2(ad)	1	1	5	transition metal	V-4H2.xyz
Cr^{+}−1 H_2(ad)	1	1	6	transition metal	Cr-1H2.xyz
Cr^{+}−2 H_2(ad)	1	1	6	transition metal	Cr-2H2.xyz
Cr^{+}−3 H_2(ad)	1	1	6	transition metal	Cr-3H2.xyz
Cr^{+}−4 H_2(ad)	1	1	6	transition metal	Cr-4H2.xyz
Mn^{+}−1 H_2(ad)	1	1	7	transition metal	Mn-1H2.xyz
Mn^{+}−2 H_2(ad)	1	1	7	transition metal	Mn-2H2.xyz
Mn^{+}−3 H_2(ad)	1	1	7	transition metal	Mn-3H2.xyz
Mn^{+}−4 H_2(ad)	1	1	7	transition metal	Mn-4H2.xyz
Fe^{+}−1 H_2(ad)	1	1	6	transition metal	Fe-1H2.xyz
Fe^{+}−2 H_2(ad)	1	1	6	transition metal	Fe-2H2.xyz
Fe^{+}−3 H_2(ad)	1	1	6	transition metal	Fe-3H2.xyz

Table C.5 continued from previous page

Molecule	# H2	Charge	Multiplicity	Category	Filename
Fe^{+}−4 H_2(ad)	1	1	6	transition metal	Fe-4H2.xyz
Co^{+}−1 H_2(ad)	1	1	3	transition metal	Co-1H2.xyz
Co^{+}−2 H_2(ad)	1	1	3	transition metal	Co-2H2.xyz
Ni^{+}−1 H_2(ad)	1	1	2	transition metal	Ni-1H2.xyz
Ni^{+}−2 H_2(ad)	1	1	2	transition metal	Ni-2H2.xyz
Cu^{+}−1 H_2(ad)	1	1	1	transition metal	Cu-1H2.xyz
Cu^{+}−2 H_2(ad)	1	1	1	transition metal	Cu-2H2.xyz
Cu^{+}−3 H_2(ad)	1	1	1	transition metal	Cu-3H2.xyz
Zn^{+}−1 H_2(ad)	1	1	2	transition metal	Zn-1H2.xyz
Zn^{+}−2 H_2(ad)	1	1	2	transition metal	Zn-2H2.xyz
Zn^{+}−3 H_2(ad)	1	1	2	transition metal	Zn-3H2.xyz
Zn^{+}−4 H_2(ad)	1	1	2	transition metal	Zn-4H2.xyz
H_2(ad)	1	0	1	NA	ad_H2.xyz

Table (C.5) shows all the data points in the H2Bind275 dataset. This table along with the Geometries.zip file which contains the .xyz files for all the geometries can be used to compute all the density functional and ab initio interaction energies. For the vertical interaction energy subset, the binding motif and the H_2(s) are contained in a single file. For example, the file Ca-2H2.xyz contains the geometry of Ca^{2+} binding two molecular hydrogens. The each .xyz file contains the binding motif geometry first followed by each of the bound H_2(s). For the adiabatic interaction energy subset, the interaction energy for the binding motif M bound to n H_2(s) is computed as shown in Eq. (C.1)

$$\Delta E_{\text{int}} = E(\text{M} - \text{nH}_2) - E(\text{M} - (\text{n-1})\text{H}_2) - E(\text{H}_2(\text{ad})) \tag{C.1}$$

H_2(ad) is the relaxed geometry of H_2. For binding motifs containing only one H_2, the relaxed ligand geometry is contained in the file prefixed by "ad_".

Table C.6: Regularized maximum absolute percentage errors for each category of the dataset as well as the total.

