Title
Molecular Orbitals and Where to Find Them

Author
Aldossary, Abdulrahman N.

Publication Date
2022

Peer reviewed|Thesis/dissertation

Molecular Orbitals and Where to Find Them

by

Abdulrahman N. Aldossary

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 Head-Gordon, Chair
Associate Professor Eric Neuscamman
Professor Kristin Persson

Fall 2022

Molecular Orbitals and Where to Find Them

Abstract

Molecular Orbitals and Where to Find Them

by

Abdulrahman N. Aldossary

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Martin Head-Gordon, Chair

In this thesis, new ways of generating orbitals to accelerate quantum mechanical calculations or gain chemical insight from such calculations are presented. While antibonding orbitals are integral to our understanding of chemistry, finding them is not trivial because they cannot be uniquely defined. A new way to find them is demonstrated along with many uses they have in valence space methods such as complete active space self-consistent field (CASSCF) theory. New molecule-adapted atomic orbitals can also be obtained from the antibonding orbitals which is used for atomic charge analysis. Furthermore, new oxidation state localized orbitals (OSLOs) for assigning oxidation states are also presented and assessed across a wide range of systems to demonstrate improved performance on challenging systems. Lastly, an approach to decompose the intermolecular forces based on the absolutely localized molecular orbital energy decomposition analysis (ALMO-EDA) is described. This force decomposition analysis (FDA) separates different physically distinct contributions (such as those from permanent electrostatics and Pauli repulsion, those from the effect of induced electrostatics, and those from charge transfer or dative interactions). A series of chemical examples ranging from ion water interactions to activation of carbon dioxide by gold (and silver) anions are studied to explore the chemical insight that can be gained. The FDA may also be useful for the development and testing of molecular mechanics force fields. As an example, we use force decomposition to further validate the promising polarizable MB-UCB force field for water.

To My Parents

Contents

List of Figures

List of Tables

Acknowledgments

Foremost, I thank my advisor, Martin Head-Gordon, for mentoring me throughout graduate school and helping me be the scientist I have become today. Our interactions and scientific discussions taught me to appreciate how difficult it is to do good science and how important it is to stay curious. Thank you for helping me in every step during my Ph.D., for working hard on securing the research funds, and for giving me the freedom to explore research in many topics. I could never ask for more, and for that I am indebted. Imitation is the sincerest form of flattery, and many members of the group entrench the work-life balance and aspire to be as diligent yet patient and respectful to others as we have learned from Martin. I am very proud to be inducted into the long list of scientists trained under his supervision.

I also would like to acknowledge the many people I have worked with during my Ph.D. time, which there are too many to include here. Drs. Luke Bertels, Daniel Levine, and Norman Tubman helped me get acclimated in the research group and taught many of the basics. Dr. Yuezhi Mao was patient enough to help me a lot with my first project as well as throughout my graduate school years and taught me many intricacies of energy decomposition analysis. Dr. Susi Lehtola sat with me for many hours and taught me many subtleties of valence bond theory. And of course, thanks to Dr. Diptarka Hait for answering my many many questions on electronic structure and being always happy to talk about density functional theory at any time of the day or week. Our weekly meetings with Dr. Akshaya Das and Prof. Teresa Head-Gordon taught me everything I know about force field methods. This list cannot be complete without thanking (soon-to-be Doctor) Martí Gimferrer and Prof. Pedro Salvador for teaching me everything I know about oxidation state methods and for working closely with me on the oxidation state work.

This would not have been possible without such a friendly, helpful, and collaborative environment. There are many more people who taught me the various technicalities of quantum chemistry such as Dr. Srimukh Prasad, Brad Ganoe, Prof. Tim Neudecker, Dr. Jérôme Gonthier, Dr. David Small, Prof. Farnaz Heidar-Zadeh, Prof. Christopher Stein, and Dr. Alex Zech. I would also like to thank Prof. Eric Neuscamman for being my qualifying exam chair, for advising me throughout graduate school, and for carefully reading every word of this thesis.

Many people around me kept a high spirit of camaraderie, which made graduate school much more enjoyable. This included many people from the research group and the Gilman basement which there are too many to include. I have enjoyed the long conversations I had with Diptarka Hait about history, politics, and science, the long bike trips I had with Adam Rettig and Juan Arias-Martinez, and the daily afternoon coffee walks as well as the nightly tipsy conversations with the group. Jonathan Wong, Bryan Place, and David Lunderberg kept me sane throughout these years with our many adventures and road trips, for which I am very grateful.

I would not have been able to do this without the endless love and support from my wife, Raida. Thank you for all the years we have spent together, and I am excited for many more

to come. I am also very grateful to my family for their continuous support throughout my academic career. I hope I will make you proud one day. I have always looked forward to visiting home and spending time with you all. I hope I will be able to do so much more in the future.

Chapter 1

Introduction

1.1 Quantum Mechanics Review

Quantum chemistry involves making predictions (energies, molecular geometries, and other observables) about molecules using the principles of quantum mechanics. Here we will go through the relevant principles and their ramifications in chemistry.

Principles of Quantum Mechanics

In the early 20th century, experiments such as the double-slit experiment, photoelectron effect, and Stern-Gerlach made it clear that on the atomic scale, particles behave differently than predicted by classical mechanics. Most notably, the particle-wave duality in quantum mechanics means that one needs a wavefunction to describe particles. These wavefunctions need to obey other quantum mechanical postulates, such as:

1. The wavefunction often denoted by Ψ describes the state of the system completely.
2. Observables are described by Hermitian operators, $\hat{O}$. The expectation value of measuring operator $\hat{O}$ for a system in state Ψ is $\int \Psi^* \hat{O} \Psi$.
3. The position and momentum operators are $\hat{x} = x$ and $\hat{p} = -i\hbar\nabla$, respectively, in the position representation, where $\hbar$ is the reduced Planck's constant.
4. The possible outcomes of measuring a Hermitian operator are the eigenvalues of the operator, $\hat{O}\Psi = \epsilon_i \Psi$. Upon a measurement of Ψ yielding ϵ_i, the wavefunction collapses into the corresponding eigenstate.
5. Wavefunctions with Hamiltonian $\hat{H}$ evolve in time according to the Schrödinger equation, $i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi$.
6. Exchange of two fermionic particles in the wavefunction results in a change of sign of the wavefunction.

Based on the third postulate, other operators are taken from classical mechanics and directly used in quantum mechanics. For example, the classical kinetic energy, $p^2/2m$, is now an operator $\hat{T} = \hat{p}^2/2m$. The kinetic energy of a particle with some known wavefunction is $\int \Psi^* \hat{T} \Psi$. Another relevant quantity is the electric potential energy between two charged particles given by $\hat{V} = q_1 q_2 / r 4\pi\epsilon_0$, where q_1 and q_2 are the charges of the two particles, r is the distance between them, and ϵ_0 is the vacuum permittivity. These two operators lay the groundwork for the forthcoming sections.

Molecular Hamiltonian

The molecular Hamiltonian operator contains all the energy terms in a molecule. It is simply the sum of the kinetic and electric potential energies of nuclei and electrons and is given by

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{ne} + \hat{V}_{nn}, \tag{1.1}$$

where $\hat{T}_e$ and $\hat{T}_n$ are the kinetic energy of the electrons and nuclei, respectively, and $\hat{V}_{ee}$, $\hat{V}_{ne}$, and $\hat{V}_{nn}$ are the electron-electron, electron-nuclei, and nuclei-nuclei electric potential energy, respectively. The fourth postulate stipulates that solving the Hamiltonian of a system means finding all the possible energies and states of that system. The solution for the eigenvectors and eigenvalues of Eq. 1.1 solves all of chemistry and is why the famous physicist Paul Dirac in 1929 declared [1]:

> "the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

It has since been the duty of chemists to study these equations, their approximations, and their consequences.

Hydrogen Atom and Atomic Units

The simplest form of the Eq. 1.1 is for a system consisting of one electron and one nucleus. Transforming the nucleus' coordinates into the frame of reference makes the Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2\mu}\nabla^2 - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r}, \tag{1.2}$$

where e is the elementary charge of an electron or a proton, and μ is the reduced mass which is half the harmonic mean or inverse of the sum of mass inverses. Note that we are ignoring relativity, as it will not be discussed in this thesis. We shall introduce another form of Eq. 1.2 using atomic units before moving forward. In atomic units, we set $\hbar$, e, m_e, and

the Bohr radius, $a_0 = 4\pi\epsilon_0\hbar^2/m_e e^2$ to unity. Using this system of units is very effective in writing equations, and Eq. 1.2 for instance is now

$$\hat{H} = -\frac{1}{r} - \frac{1}{2}\nabla^2. \tag{1.3}$$

The solution to the time-independent Schrödinger equation of the hydrogen and hydrogen-like atom is well known and discussed in many standard quantum mechanics textbooks [2]. In brevity, we show the general form of the solution,

$$\psi_{n,l,m}(r,\theta,\phi) = C_{n,l} r^l L_{n-l-1}^{2l+1}(r) e^{-r/na_0} \mathbf{Y}_l^m(\theta,\phi) \tag{1.4}$$

where n, l, and m are the principal, angular, and magnetic quantum numbers, $C_{n,l}$ is some normalization constant, $L_{n-l-1}^{2l+1}(r)$ is the generalized Laguerre polynomial, and $\mathbf{Y}_l^m(\theta,\phi)$ is a spherical harmonic.

Born-Oppenheimer Approximation

Up to this point, the Schrödinger equation for any molecule is

$$\big(-\sum_i \frac{1}{2}\nabla_i^2 - \sum_A \frac{1}{2m_A}\nabla_A^2 + \sum_{i>j}\frac{1}{r_{ij}} - \sum_{i,A}\frac{Z_A}{r_{iA}} + \sum_{A>B}\frac{Z_A Z_B}{r_{AB}}\big)\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R}) \tag{1.5}$$

where i and j are indices of electrons, A and B are indices of nuclei, m_A is the mass of atom A, r_{xy} is the euclidean distance between the coordinates of particles x and y, Z_A is the nuclear charge of atom A, $\mathbf{r}$ and $\mathbf{R}$ correspond to positions of electrons and nuclei, respectively, and E is the energy eigenvalue of the system. The solution to this partial differential equation gives us the energy, E, and the state of any molecular system, $\Psi(\mathbf{r},\mathbf{R})$, but the solution is far too difficult. Invoking the Born-Oppenheimer approximation allows for factorizing the wavefunction into electronic and nuclear parts [3]. This is justified because the forces acting on electrons and nuclei are similar, yet they have masses that are different by more than three orders of magnitude. This vast difference in their masses is reflected in their velocities, hence why we treat nuclei as fixed charges compared to electrons.

Under the Born-Oppenheimer approximation, we consider the solution for the electronic part of the wavefunction where the nuclei are at fixed positions then consider the nuclear wavefunction for a fixed electronic wavefunction. The first part of this yields the electronic Schrödinger equation,

$$\big(-\sum_i \frac{1}{2}\nabla_i^2 + \sum_{i>j}\frac{1}{r_{ij}} - \sum_{i,A}\frac{Z_A}{r_{iA}}\big)\Psi(\mathbf{r};\mathbf{R}) = E_{\text{elec}}(\mathbf{R})\Psi(\mathbf{r};\mathbf{R}) \tag{1.6}$$

where $\Psi(\mathbf{r};\mathbf{R})$ is the electronic wavefunction for a given set of (clamped) coordinates of the nuclei, referred to hereafter as $\Psi(\mathbf{r})$ or just Ψ. Changing the nuclear coordinates and evaluating the energy produces hypersurfaces, often called potential energy surfaces in chemistry.

Given the electronic wavefunctions of the system, one can evaluate the vibrational and rotational energies and states of the system. The study of these quantities is important for spectroscopic measurements and thermodynamics but is not relevant to this thesis. The potential energy surfaces define adiabatic states that can intersect each other. When two surfaces come close in energy, the adiabatic approximation breaks down [3] for example in the stretching of the LiF molecule. Treatments of nonadiabatic effects will not be discussed in this thesis as well.

Slater Determinant

Consider a permutation operator, $\hat{P}_{12}$ that swaps the coordinates of two electrons in a two-electron wavefunction, $\Psi(\mathbf{r}_1, \mathbf{r}_2)$. Since the particles are identical, swapping the dummy indices does not change the Hamiltonian eigenstate. That is, the permutation operator commutes with the Hamiltonian, $[\hat{H}, \hat{P}_{12}] = 0$. Thus, the action of $\hat{P}_{12}$ on some eigenstate of the Hamiltonian,

$$\hat{P}_{12}\Psi(\mathbf{r}_1, \mathbf{r}_2) = \lambda\Psi(\mathbf{r}_2, \mathbf{r}_1), \tag{1.7}$$

where λ is some unknown eigenvalue. The repeated action of this operator recovers the original wavefunction,

$$\hat{P}_{12}\lambda\Psi(\mathbf{r}_2, \mathbf{r}_1) = \lambda^2\Psi(\mathbf{r}_1, \mathbf{r}_2) = \Psi(\mathbf{r}_1, \mathbf{r}_2). \tag{1.8}$$

Hence, λ^2 has to be equal to unity, with two possible solutions, $\lambda = \pm 1$. Particles with $\lambda = 1$ are bosons and have symmetric wavefunctions, whereas particles with $\lambda = -1$ are fermions and have antisymmetric wavefunctions, following from the sixth postulate discussed above. Most relevant here, electrons are fermions and can have wavefunctions of the form

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}(\psi_A(\mathbf{r}_1)\psi_B(\mathbf{r}_2) - \psi_A(\mathbf{r}_2)\psi_B(\mathbf{r}_1)). \tag{1.9}$$

Notice that this form produces a negative sign upon permutation of the coordinates. Also notice that the wavefunction vanishes when $\psi_A = \psi_B$, i.e., two electrons occupy the same space, which is colloquially known as the Pauli exclusion principle. The product of the two single-electron wavefunctions, $\psi_A\psi_B$, is often called a Hartree product.