DFA	s-block ions	Salts	Organic ligands	Transition metals	Total
SPW92	75.5	95.4	40.9	192.1	192.1
SVWN5	75.4	95.4	40.9	192.0	192.0
PBE	12.5	33.0	26.4	81.4	81.4
RPBE	36.8	89.5	54.6	65.0	89.5
PBE-D3(0)	17.7	19.2	12.9	89.1	89.1
BLYP	46.3	87.2	66.8	68.7	87.2
PW91	18.9	25.2	18.1	86.5	86.5
revPBE	48.0	100.5	65.8	73.7	100.5
revPBE-D3(op)	33.5	20.6	4.4	77.0	77.0
BLYP-D3(op)	25.0	31.6	8.9	72.9	72.9
GAM	25.3	16.4	5.7	264.3	264.3
B97-D3(0)	15.9	41.0	5.9	185.8	185.8
B97-D3(BJ)	19.1	32.5	6.8	166.1	166.1
BP86-D3(BJ)	13.2	19.2	20.5	86.1	86.1
TPSS	18.8	54.5	42.7	71.0	71.0
TPSS-D3(BJ)	18.8	12.3	7.4	77.9	77.9
revTPSS	11.3	40.0	35.9	68.5	68.5
SCAN	11.4	14.9	14.6	76.6	76.6
SCAN-D3(BJ)	10.5	20.5	5.7	77.1	77.1
MS2	14.1	19.3	8.7	84.2	84.2
B97M-V	6.8	11.2	3.3	65.8	65.8
B97M-rV	6.6	11.2	2.0	66.4	66.4
MS2-D3(op)	22.4	26.6	11.5	85.1	85.1
M06-L	31.8	39.0	10.1	136.6	136.6
MN15-L	76.7	66.0	41.5	85.2	85.2
mBEEF	38.4	45.0	26.8	63.6	63.6
oTPSS-D3(BJ)	20.2	25.0	13.8	68.5	68.5
B3LYP	29.7	64.1	51.2	53.2	64.1
PBE0	8.1	31.4	25.8	64.3	64.3
PBE0-D3(BJ)	18.1	13.5	4.5	69.1	69.1
MN15	61.0	44.0	26.3	72.9	72.9
TPSSh	17.5	52.0	41.1	64.9	64.9
TPSSh-D3(BJ)	16.2	12.3	7.5	71.2	71.2
MVSh	28.3	40.0	27.4	38.1	40.0
SCAN0	7.4	15.1	13.3	61.2	61.2
B3LYP-D3(0)	9.9	16.4	3.3	90.6	90.6
M06-2X	19.5	31.1	5.3	103.8	103.8
M06-2X-D3(0)	20.3	32.4	7.3	103.8	103.8
M06	52.1	47.8	7.8	127.5	127.5
revM06	34.5	34.9	10.3	61.4	61.4
ωB97M-V	7.3	10.5	3.3	46.7	46.7
ωB97X-D	55.3	13.5	3.7	41.7	55.3
ωB97X-D3	19.3	13.0	6.8	43.3	43.3
ωB97X-V	8.2	9.6	3.0	47.0	47.0
M11	40.5	36.3	5.8	43.0	43.0
revM11	50.1	50.1	18.5	58.5	58.5
HSE-HJS	8.2	30.2	24.9	65.0	65.0
MN12-SX	27.3	22.3	10.2	80.2	80.2
ωB97M(2)	11.7	29.7	4.5	47.5	47.5
DSD-PBEPBE-D3(BJ)	5.6	15.8	3.1	35.0	35.0
XYG3	11.5	39.1	8.7	31.5	39.1
B2PLYP-D3(BJ)	20.2	10.9	5.3	59.1	59.1
PTPSS-D3(0)	13.8	15.0	4.5	47.0	47.0
PBE0-DH	5.0	19.7	17.9	55.9	55.9
XYGJOS	15.4	16.8	5.6	42.5	42.5

Table C.7: Performance of the density functionals considered in this study on the MGCDB84 database consisting of $\sim$ 5000 data points. This table also shows the number of semi-empirical parameters optimized in the density functionals considered in this study after counting out the number of constraints enforced.

DFA	MGCDB84 rank	Num. of parameters
SPW92	49	3
SVWN5	50	3
PBE	44	0
RPBE	47	0
PBE-D3(0)	40	2
BLYP	48	1
PW91	45	NA
revPBE	46	1
revPBE-D3(op)	NA	5
BLYP-D3(op)	NA	5
GAM	37	26
B97-D3(0)	35	11
B97-D3(BJ)	32	12
BP86-D3(BJ)	39	NA
TPSS	43	0
TPSS-D3(BJ)	30	3
revTPSS	42	1
SCAN	29	0
SCAN-D3(BJ)	27	3
MS2	36	3
B97M-V	14	12
B97M-rV	13	12
MS2-D3(op)	NA	7
M06-L	33	34
MN15-L	34	58
mBEEF	25	64
oTPSS-D3(BJ)	31	8
B3LYP	41	3
PBE0	22	0
PBE0-D3(BJ)	18	3
MN15	12	59
TPSSh	38	1
TPSSh-D3(BJ)	16	4
MVSh	24	0
SCAN0	28	0
B3LYP-D3(0)	20	2
M06-2X	15	29
M06-2X-D3(0)	11	30
M06	21	33
revM06	NA	32
ωB97M-V	2	12
ωB97X-D	9	15
ωB97X-D3	7	16
ωB97X-V	3	10
M11	17	40
revM11	NA	25
HSE-HJS	23	NA
MN12-SX	26	58
ωB97M(2)	1	14
DSD-PBEPBE-D3(BJ)	8	6
XYG3	6	3
B2PLYP-D3(BJ)	10	5
PTPSS-D3(0)	4	10
PBE0-DH	19	0
XYGJOS	5	4