The action of a permutation operator of two electrons on an arbitrary many-electron wavefunction produces a change of sign, i.e.,

$$\hat{P}_{ij}\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_j, \cdots, \mathbf{r}_n) = -\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_j, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_n). \tag{1.10}$$

Taking the sum of all possible permutations of a system gives the antisymmetrizer

$$\mathcal{A} = \frac{1}{N!}\sum_{\pi \in S_N} \hat{P}_\pi, \tag{1.11}$$

where S_N is the group of all permutations for N particles and π is an element of the group. Applying the antisymmetrizer operator on a Hartree product of many spin-orbitals (wavefunction with position and spin, vide infra) produces a Slater determinant which enforces the antisymmetry. Slater determinants are the simplest form of an antisymmetric (fermionic) many-electron wavefunction and describe one configuration of the electrons and are referred to as a "configuration" [4]. They take the form

$$\Phi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \ldots & \chi_1(\mathbf{x}_N) \\ \chi_2(\mathbf{x}_1) & \chi_2(\mathbf{x}_2) & \ldots & \chi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(\mathbf{x}_1) & \chi_N(\mathbf{x}_2) & \ldots & \chi_N(\mathbf{x}_N) \end{vmatrix}, \tag{1.12}$$

where $\chi(\mathbf{x}_i)$ is a spin-orbital. These are defined to be

$$|\chi_i(\mathbf{x})\rangle = |\psi_i(\mathbf{r})\sigma_i(\omega)\rangle\,, \tag{1.13}$$

where $\psi_i(\mathbf{r})$ are spatial orbitals with coordinate $\mathbf{r}$ that spans the real space, i.e., $\mathbf{r} \in \mathcal{R}^3$, while σ_i are spin functions with coordinate ω, spanning two values: α and β, i.e., $\omega \in \{\alpha, \beta\}$. It is then understood that coordinate $\mathbf{x}$ in Eq. 1.13 spans both spatial and spin coordinates. The spin function, $|\sigma(\omega)\rangle$, are orthogonal with respect to their argument, i.e.,

$$\langle \sigma(\omega_i) | \, \sigma(\omega_j) \rangle = \begin{cases} 0 & \text{if } \omega_i \neq \omega_j \\ 1 & \text{otherwise.} \end{cases} \tag{1.14}$$

1.2 Wavefunction-Based Methods

Hartree-Fock

The electronic Schrödinger equation in Eq. 1.6 for more than one electron is very difficult to solve and no analytical solution exists to date. One often reverts to mean-field approximations where the exact electron-electron interaction is approximated. Consider a trial wavefunction (ansatz) that is a Slater determinant and minimize the expectation value of the energy, i.e.,

$$E_{\text{HF}} = \min_{\mathbf{C}} \langle \Phi_0(\mathbf{C}) | \sum_i -\frac{1}{2}\nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} |\Phi_0(\mathbf{C})\rangle\,, \tag{1.15}$$

where $\mathbf{C}$ is the coefficient matrix describing the spin-orbitals of the trial wavefunction. Note that the minimization must be performed subject to the orthonormality constraint, $\langle \chi_i | \, \chi_j \rangle = \delta_{ij}$. The first two operators in Eq. 1.15 are one-electron operators, and, combined, they are called the core Hamiltonian, $\hat{h}$. The last operator is a two-electron operator of the electron-electron interaction.

Evaluating a one-electron operator for a Slater determinant, such as the core Hamiltonian, gives [4]

$$\langle \Phi_0 | \, \hat{h} \, | \Phi_0 \rangle = \sum_i \langle \chi_i | \, \hat{h} \, | \chi_i \rangle = \sum_i \mathbf{h}_{ii}, \tag{1.16}$$

where the $\mathbf{h}_{ii}$ is a diagonal element of the core Hamiltonian. Meanwhile, two-electron operators evaluate to [4]

$$\langle \Phi_0 | \sum_{i>j} \frac{1}{r_{ij}} \, | \Phi_0 \rangle = \sum_{i>j} \langle \chi_i(x_1)\chi_j(x_2) | \, \frac{1}{r_{12}}(1 - \hat{P}_{12}) \, | \chi_i(x_1)\chi_j(x_2) \rangle = \frac{1}{2} \sum_i \mathbf{J}_{ii} - \mathbf{K}_{ii}, \tag{1.17}$$

where the $\hat{P}_{12}$ is the permutation operator introduced earlier, and it shows up here due to the antisymmetry of the trial wavefunction, Φ_0. Furthermore, $\mathbf{J}$ is the Coulomb matrix which describes a simple charge density interaction with another, while $\mathbf{K}$ is called the exchange matrix due to exchanging the two coordinates of the ket side. Both $\mathbf{J}$ and $\mathbf{K}$ matrices describe the potential of the trial wavefunction and are a function of it. Evaluating the energy of one Slater determinant using Eq. 1.16 and Eq. 1.17 produces

$$\langle \Phi_0(\mathbf{C}) | \, \hat{h} + \sum_{i>j} \frac{1}{r_{ij}} \, | \Phi_0(\mathbf{C}) \rangle = \sum_i \mathbf{h}_{ii} + \frac{1}{2}\big(\mathbf{J}_{ii} - \mathbf{K}_{ii}\big). \tag{1.18}$$

For a Slater determinant, the energy associated with one electron in orbital χ_i can be calculated with the Fock operator, $\hat{F}$,

$$\langle \chi_i | \, \hat{F} \, | \chi_i \rangle = \mathbf{F}_{ii}, \tag{1.19}$$

where $\mathbf{F}_{ii}$ is the diagonal element of the Fock operator in matrix form, where the Fock matrix is

$$\mathbf{F} = \mathbf{h} + \mathbf{J} - \mathbf{K}. \tag{1.20}$$

Finding the eigenstates of the Fock operator (the mean-field Hamiltonian operator) constitutes the eigenvalue problem:

$$\hat{F} \, | \chi_i \rangle = \epsilon_i \, | \chi_i \rangle \tag{1.21}$$

where χ_i is the eigenstate (eigenvector) or simply the orbital, and ϵ_i is the orbital energy (eigenvalue). To solve the eigenvalue problem in Eq. 1.21, one can project it onto a set of basis functions (Galerkian projection), producing the Roothaan-Hall equations,

$$\mathbf{F}(\mathbf{C})\mathbf{C} = \mathbf{SCE}, \tag{1.22}$$

where $\mathbf{F}$ is the Fock matrix evaluated for the basis functions, $\mathbf{C}$ is the coefficient matrix (eigenvector matrix) describing the molecular orbitals in the basis functions, $\mathbf{S}$ is the overlap between the basis functions, and $\mathbf{E}$ is a diagonal matrix of orbital energies (eigenvalues). Once the generalized eigenvalue problem is solved, the orbitals are filled in ascending order of their energies following the aufbau principle, creating the occupied space, while the empty

ones are called virtual orbitals. The occupied orbitals from the new coefficient matrix, $\mathbf{C}$, constitute the new trial wavefunction, $\Phi_0(\mathbf{C})$, and the procedure is repeated until the density does not change. Although Eq. 1.22 is a generalized eigenvalue problem, the Fock matrix $\mathbf{F}(\mathbf{C})$ is a function of $\mathbf{C}$, and only upon self-consistency, the total energy, E_{HF}, is made stationary. This is called the self-consistent field or the Hartree-Fock (HF) method in quantum chemistry [4–6].

The HF method is a simple self-consistent method that minimizes the energy of one Slater determinant. It is often used as a starting point for more advanced methods and is very important in quantum chemistry. Most importantly, it obeys the variational principle, i.e.,

$$\mathcal{E}_{\text{HF}} \geq \mathcal{E}. \tag{1.23}$$

In other words, the HF energy cannot be lower than the ground-state electronic energy. In fact, the difference between them was termed the *correlation energy* by Löwdin [7] and is a non-positive quantity

$$\mathcal{E}_{\text{corr}} = \mathcal{E} - \mathcal{E}_{\text{HF}} \leq 0. \tag{1.24}$$

To recover the correlation energy, consideration of more than one Slater determinant is required. In particular, once an HF wavefunction is obtained, excited determinants can be constructed by the promotion of an electron from an occupied orbital (indexed $i, j, k, ...$) into a virtual (empty) orbital (indexed $a, b, c, ...$). These substituted or excited Slater determinants are commonly denoted by the notation $|\Phi_{ijk...}^{abc...}\rangle$ to describe electrons occupying orbitals $ijk...$ being promoted to virtual orbitals $abc...$. Then, the correlation energy can be recovered either by perturbation theory, exact diagonalization, or other methods that will be discussed further below.

Basis Sets

To produce the generalized eigenvalue problem in Eq. 1.22, a set of basis functions, i.e. a finite basis set, must be introduced. The basis functions must be capable of describing any electron in the Hilbert space which is infinite. Because electrons in molecules 'feel' the field of multiple atoms, a common choice for chemists is a linear combination of atomic orbitals (LCAO). The simplest form of the atomic orbitals (AOs) is Slater-type orbitals STOs which are the hydrogen-like orbitals introduced in Eq. 1.4 but without the Laguerre polynomials. Using a linear combination of Gaussian-type orbitals (GTOs), to describe one Slater orbital has proven to be very efficient, especially for evaluating matrix elements of $\mathbf{J}$ or $\mathbf{K}$ from Eq. 1.17 since the multiplication of two Gaussian orbitals yields another Gaussian orbital [4]. Gaussian-type orbitals have the form

$$\psi_{n,l,m}(\mathbf{r}, \theta, \phi) = \sum_k c_k r^l e^{-\zeta_k |\mathbf{r}-\mathbf{R}_A|^2} \mathbf{Y}_l^m(\theta, \phi), \tag{1.25}$$

where c_k and ζ_k are optimizable parameters and $\mathbf{R}_A$ is the coordinate of the atom that the basis function is centered on. Many techniques have been developed for producing the set of ζ_k's and c_k's to reproduce the electronic energy efficiently [8].

The set of atomic orbitals that can describe simple bonding includes each atom's outermost valence shell, often called the minimal basis set. Since the early days of quantum chemistry, it has been noted that atomic orbitals change upon bonding and named the modified atomic orbitals by Mulliken [9, 10]. Atoms with stronger electronegativity attract electrons more and result in orbitals polarized towards them. Hence, even for the HF method, many more basis functions are required beyond the minimal basis set to accurately describe the wavefunction. To capture that polarization, basis sets often include the next shell of atomic orbitals, often called double-ζ (DZ), or N more additional shells, N-ζ. The Hartree-Fock energy converges with the exponential of the cardinal number of extra valence shells in the basis, $\sim e^{-N}$, while the correlation energy converges with the inverse of the cube, $\sim N^{-3}$ [11, 12].

Many basis sets (or sets of GTOs) have been optimized and made to allow custom inclusion of different parts of the Hilbert space as necessitated by the chemistry in question. For example, calculations involving core electrons require the addition of extra core (very large ζ) basis functions, while long-range interactions benefit from the addition of diffuse (very small ζ) basis functions. This customizability allows chemists to achieve the highest accuracy with the lowest computational cost. Although GTOs are advantageous in many ways, they are not required. Plane waves [13] (eigenfunctions of the kinetic energy operator) or real-space [14] methods are not uncommon choices of basis functions. All of these incomplete bases converge upon expansion to the complete basis set (CBS) limit.

Two dimensions of the accuracy of approximate electronic energy have been discussed. To get the exact energy, one needs not only to use methods capable of capturing more correlation but also one needs to use large basis sets to recover the exact result. The two dimension graph of accuracy is often referred to as Pople's diagram [15] and is shown in Fig. 1.1. This leaves the third dimension for correction to some of the approximations already made, such as relativistic correction, the Born-Oppenheimer approximation, or reserving the third dimension for time propagation of the wavefunction or nuclei. To estimate the exact answer (top right of Fig. 1.1), composite methods popularized by Pople and coworkers exploit energy differences between methods of increasing correlation accuracy with decreasing size of the basis set [16]. For example, the "composite method" in Fig. 1.1 uses the energy of the HF in the QZ basis, MP2 in the TZ basis, and CCSD in the DZ basis to estimate the full CI energy, where these correlation methods are discussed in detail below.

Full Configuration Interaction

To recover the correlation energy for a given basis, we minimize the energy of Ψ that is a linear combination of *all* other substituted determinants, producing

$$\mathbf{HC} = \mathbf{CE} \tag{1.26}$$

where the matrix $\mathbf{H}$ is known as the configuration interaction (CI) Hamiltonian with matrix elements,

$$\mathbf{H}_{pq} = \langle \Phi_p | \, \hat{H} \, | \Phi_q \rangle \tag{1.27}$$

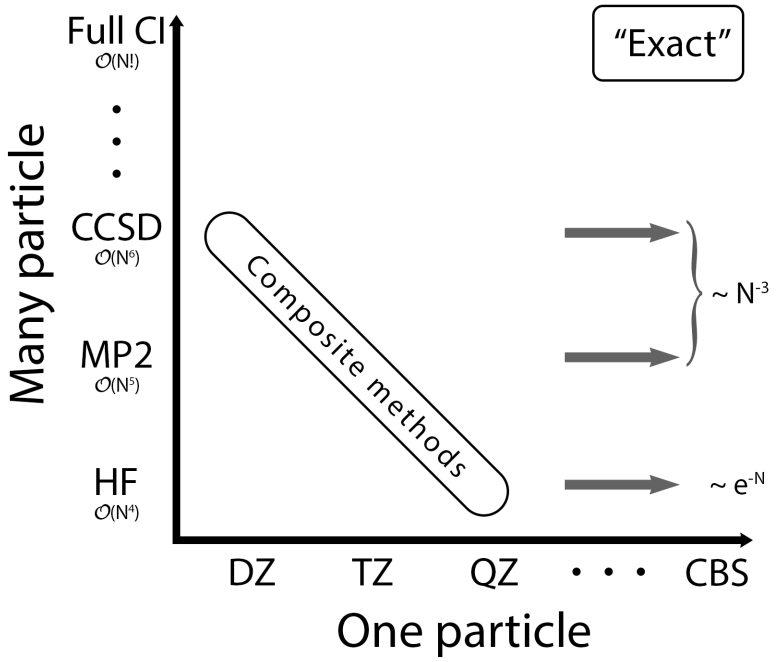


Figure 1.1: Pople's diagram where the x-axis is the basis set ("one particle") and the y-axis is the correlation energy recovered ("many particle") with the scaling of each method indicated underneath it. The horizontal arrows indicate the convergence with the basis set, where N is the cardinal number of the basis set. The "Exact" in the top-right corner denotes the exact non-relativistic Born-Oppenheimer electronic energy.

where Φ_p and Φ_q are determinants. In Eq. 1.26, $\mathbf{C}$ is a matrix with a coefficient for each true state of the exact Hamiltonian, with each column called the CI vector. The matrix $\mathbf{E}$ in Eq. 1.26 is a diagonal matrix with exact eigenenergies, in which the lowest is the exact ground state energy. This method of exact diagonalization is called the full configuration interaction (full CI) method. Taking the eigenvector of $\mathbf{C}$ with the lowest eigenvalue allows us to express the ground state wavefunction as a linear combination of determinants,

$$\Psi = c_0\Phi_0 + \sum_{ia} c_{ia}\Phi_i^a + \sum_{ijab} c_{ijab}\Phi_{ij}^{ab} + \cdots . \tag{1.28}$$

Notice that the choice of constructing excited determinants from the same orbitals is not required, but preferable to maintain orthogonality between all excited determinants. Breaking that assumption gives rise to a generalized eigenvalue problem [17] as the overlap will be needed in Eq. 1.26.