Semi-empirical density functionals can consist of parameters borrowed without further optimization from other functionals, parameters optimized by fitting to the training data, and a some parameters that are set to meet theoretical constraints. In addition to this, some density functionals containing empirical dispersion corrections have their own set of optimized parameters. Table (C.7) counts the number of empirical parameters optimized by fitting to training data after subtracting the number of constraints imposed, thereby only counting the number of free parameters.

Appendix D

Supporting Information: Benchmarking Density Functionals for Predicting Potential Energy Curves

Table D.1: Parameters for treatment of MP2 correlation energy for all double hybrid functionals considered in this work.

Density functional	Density fitting employed?	Level of frozen core approximation
ωB97M(2)	yes	all core electrons
DSD-PBEPBE-D3(BJ)	yes	all core electrons
XYG3	no	all core electrons
B2PLYP-D3(BJ)	yes	no core electrons were frozen
PTPSS-D3(0)	yes	all core electrons
PBE0-DH	no	no core electrons were frozen
XYGJOS	yes	same as the coupled cluster reference

Table D.2: Performance of all 55 density functional approximations (DFAs) for non-equilibrium geometries ($0.75r_{eq}$, $0.9r_{eq}$, $1.1r_{eq}$, $1.25r_{eq}$, and $1.5r_{eq}$)

Rank	DFA	RegMAPE (%)	DFA	RMSE (in kJ/mol)
1	PBE0-DH	4.7	PBE0-DH	2.8
2	ωB97X-V	5.6	DSD-PBEPBE-D3(BJ)	3.9
3	ωB97M-V	6.2	PBE0	4.1
4	XYGJ-OS	6.7	ωB97X-V	4.2
5	DSD-PBEPBE-D3(BJ)	7.0	ωB97X-D	4.2
6	PBE0	7.0	HSE-HJS	4.2
7	HSE-HJS	7.1	MVSh	4.4
8	ωB97M(2)	7.2	XYGJ-OS	4.6
9	B2PLYP-D3(BJ)	8.0	ωB97M-V	4.7
10	XYG3	8.6	PTPSS-D3(0)	4.9
11	SCAN0	8.7	XYG3	4.9
12	PTPSS-D3(0)	9.1	ωB97X-D3	5.0
13	ωB97X-D	9.4	ωB97M(2)	5.1
14	B97M-rV	9.6	PBE0-D3(BJ)	5.1
15	B97M-V	9.7	MN15	5.5
16	ωB97X-D3	9.9	SCAN0	5.8
17	MVSh	10.4	B2PLYP-D3(BJ)	5.8
18	TPSSh	10.5	TPSSh	6.2
19	revTPSS	11.0	revM11	6.5
20	PBE0-D3(BJ)	11.3	mBEEF	6.8
21	revM06	12.0	revM06	7.1
22	M11	12.5	revTPSS	7.6
23	TPSS	12.6	B3LYP	7.7
24	B3LYP-D3(0)	14.0	TPSS	7.8
25	MN15	14.4	B97M-rV	7.8
26	TPSSh-D3(BJ)	14.4	B97M-V	7.8
27	TPSS-D3(BJ)	14.4	B3LYP-D3(0)	7.9
28	oTPSS-D3(BJ)	14.4	oTPSS-D3(BJ)	7.9
29	revM11	14.4	MN15-L	8.3
30	PBE	14.6	TPSS-D3(BJ)	8.5
31	BLYP-D3(op)	15.1	TPSSh-D3(BJ)	8.5
32	B3LYP	15.2	revPBE-D3(op)	8.7
33	SCAN	15.3	MN12-SX	8.9
34	MN12-SX	15.9	revPBE	9.4
35	SCAN-D3(BJ)	16.1	M11	9.5
36	mBEEF	16.2	RPBE	9.6
37	M06	16.3	B97-D3(BJ)	9.8
38	revPBE-D3(op)	16.8	BLYP-D3(op)	10.1
39	MS2	16.9	PBE	10.3
40	PW91	17.0	BLYP	10.4
41	BP86-D3(BJ)	17.8	M06-L	10.5
42	M06-2X	18.4	M06	10.5
43	M06-2X-D3(0)	18.9	BP86-D3(BJ)	10.9
44	M06-L	19.7	B97-D3(0)	11.0
45	MS2-D3(op)	19.8	PBE-D3(0)	11.1
46	PBE-D3(0)	20.0	PW91	11.2
47	MN15-L	20.1	GAM	11.6
48	RPBE	20.8	MS2	12.0
49	BLYP	21.6	MS2-D3(op)	12.2
50	revPBE	22.9	SCAN	13.1
51	GAM	23.9	SCAN-D3(BJ)	13.3
52	B97-D3(BJ)	24.2	M06-2X	14.0
53	B97-D3(0)	27.9	M06-2X-D3(0)	14.0
54	SPW92	62.9	SPW92	32.8
55	SVWN5	63.0	SVWN5	32.9