How many elements are in the matrix $\mathbf{H}$ like the one shown in Eq. 1.27? The answer is too many. For a system with M orbitals and N electrons, it is possible to construct $\binom{M}{N}$

determinants, which scales factorially with the system size. Luckily, many of these terms evaluate to zero trivially by the Slater-Condon rules [18, 19], which include $\mathbf{H}_{pq} = 0$ if the two determinants differ by more than two orbitals since our Hamiltonian couples two particles at most. The size of the $\mathbf{H}$ matrix, however, is still factorial. As a consequence, for a given AO basis set, the compute effort to solve the full CI problem grows exponentially with the molecule size. This unfortunate fact of nature gives rise to many interesting endeavors to solve the many-body problem.

Truncated Configuration Interaction

The simplest and earliest solution to improve upon the HF energy is to consider a small set of excitations. For example, it is possible to include only single and double excitations, named configuration interaction singles and doubles (CISD) [4, 6, 11]. Although such treatment is variational, they suffer from a lack of size consistency. A method is called size consistent when it recovers the sum of individual molecules' energies upon infinite separation. A CISD wavefunction overentangles two molecules by requiring one molecule to be fully unexcited for the other molecule to be doubly excited, for example. This is often resolved by adding some quadruply excited determinants as in QCISD [20]. However, other methods were introduced that capture more correlation with similar cost [21], surpassing truncated CI methods in use.

Møller-Plesset Perturbation Theory

All hope is not lost. Perturbation theory is a very effective approach capable of recovering some of the correlation energy. Consider a small perturbation, λ, to the Hamiltonian, i.e.,

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}', \tag{1.29}$$

which modifies the eigenvectors and eigenvalues of $\hat{H}_0$ slightly, such that

$$\big(\hat{H}_0 + \lambda \hat{H}'\big)\big(\Psi^{(0)} + \lambda \Psi^{(1)} + \cdots\big) = \big(E^{(0)} + \lambda E^{(1)} + \cdots\big)\big(\Psi^{(0)} + \lambda \Psi^{(1)} + \cdots\big). \tag{1.30}$$

Collecting powers of λ, gives the following sets of equations

$$\hat{H}_0 \Psi^{(0)} = E^{(0)} \Psi^{(0)} \tag{1.31a}$$

$$\hat{H}_0 \Psi^{(1)} + \hat{H}' \Psi^{(0)} = E^{(0)} \Psi^{(1)} + E^{(1)} \Psi^{(0)} \tag{1.31b}$$

$$\hat{H}_0 \Psi^{(2)} + \hat{H}' \Psi^{(1)} = E^{(0)} \Psi^{(2)} + E^{(1)} \Psi^{(1)} + E^{(2)} \Psi^{(0)} \tag{1.31c}$$

$$\vdots$$

$$\hat{H}_0 \Psi^{(\mathrm{n})} + \hat{H}' \Psi^{(\mathrm{n}-1)} = \sum_{k=0}^{n} E^{(\mathrm{k})} \Psi^{(\mathrm{n}-\mathrm{k})}, \tag{1.31d}$$

showing up to the second-order expansion explicitly and the general form in Eq. 1.31d. Left projection by different orders of the wavefunction, $\Psi^{(\mathrm{n})}$, allows solving the perturbation series in ascending order, with increasing corrections of the wavefunction and the energy.

So far, this formalism is the Rayleigh-Schrödinger perturbation theory. In the many-body perturbation theory, we choose the 0th order wavefunction as the HF wavefunction, i.e., $\Psi^{(0)} = \Phi_0$, while higher order wavefunctions are to be the excited determinants. The Møller-Plesset perturbation theory [22] makes a special choice of the 0th order Hamiltonian, $\hat{H}_0$, to be the Fock operator, $\hat{F}$, of Eq. 1.19. This choice is very important as it makes the method size extensive, yet problematic since the perturbation is not small for this formalism to be suitable. The $\hat{H}'$ of Eq. 1.29 following from that choice becomes

$$\hat{H}' = \hat{H} - \hat{F}. \tag{1.32}$$

Using Eq. 1.31a, the zeroth order of the energy is

$$E^{(0)} = \langle \Phi_0 | \hat{F} | \Phi_0 \rangle = \sum_i \epsilon_i, \tag{1.33}$$

which is higher than the HF energy since it double counts the electron-electron repulsion. Using the next power in Eq. 1.31b gives the first order of the energy,

$$E^{(1)} = \langle \Phi_0 | \hat{H} - \hat{F} | \Phi_0 \rangle = E_{\mathrm{HF}} - E^{(0)}, \tag{1.34}$$

where the first term is identical to Eq. 1.18, while the second one is identical to the one in Eq. 1.33. Left projecting Eq. 1.31b by an excited determinant, Φ_q, which are also eigenvectors of $\hat{F}$,

$$\hat{F} | \Phi_p \rangle = E_p^{(0)} | \Phi_p \rangle , \tag{1.35}$$

and expanding $\Psi^{(1)}$ in terms of other Φ_p's,

$$| \Psi^{(1)} \rangle = \sum_p t_p | \Phi_p \rangle , \tag{1.36}$$

yields

$$\langle \Phi_q | \hat{F} \sum_p t_p | \Phi_p \rangle + \langle \Phi_q | \hat{H}' | \Phi_0 \rangle = E^0 \langle \Phi_q | \sum_p t_p | \Phi_p \rangle \tag{1.37}$$

where the summation index p goes over all excited determinants. After rearranging, the famous t amplitudes emerge,

$$t_q = \frac{\langle \Phi_q | \hat{H} | \Phi_0 \rangle}{E_0 - \langle \Phi_q | \hat{F} | \Phi_q \rangle}. \tag{1.38}$$

Lastly, the second-order Møller-Plesset perturbation theory (MP2) energy can be obtained through the left projection of Eq. 1.31c by the HF wavefunction Φ_0. It has the form,

$$E^{(2)} = \langle\Phi_0|\,\hat{H} - \hat{F}\,|\Psi^{(1)}\rangle \tag{1.39a}$$

$$= \langle\Phi_0|\,\hat{H}\sum_p t_p\,|\Phi_p\rangle \tag{1.39b}$$

$$= \sum_p \frac{\langle\Phi_0|\,\hat{H}\,|\Phi_p\rangle\,\langle\Phi_p|\,\hat{H}\,|\Phi_0\rangle}{E_0 - \langle\Phi_p|\,\hat{F}\,|\Phi_p\rangle} \tag{1.39c}$$

$$= \sum_{i>j}\sum_{a>b} \frac{|\,\langle ij|\,|ab\rangle\,|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \tag{1.39d}$$

where $\langle ij|\,|ab\rangle$ are the matrix elements of the electron-electron interaction given by

$$\langle ij|\,|ab\rangle = \int\int \chi_i(\mathbf{x}_1)\chi_j(\mathbf{x}_2)\frac{1}{r_{12}}(1-\hat{P}_{12})\chi_a(\mathbf{x}_1)\chi_b(\mathbf{x}_2)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2. \tag{1.40}$$

Since our molecular orbitals are constructed from M LCAOs, each term like Eq. 1.40 requires M^4 integrals to be evaluated, resulting in $N^2(M-N)^2M^4$ scaling, for a wavefunction of N electrons. This scaling is typically brought down by a common "trick" of reusing the M^4 integrals and doing the transformation one at a time. By doing so, we have four $\mathcal{O}(M^5)$ operations to make. This reordering of summations to reduce scaling is one example of dynamic programming.

There are a few things to note about the MP2 method. One, it is not variational and thus does not obey Eq.1.23. This drawback is not usually relevant for chemistry since the relative energy is improved. Furthermore, the MP series equation tends to oscillate above and below the exact energy [23]. This has led to the development of methods that take the mean between MP2 and MP3 [24]. Recent advances in MP2 include multiplying the same spin and opposite spin by scalar factors deduced empirically [25, 26] as well as regularizing the denominator to avoid the singularity in case of the degeneracy of the occupied and virtual molecular orbitals such as level shift [27, 28] and κ- and σ-MP2 [29].

Notice that Eq. 1.39d is obtained by applying the Slater-Condon rules, leaving only double excitations. The second assumption is that the different excited determinants, which are each made of orthogonal orbitals, are orthogonal. In some instances of MP2, breaking this convention is advantageous for faster calculations as is discussed in the next section.

Hylleraas Functional

The Hylleraas functional is a different way to solve the equations of the perturbative series [11, 30–32]. First, consider the general form of the perturbative series in Eq. 1.31d,

$$\hat{H}_0\Psi^{(\mathrm{n})} + \hat{H}'\Psi^{(\mathrm{n}-1)} - \sum_{k=0}^{n} E^{(\mathrm{k})}\Psi^{(\mathrm{n}-\mathrm{k})} = 0. \tag{1.41}$$

By some algebraic manipulation, Wigner's $2n+1$ rule can be found

$$E^{(2n+1)} = \langle \Psi^{(n)} | \hat{H}' | \Psi^{(n)} \rangle - \sum_{k,l=0}^{n} E^{(2n+1-k-l)} \langle \Psi^{(k)} | \Psi^{(l)} \rangle, \tag{1.42}$$

which asserts that the nth order wavefunction is the highest order wavefunction needed to calculate the $(2n+1)$th order of the energy. Hylleraas was the first to apply the variational principle back into Eq. 1.41 with Lagrange multipliers [30, 31], assuming the nth wavefunction is some trial wavefunction. Taking the derivative of the Lagrangian reproduces Eq. 1.42, making it a formal upper bound to the $2n+1$th order energy, $E^{(2n+1)}$.

Although the formalism developed by Hylleraas has been around since the 1930s [30, 31], it was not common in the quantum chemistry literature until the seminal work of Peter Pulay and Svein Saebø in the 1980s [33, 34]. Realizing that this formalism can be adapted to any set of orbitals, Pulay and Saebø were able to reproduce MP2 energy at a much lower cost. Here we show the second-order energy functional

$$h_2[\Psi^{(1)}] = \langle \Psi^{(1)} | \hat{H}_0 - E^{(0)} | \Psi^{(1)} \rangle + 2 \langle \Psi^{(1)} | \hat{H}' - E^{(1)} | \Psi^{(0)} \rangle \tag{1.43a}$$

$$= \sum_{ijab} \sum_{i'j'a'b'} t_{ijab} (\mathbf{H}_0)_{ijab,i'j'a'b'} t_{i'j'a'b'} + 2 \sum_{ijab} t_{ijab} \langle ij | ab \rangle \geq E^{(2)} \tag{1.43b}$$

Notice that taking the derivative with respect to $\mathbf{t}$ reproduces the MP2 energy expression in Eq. 1.39d, which is the minimum of the Hylleraas functional.

Coupled Cluster

As an alternative to the MP series, the coupled cluster (CC) method was developed in the 1970s [11, 35–37]. It uses an exponential ansatz

$$|\Psi\rangle = e^{\hat{T}} |\Phi_0\rangle \tag{1.44}$$

where $\hat{T}$ is the excitation operator, which can be written in second quantization

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots \tag{1.45a}$$

$$= \sum_{ia} t_{ia} a_a^\dagger a_i + \sum_{ijab} t_{ijab} a_b^\dagger a_a^\dagger a_j a_i + \sum_{ijkabc} t_{ijkabc} a_c^\dagger a_b^\dagger a_a^\dagger a_k a_j a_i + \cdots, \tag{1.45b}$$

where $a_a^\dagger$ and a_i are the raising and lower operators of the ath and ith orbitals, respectively, and the t's are expansion coefficients of the excited determinants. The excitation operators $\hat{T}_1$, $\hat{T}_2$, $\hat{T}_3$, etc. are called singles, doubles, triples, etc. excitation operators. Minimizing the energy of this exponential ansatz produces,

$$\min_{\mathbf{t}} \langle \Phi_0 | e^{\hat{T}^\dagger} \hat{H} e^{\hat{T}} | \Phi_0 \rangle = E_{\mathrm{VCC}} \tag{1.46}$$

the variational coupled cluster equations and energy. Despite being variational, it scales worse than full CI and is never used in practice.

Because the exact ground state is expected to have a high overlap with the HF wavefunction, Φ_0, the conventional coupled cluster energy is evaluated as

$$\langle \Phi_0 | \, \hat{H} e^{\hat{T}} \, | \Phi_0 \rangle = E_{\mathrm{CC}} \tag{1.47}$$

leaving all the t-amplitudes unknown. A subspace projection onto the space of excited determinants can produce

$$\langle \Phi_{ijk\ldots}^{abc\ldots} | \, \hat{H} e^{\hat{T}} \, | \Phi_0 \rangle = E_{\mathrm{CC}} \, \langle \Phi_{ijk\ldots}^{abc\ldots} | \, e^{\hat{T}} \, | \Phi_0 \rangle \tag{1.48}$$

which is the usual (or conventional) coupled cluster equations. Since the exponential operator contains infinite order excitations, the equations are only soluble upon truncating the number of $\hat{T}$ operators and limiting the set of excited determinants projected from the left. The equations resulting from that are non-linear and require iterations until convergence is achieved. Upon limiting the excitation operators to only $\hat{T}_1$ and $\hat{T}_2$, this method is called CCSD, while inclusion of $\hat{T}_3$ operator makes it CCSDT, and so on.

Unfortunately, this conventional CC approach is no longer variational, but it is very useful for its size consistency, size extensivity, and ability to capture more correlation than CISD at the CCSD level. CCSD is capable of reproducing full CI results for 2-electrons, which makes it exact for the simplest bond in chemistry, H_2. Furthermore, this approach is polynomially scaling for different truncations of the excitation operator, e.g., CCSD is $\mathcal{O}(M^6)$ while CCSDT is $\mathcal{O}(M^8)$, etc. Lastly, perturbative corrections have been successfully derived and implemented, most famously the coupled cluster singles and doubles with triple corrections, known as CCSD(T) [38]. This method is very accurate for a wide range of stable organic compounds and earned the name of the "gold standard" of quantum chemistry.

Other Wavefunction-Based Methods

There are a few approaches to reduce the work needed for *ab initio* methods like CI, MP, and CC. One approach discussed above is to limit the work to a specific order of excitation, such as CISD or CCSD, etc. The next simple thing to do is to restrict the set of correlated orbitals, for example by constraining the summation on the occupied or virtual orbitals. This approach can be justified by the denominator of the MP2 energy term, Eq. 1.39d, being only appreciable for occupied and virtual orbitals close in energy. Such methods are named active space methods whereas including all excitations (singles, doubles, triples, etc.) in an orbital active space is called complete active space, for example, complete active space CI (CASCI). These approaches clearly reduce the prefactor, but not the exponents of the scaling of these methods. Lastly, a class of active space (and other) methods emerges where the molecular orbitals are variationally optimized to reduce the overall energy. Application of these on CASCI gives rise to complete active space SCF (CASSCF) [39–41], while applying them to

MP2 and CCD gives rise to orbital optimized MP2 (OOMP2) [29, 42] and optimized doubles (OD) [43, 44], respectively.