Table D.3: Performance of density functional approximations (DFAs) with and without empirical dispersion corrections shown with RegMAPE (%), RMSE (in kJ/mol), MSE (in kJ/mol), and their ranking by RegMAPE metric among the 55 DFAs assessed.

Parent DFA					DFA with dispersion				
DFA	RegMAPE	RMSE	MSE	Rank	DFA with dispersion	RegMAPE	RMSE	MSE	Rank
PBE	15.3	10.3	-5.1	29	PBE-D3(0)	20.1	11.0	-7.1	45
revPBE	23.4	9.1	3.8	50	revPBE-D3(op)	17.0	8.5	-3.5	37
BLYP	22.1	10.1	2.4	49	BLYP-D3(op)	15.4	9.8	-2.0	31
TPSS	13.6	7.7	-1.0	23	TPSS-D3(BJ)	14.7	8.5	-4.2	27
SCAN	15.6	12.9	-6.7	33	SCAN-D3(BJ)	16.3	13.2	-7.3	35
MS2	17.4	12.1	-7.1	39	MS2-D3(op)	20.3	12.4	-7.8	46
B3LYP	15.5	7.4	2.7	32	B3LYP-D3(0)	14.0	7.5	-0.8	25
PBE0	7.6	4.2	-1.3	7	PBE0-D3(BJ)	11.5	5.2	-3.5	19
TPSSh	11.3	6.0	0.1	18	TPSSh-D3(BJ)	13.8	8.0	-3.7	24
M06-2X	19.3	13.4	0.6	42	M06-2X-D3(0)	19.8	13.4	0.5	43

Table D.4: Performance of all 55 density functional approximations (DFAs) for predicting H_2 binding energy at equilibrium geometry shown using the adiabatic regularized mean absolute percentage error ($RegMAPE_{ad}$).

Rank	DFA	$RegMAPE_{ad}$ (%)
1	ωB97X-V	4.7
2	DSD-PBEPBE-D3(BJ)	4.7
3	PBE0-DH	5.3
4	ωB97M-V	6.0
5	XYGJ-OS	7.1
6	B97M-rV	7.2
7	B97M-V	7.2
8	ωB97M(2)	7.4
9	ωB97X-D	7.8
10	XYG3	7.9
11	B2PLYP-D3(BJ)	8.3
12	PBE0	8.4
13	HSE-HJS	8.4
14	ωB97X-D3	9.5
15	SCAN0	9.7
16	MVSh	9.9
17	PTPSS-D3(0)	10.4
18	M11	11.1
19	PBE0-D3(BJ)	11.7
20	TPSSh-D3(BJ)	12.0
21	TPSSh	12.3
22	revTPSS	13.5
23	revM06	13.7
24	oTPSS-D3(BJ)	14.4
25	TPSS	14.6
26	TPSS-D3(BJ)	14.8
27	BLYP-D3(op)	15.6
28	B3LYP	15.7
29	B3LYP-D3(0)	16.1
30	PBE	16.3
31	SCAN	16.4
32	SCAN-D3(BJ)	17.2
33	revPBE-D3(op)	17.3
34	MN12-SX	17.3
35	MN15	17.5
36	PW91	17.9
37	revM11	17.9
38	RPBE	18.1
39	MS2	19.1
40	M06	19.3
41	BP86-D3(BJ)	19.4
42	mBEEF	19.9
43	MN15-L	20.1
44	PBE-D3(0)	20.6
45	revPBE	21.6
46	BLYP	21.6
47	MS2-D3(op)	22.0
48	M06-2X	22.3
49	M06-2X-D3(0)	22.6
50	M06-L	24.1
51	B97-D3(BJ)	24.4
52	GAM	26.5
53	B97-D3(0)	29.4
54	SPW92	73.9
55	SVWN5	73.9