One way to reduce the work of calculating the correlation energy is the independent electron pair approximation (IEPA) [45–48]. It treats every pair of electrons independently from every other pair (decoupling), and a CID matrix with only two electrons promoted is diagonalized. Because the equations are solved for each pair independently, this method is relatively cheap. It however suffers from serious drawbacks including poor accuracy among others discussed elsewhere [4, 49]. Most importantly, the method is not invariant to occupied-occupied orbital mixing, unlike all the methods discussed so far. A better and more accurate alternative is the coupled electron pair approximation (CEPA) [50], which reintroduces an approximate coupling between different pairs, while keeping the equations linearized, making the method still size extensive.

Valence bond theory offers an alternative picture to the molecular orbital theory discussed so far [51]. While molecular orbitals tend to spread over the whole molecule and take its symmetry, valence bond theory is pivoted around atomic orbitals combining to form bonds. Despite not becoming as popular as the molecular orbitals theory, the history of its development is very interesting [52, 53]. It was Heitler and London [54] who explained H_2 bond formation by combining two hydrogen atomic orbitals,

$$\Phi_0 = \frac{1}{2}(\psi_A\psi_B + \psi_B\psi_A)(\sigma(\alpha)\sigma(\beta) - \sigma(\beta)\sigma(\alpha)), \tag{1.49}$$

where the order of the orbitals (spatial and spin) indicates the electron in question. Heitler and London showed that the exchange energy is important for obtaining a qualitatively accurate bond length and energy of H_2. To get a quantitative answer, it is important to include the ionic (antibonding) wavefunction (doubly excited determinant),

$$\Phi_{\mathrm{I}} = \frac{1}{2}(\psi_A\psi_A + \psi_B\psi_B)(\sigma(\alpha)\sigma(\beta) - \sigma(\beta)\sigma(\alpha)). \tag{1.50}$$

Together, the overall wavefunction becomes,

$$\Psi = c_0\Phi_0 + c_{\mathrm{I}}\Phi_{\mathrm{I}}, \tag{1.51}$$

where c_0 and c_{I} are variational parameters. Coulson and Fischer [55] proposed using new atomic orbitals, ϕ_A, ϕ_B with a variational parameter, c, to improve the Heitler-London wavefunction, Φ_0,

$$\phi_A = \psi_A + c\psi_B \tag{1.52}$$

$$\phi_B = \psi_B + c\psi_A. \tag{1.53}$$

Rewriting Eq. 1.49 with the new modified orbitals recovers Eq. 1.51, i.e.,

$$\Psi = \frac{1}{2}(\phi_A\phi_B + \phi_B\phi_A)(\sigma(\alpha)\sigma(\beta) - \sigma(\beta)\sigma(\alpha)) \tag{1.54a}$$

$$= N\big((\psi_A + c\psi_B)(\psi_B + c\psi_A) + (\psi_B + c\psi_A)(\psi_A + c\psi_B)\big)(\sigma(\alpha)\sigma(\beta) - \sigma(\beta)\sigma(\alpha)) \tag{1.54b}$$

$$= N\big((1 + 2c^2)\Phi_0 + 2c\Phi_{\mathrm{I}}\big), \tag{1.54c}$$

where N is a normalization constant. Optimizing the atomic orbitals produces a wavefunction capable of describing bond dissociation. However, this uses only two basis function orbitals (minimal basis set), which is typically insufficient to describe highly polarized bonds. Hurley and co-workers [56] in the 50s developed the perfect pairing (PP) approach which works for multiple bonds each with its singlet excitation. Later, Goddard and co-workers [57] in the 70s generalized this approach for any basis set, i.e. Eq. 1.52 included all basis functions. This method was called generalized valence bond (GVB) theory, while Gerratt et al. developed an almost identical formalism named spin-coupled valence bond (SCVB) [58, 59] at about the same time. Note that the spin-coupling between pairs is added as an additional variational parameter, and removing the spin-coupling parameters from the GVB/SCVB gives back the PP method [60] which had a reduced compute cost compared to the "full GVB". Cullen reformulated the PP wavefunction using the (conventional) coupled-cluster ansatz [61]. Head-Gordon and co-workers explored reviving other GVB methods and variants with the coupled-cluster ansatz allowing more practical calculations [61–65].

VB methods without the coupled-cluster approach resemble CASSCF in their computational cost. Yet, they offer the advantage of being closely related to chemical concepts, allowing chemical interpretation of accurate wavefunctions [66]. Both VB methods and CASSCF recover the static correlation [67], which is the part of the correlation associated with bond breaking or the polyradical nature. Static correlation usually resides in the valence space and is characteristic of the system, while dynamic correlation requires many virtual orbitals and is more universal [11, 68]. Nonetheless, many definitions exist for the difference between the two types of correlation [69]. Although VB methods describe single bond breaking very well, they suffer from missing the dynamic correlation and dispersion, making these methods not widely used for chemical applications.

Accordingly, there have been many efforts to reproduce the energy of the methods discussed to any arbitrary accuracy [47, 68] by exploiting the locality of the electron-electron integrals, which are the basis of the difficult term in the Hamiltonian, r_{12}^{-1}. Note that these methods are only applicable to insulating systems where the occupied and virtual energies are well separated. Such treatments are capable of reducing the asymptotic scaling of any of the methods mentioned earlier to linear in principle while achieving that is rather daunting in practice.

1.3 Density Functional Theory

Density functional theory (DFT) starts with two Hohenberg-Kohn (HK) theorems [70]. The first asserts that there is a one-to-one mapping between any V_{ext} potential and the ground state density. This means knowing the potential gives the density and vice versa. The second applies the variational principle, proving that the ground state density is unique for a given potential.

The formulation for DFT is somewhat different from the wavefunction-based methods discussed earlier, and we will only give a brief summary here and refer interested readers to

Refs. [71–76]. It is implied from the first HK theorem that the ground state wavefunction is implicitly described by its density. As such, the exact ground state energy can in theory be obtained from the density, $\rho(\mathbf{r})$, by evaluating the energy functional of the density. In this formalism, the electronic Schrödinger equation from Eq. 1.6 is replaced by

$$E_0 = \min_{\rho} E[\rho] = \min_{\rho} \big(T[\rho] + V_{\text{ee}}[\rho] + V_{\text{ext}}[\rho]\big), \tag{1.55}$$

where V_{ext} is the external potential energy, normally the mere nuclear attraction energy. Without knowing the exact form of the electron-electron interaction energy functional, $V_{ee}[\rho]$, it is possible to partition it to:

$$V_{\text{ee}}[\rho] = V_{\text{J}}[\rho] + V_{\text{x}}[\rho] + V_{\text{c}}[\rho], \tag{1.56}$$

where V_{J} is the known classical Coulomb repulsion, while V_{x} and V_{c} are the unknown exchange and correlation energy functionals, respectively, which are the quantum terms. These two terms are often combined into one exchange-correlation functional, V_{xc}.

Early attempts estimated different parts of the functional for the uniform electron gas (UEG), most famously the Thomas-Fermi-Dirac model for the UEG [77–79]. In that limit, the kinetic energy and the exchange energy can analytically be shown to be:

$$T[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho(\mathbf{r})^{5/3} d\mathbf{r} \tag{1.57}$$

$$V_{\text{x}}[\rho] = -\frac{3}{4}\Big(\frac{3}{\pi}\Big)^{1/3} \int \rho(\mathbf{r})^{4/3} d\mathbf{r} \tag{1.58}$$

Exciting at first, these models failed to do any useful chemistry, primarily because of large errors in the kinetic energy density functional. Using the Hohenberg-Kohn energy functional is known as orbital-free DFT since treatment has been limited to the density, with no care to orbitals.

In a similar spirit, a breakthrough was made by Kohn and Sham (KS) reintroduced *fictitious* orbitals [80]. Kohn-Sham DFT (KSDFT) formalism is based on minimizing the energy of one Slater determinant, Φ^{KS}, made of orbitals, yet the energy evaluation uses a functional that is blind to these orbitals. The orbitals are found by modifying the Fock operator to be

$$\Big(-\frac{1}{2}\nabla^2 + \hat{V}_{\text{ext}} + \hat{J}[\rho] + \hat{V}_{xc}[\rho]\Big)\,|\chi_i\rangle = \epsilon_i\,|\chi_i\rangle\,, \tag{1.59}$$

where $\hat{V}_{\text{ext}}(\mathbf{r})$ is the external potential, $\hat{J}(\mathbf{r})$ is the Coulombic potential to the density, and $\hat{V}_{xc}$ is the exchange-correlation potential. The kinetic energy is evaluated using the KS orbitals,

$$T_s = \langle\Phi^{\text{KS}}| - \frac{1}{2}\nabla^2\,|\Phi^{\text{KS}}\rangle = \sum_i \langle\chi_i| - \frac{1}{2}\nabla^2\,|\chi_i\rangle\,, \tag{1.60}$$

which is the exact kinetic energy of electrons of the non-interacting system. The exchange-correlation functional takes the form

$$\hat{V}_{\text{xc}}[\rho] = \frac{\delta}{\delta\rho}\big(T[\rho] - T_{\text{s}}[\rho] + V_{\text{x}}[\rho] + V_{\text{c}}[\rho]\big) \tag{1.61}$$

which can in principle recover the exact energy. Note the subtraction between the two terms on the right-hand side to substitute the approximate T_s in Eq. 1.60. The Kohn-Sham formalism is transformative because it gives exact V_ext, J, and T_s, leaving only the V_xc to be unknown.

Exchange-correlation functional

Many attempts have been made to approximate the exchange-correlation functional. The different functionals are usually thought of as rungs on Jacob's ladder as suggested by John Perdew [81]. The rungs ascend from the "hell of Hartree" with no exchange or correlation onto the "heaven of chemical accuracy" of 1 kcal/mol. The rungs are briefly discussed below.

The first rung up from hell is to include some local density information, which initially used the Slater exchange, V_x in Eq. 1.58. While the simplest form of the local density approximation (LDA) is the Thomas-Fermi-Dirac model, now LDA refers to functionals that use a local function of the density, $\rho(\mathbf{r})$, to attain exactness for the UEG. Indeed, the next two rungs on Jacob's ladder are the generalized gradient approximation (GGA) [82–84] and meta-GGA (mGGA) [85–87], which use the first and second gradient of the density. Such information can in principle permit exactness for other systems, such as slowly varying electron gases. These three take the general forms

$$V_\mathrm{LDA}[\rho] = \int \rho(\mathbf{r}) f(\rho(\mathbf{r})) \mathrm{d}\mathbf{r} \tag{1.62}$$

$$V_\mathrm{GGA}[\rho] = \int \rho(\mathbf{r}) f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})) \mathrm{d}\mathbf{r} \tag{1.63}$$

$$V_\mathrm{mGGA}[\rho] = \int \rho(\mathbf{r}) f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \tau) \mathrm{d}\mathbf{r}, \tag{1.64}$$

where $f()$ is a function of the density or its gradients with some parameters, and τ is the kinetic energy density. This covers the first three rungs of the ladder. These three rungs scale $\mathcal{O}(N^3)$ with system size since they are dressed Hartree models with a larger prefactor arising from the numerical integration of the exchange-correlation functionals.

The fourth rung on the ladder contains hybrid functionals which add exact exchange to the exchange-correlation functional in Eq. 1.59. The purpose of the exact exchange is to counter-act the self-interaction error in DFT, where the electron interacts with its own density through $\hat{J}$ that is not guaranteed to cancel with $\hat{V}_\mathrm{xc}$. The self-interaction error is also referred to as the delocalization error since infinitely-separated molecules can spuriously have a fractional number of electrons. To remedy this error, some of the exact exchange is added as an extra term in Eq. 1.59 as is done in global hybrids [88]. While global hybrids suffer from less self-interaction error, the problem persists. Using the HF wavefunction and evaluating the DFT energy, as is done in density-corrected DFT [89, 90], can improve cases suffering from this error [91, 92]. Alternatively, the long-range exact exchange can be added to recover the HF limit for distant electrons [93, 94]. Known as range-separated hybrids, they

interpolate between the short-range exchange functional and the long-range exact exchange of HF using a smooth continuous function. This approach, despite being somewhat empirical, has been successful in making the most accurate density functionals based on large datasets of reactions [95]. Impressively, methods in this rung still scale similarly to HF at $\mathcal{O}(N^4)$ scaling because of the exact exchange, yet have mean errors lower than correlated methods. Of this rung, ωB97X-D [96] and ωB97X-V [97] have demonstrated very high accuracy [95] and are used in Chapter 3 and Chapter 4 of this thesis.

The fifth rung on our ascension to heaven is the inclusion of parameters that depend not only on the occupied orbitals but also on the virtual orbitals. These typically include some parametrized component of the MP2 contributions as is done in double-hybrid functionals [98]. These offer some of the highest accuracies for modern DFT although they have higher scaling at $\mathcal{O}(M^5)$ from the MP2 part. Methods in this rung are difficult to optimize due to the double-counting of the correlation energy. This has not stopped even more challenging recent proposals to minimize a multi-determinant wavefunction using the Kohn-Sham Hamiltonian [99, 100]. A depiction of Jacob's ladder with the five rungs discussed is shown in Fig. 1.2.

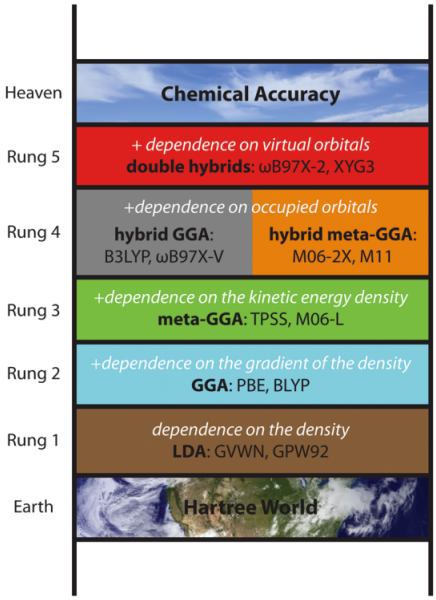


Figure 1.2: Jacob's ladder as proposed by Perdew. Figure adapted from Ref. [97].

The exchange-correlation functionals so far relate to the short-range of the correlation through approximate damping of the electron-electron repulsion. The long-range correlation such as dispersion requires a different treatment. There have been two very important advances in the non-local correlation. First, Grimme's empirical "-D" functionals, whose form is borrowed from the force field community, add an r_{AB}^{-6} attractive term between atoms (damped at a short distance), parametrized for each atom [101–103]. The second approach

was to design a van der Waals density functional (vdW-DF) which relies completely on the density [104, 105]. These take the form

$$V_{\text{nlc}}[\rho] = \int\int \rho(\mathbf{r})\Phi(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{1.65}$$

where $\Phi(\mathbf{r},\mathbf{r}')$ is the non-local correlation kernel. An important functional of that class was developed by Vydrov and Van Voorhis [106–108] which contains only 3 empirical parameters. This is an impressive accomplishment considering some exchange-correlation functionals contain more than 50 parameters [109, 110]. The parametrization fever culminated with the development of DM21 [111], training an exchange-correlation functional with neural networks containing thousands of parameters. More parameters are not necessarily better when it comes to density functional approximations [112] although today's most accurate functionals are semi-empirical and parameterized. Much work in recent years has explored the transferability to systems not seen in the training process [95] as well as the ability to produce the correct energy derivatives [113].

On the philosophical difference between ab initio and empirical methods in quantum chemistry: Initial developments within wavefunction-based methods and DFT were both motivated by ideal cases where the theory works exactly. The Slater exchange in Eq. 1.58 is exact for the non-interacting uniform electron gas, for example. The HF energy in Eq. 1.15 is the exact energy for one Slater determinant, the simplest form of an antisymmetrized wavefunction, and is exact for non-interacting electrons (as exploited in KSDFT). It is an unfortunate fact of the universe that molecules are neither uniform electron gases nor made of one Slater determinant; we have an exact answer for a different question. That exactness at one limit while deviating at others gives chemists an assessment tool for the correctness of these answers by examining the physics. For example, the natural orbital occupation numbers of the true wavefunction are a non-biased measure of the correctness of the Hartree-Fock energy. Similar arguments can be made with GVB or CASSCF when, for example, the number of electrons unpairing is known upon varying some coordinate of the chemical system or for MP2 in the weak-correlation regime presumed in the perturbative treatment. This very important indicator may potentially be lost to the supremacy of epistemology and empiricism. Parameterizing a method, whether done for the MP2 correlation or the exchange-correlation functionals, has only statistical significance, and nothing can be said about the correctness of these methods for a particular system. In other words, the dichotomy between the physics and the DFT parameters means little can be said about the correctness of a given approximation from the physics of a given system, aside from the one Slater determinant treatment. However, more can be said by the extensive benchmarking of density functionals giving us strong statistical confidence in their accuracy within certain areas of chemistry. This fundamental difference makes parameterized methods more useful as black-box methods for a wide range of chemical systems, while some wavefunction methods require a certain level of quantum chemistry knowledge and understanding to apply and extract information for a specific chosen system. It is unsettling to think that even the most

statistically accurate density functionals developed are not exact for any system, not even a hypothetical one.

1.4 Wavefunction Analysis

The previous sections were concerned with evaluating the energy of a many-electron system. Once that is known, the derivatives of the energy can tell us many properties (observables) of the system, that are difficult or impossible to evaluate accurately without quantum mechanics. Another evolving feature of quantum chemistry is to run numerical experiments to understand *why* a system exhibits the chemistry it does. While the first of these reasons relies on experimental observables, the chemical understanding is rather murky as it lacks numerical falsifiability. Although parsimony is one of the tenets of science, chemical frameworks such as Lewis structures and chemical reaction arrows have been powerful predictive tools for chemists long before the advent of modern quantum chemistry. This entices quantum chemists to quantify these chemical frameworks and make them more robust, and, perhaps, more extensible.

> Give us insight, not numbers.
>
> Charles Coulson [114, 115]

Population Analysis

The simplest description of the wavefunction for chemists is the effective charge on each atom. An early attempt made by Mulliken [116] who proposed

$$Q^A = Z^A - \sum_{\mu \in A} \big(\mathbf{PS}\big)_{\mu\mu}, \tag{1.66}$$

where Q^A is the effective charge on atom A with nuclear charge Z^A, $\mathbf{P}$ is the electronic density matrix in a given atom-centered basis and $\mathbf{S}$ is the overlap of those basis functions. The trace of $\mathbf{PS}$ gives the number of electrons, and the Mulliken population partitions the density between different atoms depending on *which* atoms the basis function is centered on. Since the trace is invariant to cyclic permutation, Eq. 1.66 can be also written as

$$Q^A(\alpha) = Z^A - \sum_{\mu \in A} \big(\mathbf{S}^\alpha \mathbf{P} \mathbf{S}^{1-\alpha}\big)_{\mu\mu} \tag{1.67}$$

where α can be any real number. Löwdin chose $\alpha = 0.5$, similar to symmetric orthogonalization [117]. Both of Mulliken and Löwdin population analyses suffer from a strong dependence on the basis set used, and in general, become worse as $\mathbf{S}$ deviates from the

identity. This naturally occurs as the rank of the GTO basis set is expanded towards completeness. Another failure mode can be constructed by providing disproportionately more basis functions to some atoms, consequently assigning more charge to these atoms, with a single-center expansion being the extreme limit.

The failure of Mulliken and Löwdin population analyses has led to the emergence of two schools of thought. First, improve the way in which orbitals are assigned to atoms, which in essence changes the representation of $\mathbf{P}$ and in particular the summation condition (basis tagging) in Eq. 1.66. Two, assign electrons to atoms through spatial numerical integration of the density, $\rho(\mathbf{r})$. The latter offers the nice property of removing the dependence on the basis set, at the expense of requiring a partition of space into atomic volumes, which itself contains an element of arbitrariness.

Of the first school of thought, Weinhold's natural population analysis is a popular one [118]. Each atom gets its own natural atomic orbitals formed by diagonalizing the block of the density matrix $\mathbf{P}$ of the basis functions of that atom. Then, occupancy-weighted symmetric orthogonalization is used instead of Löwdin's population analysis in Eq. 1.67. Mayer's effective atomic charges [119] differ by treating the atomic subblocks of $\mathbf{P}$ and $\mathbf{S}$ differently. Another subclass of methods uses a minimal basis set to assign the density, where the overlap becomes small and both Mulliken and Löwdin populations provide very similar results [120–122]. A recent method known as intrinsic atomic orbitals has gained a lot of popularity [122]. This method employs an exact projector of the density into a minimal basis set. This method is discussed in further detail in Chapter 3 of this thesis. Alternatively, the molecular atomic orbitals method makes a minimal basis with no reference minimal basis set. Its development and results are discussed in great detail in Chapter 2.

Of the density-based partitioning methods, many interesting approaches have been developed. Most prominent is Bader's quantum theory of atoms in molecules (QTAIM) approach [123, 124]. Between every pair of bonded atoms, there exists a point at which the gradient of the density is zero. This point is called the bond critical point, and it divides the electronic density between the two atoms in question. This method is very interesting and had many further developments [125] and critiques in the literature [126, 127]. Another approach is to find the charges that provide the least squares fit of the electrostatic potential grid [128]. The electrostatic potential is evaluated on the van der Waals radius of atoms, which makes atoms not represented in the molecular van der Waals radius (buried inside the chemical system) have undetermined charges. Last but not least, Hirshfeld's charges [129, 130] are found by fitting the molecular density to the density of the isolated atoms. The least squares fit interpolates between the different states of the atom, neutral and charged, with the constraint of the total charge of the system.

Although the methods discussed above were for reproducing effective charges (the atomic monopoles), they often require minimal changes to also reproduce higher moments predictions [131]. Atomic dipoles, quadrupoles in particular are of interest for molecular mechanics applications, as they better approximate the quantum mechanical density compared to sole atomic charges.

Oxidation State

Oxidation state is the effective *integer* number of electrons on an atom or ligand in a complex. In other words, every electron pair is assigned to a center or group following established rules such as those recommended by IUPAC [132, 133]. It is very useful in understanding the reactivity of metals in inorganic chemistry and catalysis. Although the concept is very prevalent in introductory chemistry courses, it is not observable and cannot be measured in the lab. However, many chemical properties relating to reactivity and spectroscopy depend directly on the commonly accepted oxidation state, making it a very relevant and respected concept in chemistry [134]. Many different approaches in quantum chemistry attempt then to give predictions that match our chemical intuition and what is taught in chemistry courses. These approaches are discussed (and extended) in Chapter 3 of this thesis. Note that rounding up (or down) the atomic charge alone is not the same as the atomic oxidation state. This is simply due to the fact that rounding a sum of numbers is not necessarily the sum of the rounded numbers.

Molecular Orbitals

If the eigenfunctions of the hydrogen atom Hamiltonian are the famous atomic orbitals, then what are the eigenfunctions of the molecular Hamiltonian? The answer is rather more difficult since the many-body wavefunction does not possess simple functions. In the case of a single Slater determinant, the eigenfunctions of the Fock operator are the canonical molecular orbitals. The energies of these orbitals describe electron ionization or affinity from these orbitals in the zeroth-order picture by Koopman's theorem [4, 135]. Similarly, the eigenfunctions can be thought of as the empty space ("hole") left by an ionized electron or the new state of ("particle") an attached electron. Both of these are approximate connections since they are missing orbital relaxation and electron correlation. The Dyson orbitals are more rigorously based in exact quantum mechanics because they incorporate both of the missing effects back and relate to observables of excited states and ionized systems [136].

A generalization of these orbitals is common in the quantum chemistry literature. By diagonalizing the 1-particle density matrix (1PDM) of any wavefunction,

$$\mathbf{P} = \mathbf{C}\Lambda\mathbf{C}^{\dagger} \tag{1.68}$$

the natural orbitals can be found as columns of the matrix $\mathbf{C}$. Note that in Eq. 1.68 the matrix $\mathbf{P}$ is assumed to be in an orthogonal representation. Otherwise, a generalized eigenvalue problem is used. The diagonal matrix Λ of occupation numbers is strictly 1's for occupied orbitals and 0's for the virtual orbitals of a single Slater determinant, such as the HF or KSDFT wavefunction. For correlated methods, this is not true, and all of the natural orbitals exhibit some occupancy. The eigenvalues of Λ can therefore indicate the multi-determinant nature of the chemical system. It is worth mentioning that methods like the correlated orbital theory fold the excitations back to the occupied orbitals to contain some amount of correlation [137].

Canonical orbitals tend to spread over the whole system and transform as the irreducible representations consistent with the symmetry of the system. Since the virtual orbitals are related to the attachment of electrons to the system, they may couple to the continuum [138, 139]. For example, in a GTO basis, the lowest unoccupied molecular orbital (LUMO) in the complete basis set limit corresponds to a free particle, of which there is an infinite number of. The canonical virtual orbitals of KSDFT contain some correlation and tend to be better behaved although their energies have a different picture [71, 140]. The description of KS canonical orbitals and energies is often discussed in the context of Janak's theorem [141].

Since the HF and KSDFT energies are invariant to occupied-occupied unitary rotations, many alternative pictures can emerge by specifying a quantity to optimize and fixing these degrees of freedom. Boys localization minimizes the variance of orbitals [142], Pipek-Mezey localization minimizes the population spread of orbitals [143], and Edmiston-Ruedenberg maximizes the orbital's self-repulsion [144], to name a few. Interestingly, all of these localization procedures give core, lone pair, and bonding orbitals, restoring the valence shell electron pair repulsion (VSEPR) picture that is taught in introductory chemistry courses to predict molecular geometries and interactions. The use of these bonding orbitals is very relevant to the VB theory discussed earlier. Since orbitals are not observables, quantum chemists often say: "orbitals are not real." Some have even called bonding orbitals unicorns, often talked about but never observed [145]:

> "The problem was that the physical picture of a chemical bond in terms of the complicated wave function did not appeal at all to the imagination of chemists who were used to living with unicorns. The heuristic chemical bonding models were easy to write and they had proven to be very helpful as classification system and as predictive tools for experimental chemistry. Chemists had become so intimate with their unicorns that they attained the status of real living beings."

Localizing virtual orbitals is of great importance in correlation methods. For GVB theory, antibonding orbitals are needed to start and converge the calculation [146, 147] as will be discussed and demonstrated in Chapter 2. In other methods where all excitations are involved, localization of the virtual orbitals can reduce the scaling of the methods by calculating asymptotically fewer two-electron integrals. The earliest of these attempts was the work by Pulay, where the complementary projector of the occupied space is used [33, 34]. Subotnik et al. have produced orbitals that best resemble the atomic orbitals after projecting out the valence virtual orbitals [148]. Most recent attempts are occupied-specific virtual orbitals [149] and the domain-based local pair natural orbitals [150, 151]. Many more localization methods for occupied and virtual orbitals are in the literature and are discussed and contrasted elsewhere [152–162].

1.5 Non-covalent Interactions

The difference in energy between a non-covalently (or covalently) bonded complex and the infinitely separated constituent fragments is the binding energy. It determines the amount of energy required to separate the system as well as the relative population of the bonded species according to a Boltzmann distribution. While most covalent bonds are stable at ambient conditions, non-covalent interactions are typically much weaker, and with binding energies of only a few k_BT at room temperature, they are relevant to condensed phase simulations such as self-assembly, crystal engineering, and biological systems [163–166].

Non-covalent interactions are often explained by the permanent and induced electrostatics between the two fragments which can be approximated classically. Other important interactions such as charge donation and van der Waals have no classical analog [167]. While the tendency of the different interactions to increase or decrease can be predicted from observables of the system, exact numbers are more difficult to discern. For example, a larger dipole moment or polarizability typically results in a larger electrostatic interaction, and ionization potential and electron affinity give indicators of the direction and magnitude of charge donation. However, quantifying the different components of interaction can be difficult and even contentious at times. Although the 'correct' decomposition of the energy may be difficult (or even impossible) to unfold, a method with clearly erroneous predictions should not be considered. By this process of elimination, one can get to methods that give reasonable predictions for many systems. Below, we discuss two common methods for decomposing non-covalent interactions.

SAPT

In the limit where the molecules interact weakly, perturbation theory is a reasonable approximation to the interaction terms [168–170]. A common choice is the symmetry-adapted perturbation theory (SAPT) where the Hamiltonian is split between inter- and intra-molecular interactions. The double perturbation theory Hamiltonian between fragment A and B is then

$$\hat{H} = \hat{F}_A + \hat{F}_B + \lambda\big(\hat{H}_A - \hat{F}_A + \hat{H}_B - \hat{F}_B\big) + \xi\hat{V}_{AB} \tag{1.69}$$

This formalism is similar to that of MP2 in Eq. 1.29, except for separating the Hamiltonian into on-fragment terms and the intermolecular interaction term given by

$$\hat{V}_{AB} = -\sum_{i\in A, b\in B}\frac{Z_b}{r_{ib}} - \sum_{a\in A, j\in B}\frac{Z_a}{r_{aj}} + \sum_{i\in A, j\in B}\frac{1}{r_{ij}}, \tag{1.70}$$

where here the indices i, j run over electrons and indices a, b run over nuclei. Since there are two powers of expansion, energy terms are donated by $E^{(\mathrm{mn})}$, where m is for expanding in powers of ξ (intermolecular interaction) while n is for expanding in powers of λ (correlation). Furthermore, the wavefunction of the two fragments should be antisymmetrized,

$$\Psi = \mathcal{A}\Psi_A\Psi_B \tag{1.71}$$

where $\mathcal{A}$ is the antisymmetrizer of Eq. 1.11, and Ψ_X is the HF wavefunction of fragment X, antisymmetrized on X. Energy terms relevant to the antisymmetrized wavefunction are donated by 'exch' as will be discussed below.

Depending on the level of perturbation theory used, many variants exist such as SAPT0, SAPT2, SAPT2+, etc., and the interested reader should refer to Refs. [170–173]. Here we will limit the discussion to the simplest form, the SAPT0 energy,

$$E_{\text{SAPT0}} = E_{\text{elst}}^{(10)} + E_{\text{elst-exch}}^{(10)} + E_{\text{ind}}^{(20)} + E_{\text{ind-exch}}^{(20)} + E_{\text{disp}}^{(02)} + E_{\text{disp-exch}}^{(02)} + \delta E_{\text{HF}} \tag{1.72}$$

Energy terms recovered from different powers of λ are correlation terms, such as dispersion, $E_{\text{disp}}^{(02)}$. Meanwhile, expanding in powers of ξ, the first-order energy recovers electrostatics, $E_{\text{elst}}^{(10)}$, and the second-order recovers induction, $E_{\text{ind}}^{(20)}$. Each 'exch' term is the difference between the energy from the Hartree product wavefunction and the antisymmetrized wavefunction as in Eq. 1.71. Because the perturbative energy is not guaranteed to be equal to the HF interaction energy, the difference in energy without correlation terms is required. This term is given by,

$$\delta E_{\text{HF}} = E_{\text{HF}} - \left(E_{\text{elst}}^{(10)} + E_{\text{elst-exch}}^{(10)} + E_{\text{ind}}^{(20)} + E_{\text{ind-exch}}^{(20)}\right) \tag{1.73}$$

where E_{HF} is the HF energy of the two fragments. This treatment is for the HF wavefunction but has been extended to other correlated methods and KSDFT [171, 174]. Note that there have been two main uses for SAPT. The relevant one for our discussion is decomposing the interaction energy between two fragments. Alternatively, many have used SAPT to efficiently calculate the interaction energies of large clusters of weakly-interacting molecules, where calculating the interaction energy is too computationally expensive by accurate CC methods [175].

Intermolecular Interaction

In the long-range, SAPT energy terms are in agreement with how intermolecular interactions should be described [167]. The electrostatics term is the Coulombic interaction between two densities. By multipole expansion of the densities, multipole-multipole interactions can be obtained. These electrostatic terms take the form

$$E_{\text{elst}} = \sum_{l,l'} \frac{\mathbf{Q}_l^A \cdot \mathbf{C}_{ll'} \cdot \mathbf{Q}_{l'}^B}{r^{1+l+l'}} \tag{1.74}$$

where $\mathbf{Q}^X$ are multipole expansion of the charge of fragment X, l, l' are the angular momenta of the multipolar expansion of the two fragments, and $\mathbf{C}_{ll'}$ is some tensor connecting the two multipolar expansions involving the orientation information between the different fragments. The main takeaway from this is the denominator, $r^{1+l+l'}$, that can tell us about the range of the interaction. For example, a charged species interacting with a neutral fragment with a dipole will have an interaction with the leading term of r^{-2} of the inter-fragment

distance. The interaction in Eq. 1.74 is exact if the summation index is taken to infinity but often truncated to dipoles, quadrupoles, and rarely octupoles. Moreover, the dipoles and quadrupoles are observables, and hence the interaction energy can be calculated from observables. Alternatively, the quantum mechanical densities can be used to obtain the permanent electrostatics term.

Because a charge can increase the magnitude of a dipole, there is more Coulombic energy to be included. In particular, the multipoles inducing each other is a higher-order term and describes the polarization energy. The dipole polarizability describes how easily a dipole can be induced and is very relevant to the polarization energy. Again, expanding the density in multipoles, one can estimate the polarization energy to be,

$$E_{\text{ind}} \propto \sum_{l,l'} \frac{\mathbf{Q}_l^A \cdot \mathbf{T}_{ll'} \cdot \alpha_{l'}^B}{r^{2+2l+2l'}} + \frac{\alpha_l^A \cdot \mathbf{T}_{ll'} \cdot \mathbf{Q}_{l'}^B}{r^{2+2l+2l'}} \tag{1.75}$$

where the α's correspond to vector field polarizabilities and $\mathbf{T}_{ll'}$ are non-trivial coupling tensors relating to the orientations of the two fragments. Notice that this treatment arises from the second-order perturbation theory, very similar to $E_{\text{ind}}^{(20)}$ in SAPT. The induction of the multipoles gives rise to higher-order inductions, and thus a self-consistent field formalism can be used [176].

Dispersion is a quantum mechanical effect often described as instantaneous dipoles between molecules. The charge density fluctuation is attractive when aligned, so configurations with simultaneous polarization contribute to the ground state energy. Although the London dispersion force resembles polarization in its dependence on polarizability and the intermolecular distance dependence, it is purely quantum mechanical. Indeed, the leading dispersion term is r^{-6} analogous to the instantaneous dipoles. The generalization of London dispersion is the Casimir-Polder equation [177, 178], which can describe the experimentally observed attraction between two neutral metal slabs [179], known as the Casimir effect. The simplest description of dispersion in SAPT is the genuine dispersion, $E_{\text{disp}}^{(02)}$, and exchange dispersion $E_{\text{disp}-\text{exch}}^{(02)}$, similar to MP2.

The last attractive term is the charge transfer (CT) where some charge is delocalized from one fragment to another. Typically the highest occupied molecular orbital (HOMO) or other high-energy orbitals can be donated to an empty orbital of the other fragment, such as antibonding orbitals. Partial charge donation to a σ^* orbital destabilizes the bond and raises the σ orbital energy, resulting in bond elongation observable in X-ray diffraction experiments [180, 181]. Although attempts have been made to recover CT from SAPT [182], the CT term does not naturally emerge from the SAPT formalism. Instead, it gets distributed between induction and the δE_{HF} term. This shortcoming of the method has resulted in many studies denying the existence of CT, attributing the interaction to σ-hole, the electropositive area associated with the empty σ^* (virtual) orbital. The description of antibonding orbitals is further discussed in Chapter 2.

While all the terms mentioned above are attractive, molecules repel each other at the short range due to Pauli repulsion. Because electrons cannot occupy the same space, they

are pushed out from their ground state configuration as the molecules attract. This effect is purely quantum mechanical and is captured in SAPT by the exchange term, $E^{(10)}_{\text{elst}-\text{exch}}$, which is non-negative. It scales exponentially with the intermolecular distance. In general, a higher Pauli repulsion and being in the overlapping region means different energy decomposition analysis methods will typically yield different results. In this region, all the exch terms become appreciable and the perturbative treatment becomes less accurate.

ALMO-EDA

Variational alternatives of SAPT have been developed. Most relevant here is the absolutely localized molecular orbital energy decomposition analysis (ALMO-EDA) [183–185]. In the ALMO-EDA framework, the interaction energy is split into 3 components upon successive releasing of constraints. First, the electronic energy is evaluated after the fragments are put into one antisymmetrized wavefunction. This defines the frozen energy which consists of dispersion, Pauli repulsion, and permanent electrostatics. Allowing each fragment to relax within its own virtuals gives rise to polarization energy. Finally, removing this constraint gives back the energy lowering due to the charge transfer. This method is discussed in further detail in Chapter 4.

1.6 Outline

To find fantastic molecular orbitals, this thesis starts with how to generate and identify antibonding orbitals in Chapter 2. Further, by combining the antibonding orbitals with the rest of the occupied orbitals, modified atomic orbitals can be obtained and are discussed also in Chapter 2. A new way to localize the occupied orbitals is presented in Chapter 3, which is used for oxidation state assignment relevant to chemical understanding and insight. Lastly, the force decomposition analysis is developed by using the analytical nuclear displacement derivatives of the energy decomposition analysis (EDA) and is presented in Chapter 4. This method is used to get chemical insight as well as to validate a force field method. A brief summary is given below for each chapter of this thesis.

Chapter 2: Valence Antibonding Orbitals And Molecular Atomic Orbitals

While bonding molecular orbitals exhibit constructive interference relative to atomic orbitals, antibonding orbitals show destructive interference. When full localization of occupied orbitals into bonds is possible, bonding and antibonding orbitals exist in 1:1 correspondence with each other. Antibonding orbitals play an important role in chemistry because they are frontier orbitals that determine orbital interactions, as well as much of the response of the bonding orbital to perturbations. In this Chapter, we present an efficient method to construct antibonding orbitals by finding the orbital that yields the maximum opposite spin

pair correlation amplitude in second-order perturbation theory (AB2) and compare it with other techniques with increasing the basis set size. We conclude the AB2 antibonding orbitals are a more robust alternative to the Sano orbitals as initial guesses for valence bond calculations, due to having a useful basis set limit. The AB2 orbitals are also useful for efficiently constructing an active space, and work as good initial guesses for valence-excited states. In addition, when combined with the localized occupied orbitals, and relocalized, the result is a set of molecule-adapted minimal basis functions that is built without any reference to atomic orbitals of the free atom. As examples, they are applied to population analysis of halogenated methane derivatives, H-Be-Cl, and SF_6 where they show some advantages relative to good alternative methods.

This work has been published [186] in *The Journal of Chemical Physics*.

Chapter 3: Oxidation State Localized Orbitals

Oxidation states represent the ionic distribution of charge in a molecule and are significant in tracking redox reactions and understanding chemical bonding. While effective algorithms already exist based on formal Lewis structures, as well as using localized orbitals, they exhibit differences in challenging cases where effects such as redox non-innocence are at play. Given a density functional theory (DFT) calculation with chosen total charge and spin multiplicity, this Chapter reports a new approach to obtaining fragment-localized orbitals that are termed oxidation state localized orbitals (OSLO), together with an algorithm for assigning the oxidation state using the OSLOs and an associated fragment orbital localization index (FOLI). Evaluating the FOLI requires fragment populations, and for this purpose, a new version of the intrinsic atomic orbital (IAO) scheme is introduced in which the IAOs are evaluated using a reference minimal basis formed from on-the-fly superposition of atomic density (IAO-AutoSAD) calculations in the target basis set, and at the target level of theory. The OSLO algorithm is applied to a range of challenging cases including high valent metal oxide complexes, redox non-innocent NO and dithiolate transition metal complexes, a range of carbene-containing TM complexes, and other examples including the potentially inverted ligand field in $[Cu(CF_3)_4]^-$. Across this range of cases, OSLO produces generally satisfactory results. Furthermore, in borderline cases, the OSLOs and associated FOLI values provide direct evidence of the emergence of covalent interactions between fragments that nicely complement existing approaches.

This work has been published [187] in *Journal of Chemical Theory and Computation*.

Chapter 4: Force Decomposition Analysis

Computational quantum chemistry can be more than just numerical experiments when methods are specifically adapted to investigate chemical concepts. One important example is the development of energy decomposition analysis (EDA) to reveal the physical driving forces behind intermolecular interactions. In EDA, typically the interaction energy from a good-quality density functional theory (DFT) calculation is decomposed into multiple additive

components that unveil permanent and induced electrostatics, Pauli repulsion, dispersion, and charge-transfer contributions to non-covalent interactions. Herein, we formulate, implement and investigate decomposing the forces associated with intermolecular interactions into the same components. The resulting force decomposition analysis (FDA) is potentially useful as a complement to the EDA to understand chemistry, while also providing far more information than an EDA for data analysis purposes such as training physics-based force fields. We apply the FDA based on absolutely localized molecular orbitals (ALMOs) to analyze interactions of water with sodium and chloride ions as well as in the water dimer. We also analyze the forces responsible for geometric changes in carbon dioxide upon adsorption onto (and activation by) gold and silver anions. We also investigate how the force components of an EDA-based force field for water clusters, namely MB-UCB, compare to those from force decomposition analysis.

This work has been submitted to *The Journal of Physical Chemistry*.

Chapter 2

Valence Antibonding Orbitals And Molecular Atomic Orbitals

2.1 Introduction

Virtual orbitals are important in chemistry as they play a central role in molecular orbital theory. From a computational standpoint, orbital mixing between occupieds and virtuals determines the optimal occupied orbitals in mean-field Hartree-Fock theory [4–6] and Kohn-Sham density functional theory [71, 72, 75, 95]. In wavefunction theory, electron correlation is typically described by amplitudes such as the pair correlations describing the simultaneous promotion of two electrons from occupied to virtual orbitals. The virtual orbitals span the unoccupied space, and the choice of representation is important. Canonical virtual orbitals are delocalized levels that are appropriate for electron attachment. Localized virtuals, such as the redundant non-orthogonal basis of atomic orbitals projected into the virtual space [33, 34], permit development of efficient local correlation methods, because the amplitude tensors describing correlation become sparse [188]. Other prescriptions for localized orthogonal virtuals exist [148, 158, 189], as well as proposals to form sets of virtuals that are specifically optimized for correlations that involve a given occupied level, as will be discussed below.

The virtual orbitals span the entire unoccupied space, which can be contrasted with the intuitive notion of antibonding orbitals that exist in 1:1 correspondence with bonding orbitals. The 1:1 correspondence is evident from constructive and destructive interference of a pair of 1s-type functions on two hydrogen atoms in H_2:

$$\sigma = N\left(1s_A + 1s_B\right) \tag{2.1}$$

$$\sigma^\star = N^\star\left(1s_A - 1s_B\right) \tag{2.2}$$

Antibonding orbitals themselves play a central role in describing chemical reactivity [190–195] of one molecule with another through donor-acceptor interactions between a high-lying occupied of one species with a low-lying antibonding orbital of the other. Frontier orbital theory is constructed on these ideas. Antibonding orbitals also play an important role in

describing strong electron correlations. A simple example is the stretching of the H-H bond which leads, in a minimal basis, to a strong increases in the amplitude for $\sigma\bar{\sigma} \rightarrow \sigma^\star\bar{\sigma}^\star$ excitation which breaks the bond.

While the antibonding orbitals are intuitive [192–194, 196], it is nonetheless not routine to extract them from modern quantum chemistry calculations performed in extended basis sets, which return canonical orbitals. By contrast, in a minimal basis description of hydrocarbons, the space of antibonding orbitals is naturally spanned by the canonical virtual orbitals. In larger basis sets however, different methods have been developed to extract the antibonding orbitals, often by relying on projection back onto some chosen minimal basis [147, 189, 197–200], typically a tabulated one for a specific free-atom Hartree-Fock energy eigenstate. For example, Schmidt et. al. found antibonding orbitals by performing an SVD of the overlap between the virtual orbitals and a minimal basis to produce valence virtual orbitals [189]. Some methods have been developed to produce a minimal basis specifically adapted to a molecular environment [121, 201], but those are non-linear optimization procedures that are often iterative and costly. One famous method that does not rely on a reference minimal basis is the Natural Bond Orbital (NBO) procedure [190, 191], where the density matrix coupling between multiple atom-tagged orbitals is utilized to produce bonding and anti-bonding orbitals. However, atom tagging of basis functions plays a critical role in the NBO procedure – in fact, the standard NBO method is specific to atom-centered orbital (AO) basis calculations. Few methods cut the umbilical cord to the minimal basis in producing antibonding orbitals. Aside from the Sano antibonding orbitals [146] (discussed below), Foster and Boys [152] suggested oscillator orbitals which are virtual orbitals with the maximum dipole from localized occupied orbitals.

Local correlation has been intensively studied [33, 34, 151, 200, 202–209], leading to the conclusion that dynamic correlation can be well approximated using domains of localized virtual orbitals that are in the same spatial region as a localized occupied orbital [33, 34]. This reduces the 4th rank tensor of pair correlation amplitudes to an asymptotically linear number of significant elements. Nevertheless, all virtual orbitals are required for post-SCF methods such as coupled cluster theory that recover dynamic correlation, rather than just the much smaller set of valence virtual orbitals. By contrast, static or strong correlation, resides mostly in the valence virtuals (i.e. the antibonding orbitals). Thus complete active space (CAS) methods that seek to describe strong correlation require only a description of the valence virtuals. Methods in this class include CASSCF, [6, 39–41, 210] spin-coupled valence bond (VB) [53, 67, 211, 212], and approximations such as generalized valence bond (GVB) [57], coupled cluster valence bond (CCVB) [61, 64, 65], etc. CAS, GVB and CCVB methods thus need an initial guess for the antibonding orbitals. We do note that the orbitals associated with key amplitudes for strong correlation are not necessarily spatially localized [213–216].

One method used to obtain initial guess antibonding orbitals is the so-called Sano procedure [146]. In brief, after localizing a set of occupied orbitals using standard methods [142–144, 158], the Sano procedure finds the virtual orbital that has maximum exchange interaction with each given localized occupied orbital. The idea of maximizing exchange is

very old [138, 139] and comes from its predecessor, the modified virtual orbitals [217, 218] (note that modified virtual orbitals have been since used to refer to any non-canonical set of virtual orbitals [219]). The resulting orbitals are symmetrically orthogonalized to yield a set of valence antibonding orbitals. This method has worked quite well for GVB-PP and CCVB calculations in moderately sized basis sets [64, 65, 213, 220, 221]. In this work we will show that the Sano procedure shows undesirable behavior with increasing the size of the AO basis set. This motivates the need for a better behaved alternative. We suggest that finding the antibonding orbital which gives the largest first order perturbation amplitude for exciting an electron pair from a given bonding orbital is a suitable alternative. A range of numerical results confirm this to be the case. These antibonding orbitals can be viewed as a specific instance of orbital specific virtual [205–207].

2.2 Theory

Defining the set of antibonding orbitals

Solving the mean field Hartree-Fock (HF) equation self consistently gives the lowest energy single Slater determinant electronic wave function. To solve the many-body problem, one needs to include the missing correlation energy [222]. Second order Møller-Plesset (MP2) perturbation theory [22, 223] offers a useful and computationally inexpensive approximation to treat the correlation yielding the following expression in the case of restricted HF orbitals:

$$E^{(2)} = \sum_{ij}^{occ} \sum_{ab}^{virt} \tau_{ij}^{ab} \, (ia|jb) \tag{2.3}$$

where

$$\tau_{ij}^{ab} = \frac{2(ia|jb) - (ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.4}$$

This expression folds together contributions from the correlation of two electrons of opposite spin (OS), with amplitudes:

$$t_{ij}^{ab} = \frac{(ia|jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.5}$$

together with the contribution of correlations of electrons with the same spin. The two-electron repulsion integrals (ERIs) over spatial orbitals describing the interaction of each occupied with each virtual are:

$$(ia|jb) = \int d\mathbf{r}_1 \phi_i(\mathbf{r}_1)\phi_a(\mathbf{r}_1) \int d\mathbf{r}_2 r_{12}^{-1} \phi_j(\mathbf{r}_2)\phi_b(\mathbf{r}_2) \tag{2.6}$$

Let us collect the ERIs associated with occupied orbital i into the symmetric matrix $\mathbf{K}^i$, where:

$$K_{ab}^i = (ia|ib) \tag{2.7}$$

$\mathbf{K}^i$ is positive semi-definite, and thus the eigenvector belonging to its largest eigenvalue will correspond to the virtual level with the strongest exchange interaction with occupied level i. That is the Sano prescription [146] for finding the antibonding orbital associated with i.

We can likewise define a matrix of second order pair correlation amplitudes, $\mathbf{T}^i$, associated with a given occupied orbital:

$$T^i_{ab} = t^{ab}_{ii} \tag{2.8}$$

This matrix is negative semi-definite since the denominators are negative for the ground state determinant. We can therefore find the largest OS pair-correlation amplitude as the lowest eigenvalue, $t^i_{\max}$ of $\mathbf{T}^i$, and the associated virtual orbital, $|i^*\rangle = \sum_a |a\rangle\, c_{ai^*}$ is the eigenvector, with expansion coefficients c_{ai^*} in the original virtual basis:

$$\sum_b T^i_{ab} c_{bi^*} = t^i_{\max} c_{ai^*} \tag{2.9}$$

Upon repeating for each occupied level, most naturally in a localized representation, and using similar arguments to Kapuy's zeroth in the Fock and 2nd order in correlation approximation [224, 225], we suggest that this is an appropriate non-iterative way to find a set of antibonding orbitals in 1:1 correspondence with the bonding orbitals. This approach may be contrasted with Sano's suggestion to obtain the virtual orbital with maximum repulsion from the bonding orbital by solving the eigenvalue problem for each orbital using $\mathbf{K}^i$ rather than $\mathbf{T}^i$. Inclusion of orbital denominators in Eq. 2.9 provides a clear physical meaning of the antibonding orbital as having strongest pair correlation amplitude with its parent bonding orbital. As will be demonstrated numerically later, this property also dramatically improves basis set convergence relative to the Sano definition.

We will refer to these virtual orbitals as "second order antibonding" (AB2) MOs to emphasize their second order origins, and their 1:1 correspondence with bonding MOs. In terms of existing literature, the AB2s are directly related to the "orbital-specific virtual" (OSV) orbitals [205, 206] that are sometimes used to evaluate the correlation energy. Each AB2 orbital is the most important OSV for a given localized bonding orbital. Of course the reason for selecting the amplitudes associated with MP2 is computational efficiency. The exact limit of this procedure would be to diagonalize the corresponding exact (i.e. from Full CI) doubles amplitudes; T^i_{ab}, via Eq. 2.9.

A closely related alternative that has some advantages over Eq. 2.9 above is to define the space of valence antibonding orbitals from the virtual-virtual block of the MP2 one-particle density matrix [226, 227]:

$$P_{ab} = \sum_{ijc} t^{ac}_{ij} t^{bc}_{ij} \tag{2.10}$$

Upon diagonalizing, the $(M - O)$ eigenvectors with largest occupation numbers span the valence antibonding orbital space, and, together with the occupied space, complete the span of a molecule-adapted minimal basis. Localization of these valence virtual orbitals will then yield an alternative to the localized virtuals above. The advantage of this approach is for

cases where there is no simple 1:1 mapping between bonding and antibonding orbitals, as discussed more later.

The virtual orbitals obtained this way are the valence subset of the "frozen natural orbitals" (FNO) [226, 227], and we emphasize that they are not generally localized in contrast to the AB2 MOs. They are close to the virtual natural orbitals associated with P^{MP2} as defined by the gradient of the MP2 energy [228–230], with the caveat that only the virtual-virtual block is diagonalized.

Population analysis using the effective minimal basis

Finding a suitable set of antibonding orbitals provides the missing part of the valence space not spanned by the occupied orbitals. Thus the union of the occupied space and the space of antibonding orbitals spans the space of an effective minimal basis. It is well accepted that full valence CASSCF wavefunction is spanned by an effective minimal basis within the molecule for this reason [39, 40]. Accordingly, localizing the union of the occupied orbitals with the antibonding orbitals reveals a set of molecule-adapted atomic orbitals (MAOs) [9, 10]:

$$\mathbf{C}_{\mathrm{MAO}} = \{\mathbf{C}_{\text{non-bonding}}\} \oplus \{\mathrm{Localize}(\mathbf{C}_{\text{bonding}} \oplus \mathbf{C}_{\text{antibonding}})\} \tag{2.11}$$

For a given pair of well-localized bonding and antibonding orbitals (say σ and σ^*), this procedure amounts to inverting Eqs. 2.2 to discover the corresponding MAOs even though we may be using a very extended basis, or even a non-atom-centered basis, such as plane waves or a real-space grid, to perform the calculations.

The resulting MAOs, $\boldsymbol{\chi}$ are thus expressed in terms of the AO's, $\boldsymbol{\omega}$, as $\boldsymbol{\chi} = \boldsymbol{\omega}\mathbf{C}_{\mathrm{MAO}}$. The MAOs are orthogonal, and typically localize onto atoms. The MAOs exactly span the space of the occupied orbitals, and can be used for population analysis among other things [40, 116, 201, 231–236]. Let us denote p as an MAO label for χ_p, which is centered at $\mathbf{r}_p = \langle\chi_p|\mathbf{r}|\chi_p\rangle$. Using A, B as atom labels, and given that the density matrix in the MAO basis is $\mathbf{P}^{\mathrm{MAO}} = \mathbf{C}_{\mathrm{MAO}}^{\dagger}\mathbf{SPSC}_{\mathrm{MAO}}$, one can make a population analysis as follows:

$$p \in A \iff |\mathbf{r}_p - \mathbf{R}_A| = \min_B(|\mathbf{r}_p - \mathbf{R}_B|) \tag{2.12}$$

$$Q^A = Z^A - \sum_{p\in A} P^{\mathrm{MAO}}_{pp} \tag{2.13}$$

where Q^A and Z^A are the atomic charge and the nuclear charge, respectively. Such a population analysis has no dependence on atom-tagging of the underlying basis, and does not rely upon a reference minimal basis. Therefore it generalizes nicely to plane wave basis and real space methods. If the orbitals are unrestricted, we construct antibonding pairs and the MAOs for the alpha and beta spin spaces independently.

This approach to generating an MAO representation does have some limitations. First, it assumes that there is a 1:1 mapping between bonding and antibonding orbitals. One class of exceptions can be found in electron deficient molecules (e.g. LiH will not recover 2p-like

orbitals on Li, and BH_3 will not recover a $2p_z$ orbital on B). Such species can be said to have "virtual lone pairs", whose identification is a problem that we shall not address here. A second class of exceptions lie in species such as cyclopentadiene anion, where there are 3 semi-localized π occupied orbitals, but the valence space only admits 2 antibonding orbitals. Thirdly, in symmetric systems with multiple Lewis structures (e.g. C_6H_6), the MAOs will derive from localized bond and antibonding orbitals corresponding to a single Lewis structure and may not reflect the indistinguishability of the atoms. Broadly, we can say that this MAO approach is readily applicable to neutral molecules with a single dominant Lewis structure.

Implementation details

Computational efficiency is very important for quantum chemistry in order treat molecules that are as large as possible for given computational resources (computer speed, memory size, etc). Our AB2 implementation uses exact 4-center integrals in a basis of Gaussian-type atomic orbitals (other alternatives such as using auxiliary basis expansions can also be readily implemented). Each step with its computational complexity is shown in Fig. 2.1. Note that for the figure and the discussion here we use O, V, and N for the number of bonding orbitals, virtual orbitals, and AO basis functions, respectively. We start by making a pseudo-density $\mathbf{P}^i = \mathbf{C}_i\mathbf{C}_i^\dagger$ for each bonding orbital, i. To generate the two-electron integrals $(\mu\nu|\lambda\sigma)$, Q-Chem [237] only generates significant $\mu\nu$ (i.e. AO basis) pairs to some target numerical cutoff, yielding a total that we term as $(NN)_{\text{cut}}$. $(NN)_{\text{cut}}$ scales quadratically (i.e. $(NN)_{\text{cut}} \approx N^2$) for small systems but approaches linear scaling (i.e. $(NN)_{\text{cut}} \propto N$) in the limit of large system size. The integrals are made and contracted on-the-fly with the bonding orbitals' pseudo-densities to make bonding-specific exchange integrals $K^i_{\mu\nu}$ with compute effort scaling as $\mathcal{O}(O(NN)^2_{\text{cut}})$. The $K^i_{\mu\nu}$ matrices are then transformed into the virtual space as K^i_{ab} in Eq. 2.7 with compute cost scaling as $\mathcal{O}(OVN^2+OV^2N)$. Asymptotically this is the dominant step in this method unless more careful thresholding is considered [238]. Then, we divide by the appropriate denominator to get T^i_{ab} in Eq. 2.8 (with $\mathcal{O}(OV^2)$ effort). Lastly, we diagonalize $\mathbf{T}^i$ for each bonding orbital to get the AB2 antibonding orbitals as in Eq. 2.9 with $\mathcal{O}(OV^3)$ effort. Note that the last step can in principle be made $\mathcal{O}(OV^2)$ since we are only solving for the eigenvector with the largest amplitude in each matrix. We can contrast this procedure with the modified FNO approach which has a dominant computational step that scales as the 5^{th} power of molecule size: constructing P_{ab} in Eq. 2.10 with complexity of $\mathcal{O}(O^2V^3)$.

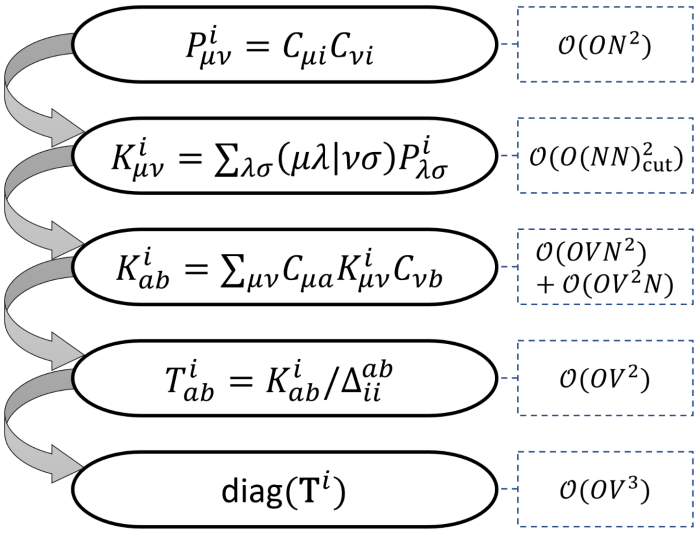


Figure 2.1: A chart illustrating the mathematical steps needed to construct AB2 orbitals with the appropriate computational complexity for each step indicated. Here, O, V, N, and $(NN)_{\text{cut}}$ refer to the number of occupied orbitals, virtual orbitals, AO basis functions, and significant AO pairs, respectively.

One reason for the efficiency of the AB2 approach compared to FNO comes from focusing on the bonding orbitals one at a time rather than the whole occupied space at once. It is then important to start by localizing the occupied space, which is known to be a cubic scaling iterative procedure for e.g. the Boys and Pipek-Mezey localization measures [239, 240]. Then, one must also distinguish between localized orbitals with different character: specifically core, bonding, and non-bonding, e.g. lone pairs. Our implementation uses an automatic bonding detection option that runs before AB2. The detection process is simply determined by Pipek's delocalization measure [241] on Mulliken charges, where measures amounting to 1 indicate an orbital localized on an atom (core or non-bonding) and measures around 2 correspond to orbitals split between two atoms.

Computational details

All methods discussed here were implemented in a developer version of Q-Chem 5 [237]. The geometries used for molecular calculations were optimized at the ωB97X-D/def2-TZVPD level of theory. All geometries are included in the Supplementary Material (SI).

2.3 Results and discussion

We will compare different approaches to generating effective antibonding orbitals: in particular we are interested in whether the second order antibonding (AB2) MOs significantly

improve upon the Sano antibonding orbitals, as measured by usage-relevant metrics obtained from a set of numerical experiments. We will first examine orbital plots, orbital energies, and orbital variances. We then test the applicability of Sano and AB2 MOs to several valence correlation methods: coupled cluster valence bond (CCVB) [64, 65], complete active space configuration interaction (CASCI) [242–245], and complete active space self-consistent field (CASSCF) [39–41, 210]. Next, we look into their uses for describing valence excited states. For basis set, we are using the Dunning basis set family [246] and Ahlrichs [247]. These are available in Q-Chem 5.3 with an automated detection of bonding orbitals.

Orbitals, orbital Energy, and orbital variance

We start by looking at the σ^* orbital of H_2, as shown in Fig. 2.2, evaluated by the Sano procedure, the AB2 approach, and CAS(2,2) (performed as 1-pair perfect pairing). It is visually clear that the Sano σ^* orbital is contracting as the basis set is improved. Fig. 2.3 displays the orbital energy (diagonal matrix element of the Fock operator) and the variance ($\langle r^2 \rangle - \langle \mathbf{r} \rangle^2$) of the σ bonding orbital, and the Sano and AB2 models of the antibonding orbital. The variance confirms that the size of the Sano σ^*-orbital contracts with basis size, while its orbital energy increases (reflecting increasing electron confinement) unsatisfactorily. By contrast the behavior of the AB2 orbital is very close to the bonding orbital, with pleasing stability in both energy and variance as the basis set is converged towards completeness. The stark difference is due to Sano orbitals including high energy orbitals to maximize the exchange interaction whereas AB2 biases against those higher energy orbitals with the denominator penalty in Eq. 2.8.

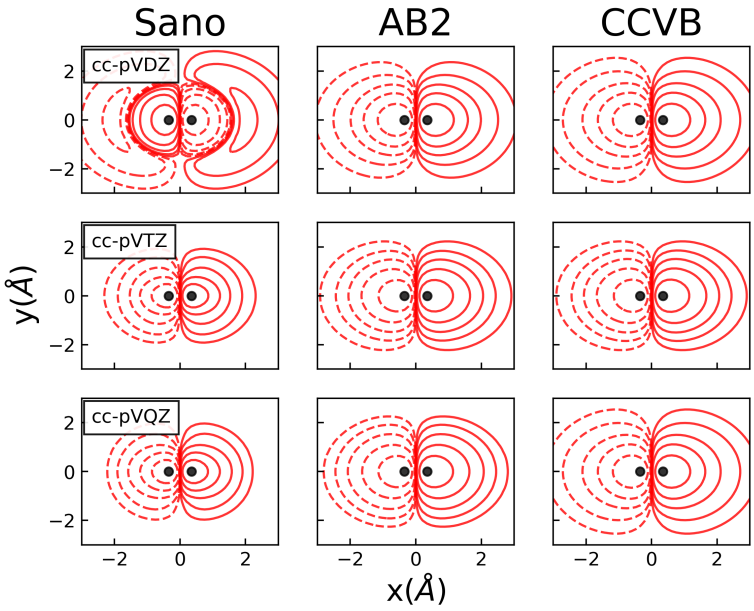


Figure 2.2: Comparison of σ^* orbitals predicted by Sano, AB2, and CCVB in H_2 with increasing size of the basis set. For this problem, CCVB is identical with (2,2) CASSCF. Orbitals were plotted with 10 contour isovalues logarithmically spaced [0.1,10], 5 for each phase.

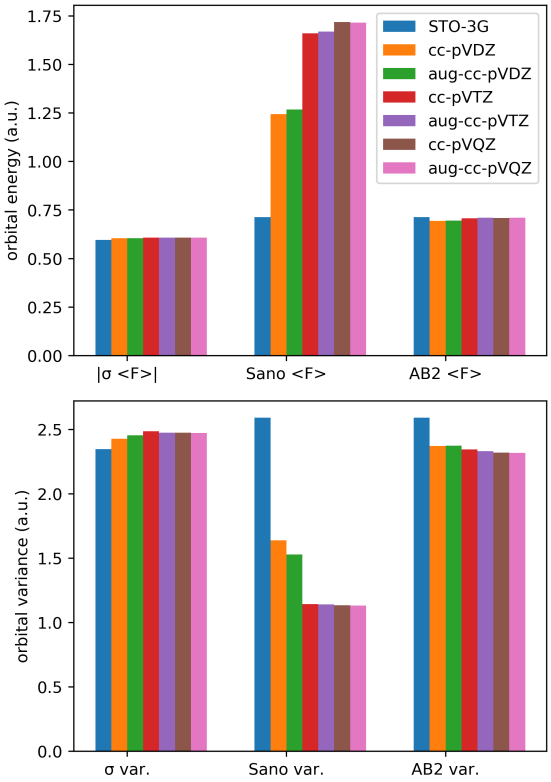


Figure 2.3: Comparison of orbital energy (diagonal matrix element of the Fock operator) for σ, Sano, and AB2 orbitals for H_2 with increasing size of the basis set. Bottom graph compares the variance.

Next we look into C_2H_4, where the localization scheme of Boys produces mixed $\sigma - \pi$ orbitals (sometimes called banana bonds), while Pipek-Mezey predicts separate σ and π orbitals. We will therefore use Pipek-Mezey orbitals whenever we encounter π orbitals. Inspecting the σ C-C bond in C_2H_4 in Fig. 2.4 shows that the shape of the occupied Pipek-Mezey and converged CCVB bonding orbitals both do not change much upon increasing the size of the basis set. By contrast, when looking at σ^* in Fig. 2.4 we see even poorer behaviour of the Sano C-C antibonding orbital as a function of basis set size than we did for H_2. This is confirmed in Fig. 2.5 where we compare the orbital energy and the orbital variance of the bonding and the antibonding C-C σ orbital in C_2H_4. The Sano σ^* orbital does not converge with the size of the basis set, with the variance decreasing, and the energy increasing. By contrast, the AB2 σ^* orbital converges rapidly both in terms of energy and variance for similar reasons to before.

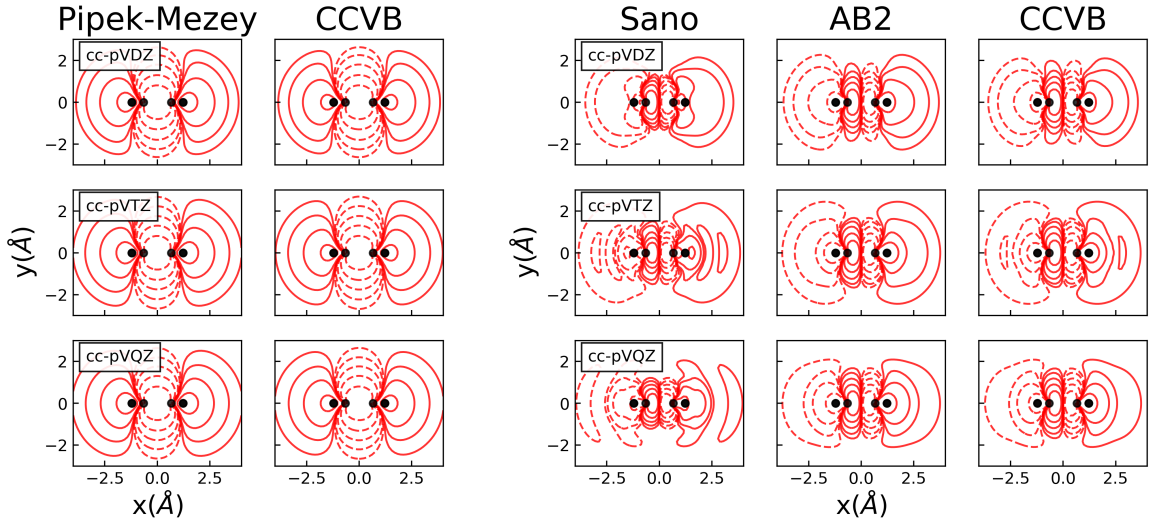
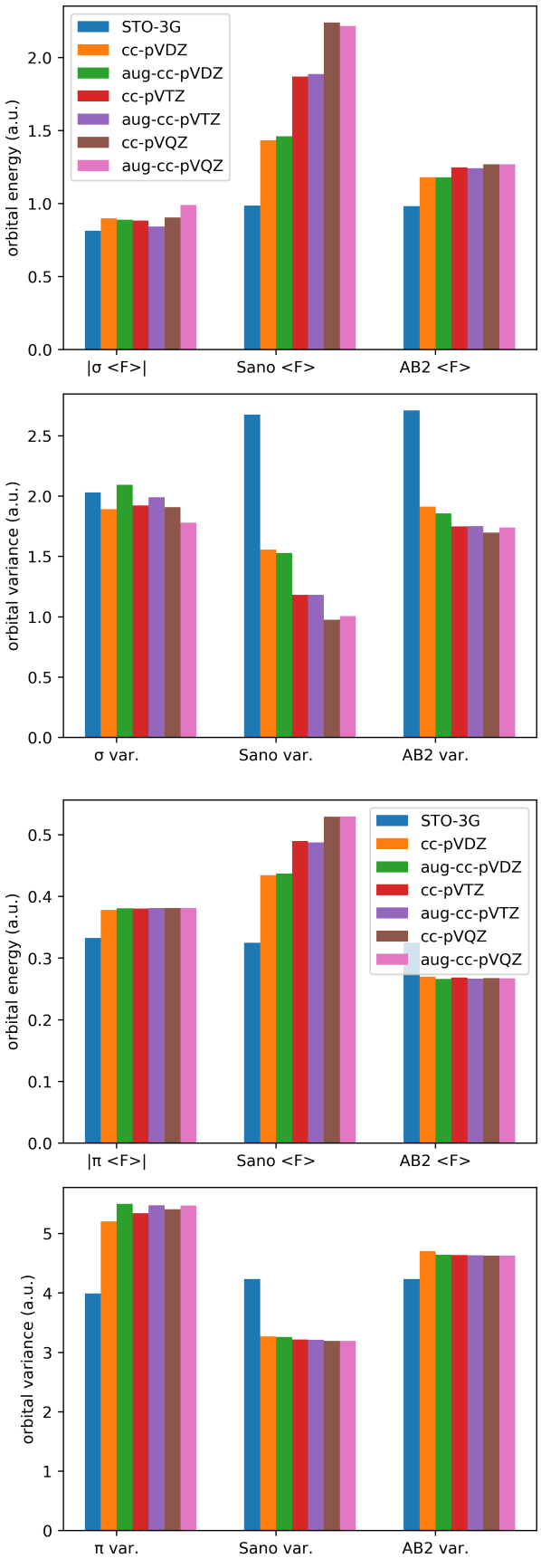


Figure 2.4: Comparison of σ orbitals predicted by Pipek-Mezey localization with those found by converging CCVB (top row) and σ^* orbitals predicted by Sano, AB2, and CCVB (bottom row) in C_2H_4 with increasing size of the basis set. Orbitals were plotted with 10 contour isovalues logarithmically spaced [0.1,10], 5 for each phase.

In Fig. 2.5 we compare the orbital energy and the orbital variance of the bonding and the antibonding orbitals for the C-C π in C_2H_4. The shortcomings of Sano seem to be much less severe in π^* orbitals. We believe this is due to the diffuse nature of the π orbitals making the maximum exchange, thus spatial locality, sufficient to describe the π^*. However, we can still see that the orbital energy and variance do not converge for Sano while they do for AB2, and converge to drastically different orbital energy and orbital variance.


STO-3G
cc-pVDZ
aug-cc-pVDZ
cc-pVTZ
aug-cc-pVTZ
cc-pVQZ
aug-cc-pVQZ
orbital energy (a.u.)
|σ <F>|
Sano <F>
AB2 <F>
orbital variance (a.u.)
σ var.
Sano var.
AB2 var.
orbital energy (a.u.)
|π <F>|
Sano <F>
AB2 <F>
orbital variance (a.u.)
π var.
Sano var.
AB2 var.

The quantitative advantage of the AB2 antibonding orbitals relative to the Sano orbitals seen so far can also become qualitative advantages in systems with more complex electronic structure. One such example is Cu_2, which, considering that the valence state of Cu can be taken as $3d^{10}4s^1$, is isoelectronic to H_2. The σ orbital (HOMO) of Cu_2 is shown in the upper panel of Fig. 2.6, along with the optimized correlating orbital from CCVB, as well as the Sano and AB2 antibonding orbitals. Maximizing exchange results in a Sano antibonding orbital that resembles an empty π-bond between the two metals. By contrast, the AB2 and CCVB orbitals look qualitatively identical.

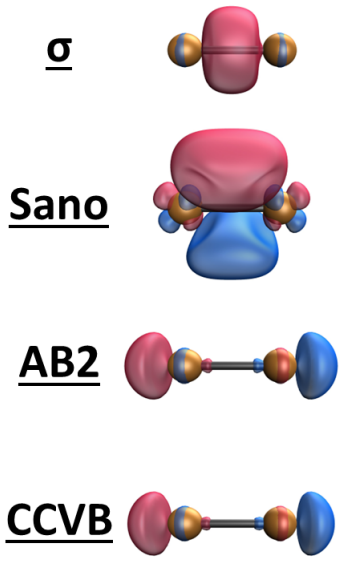


Figure 2.6: Comparison of the shape of the orbitals in Cu_2 where the σ bond is used to produce Sano and AB2 antibonding orbitals. While the AB2 method produces very similar orbitals to CCVB, the Sano approach fails to give a qualitatively correct antibonding orbital.

CCVB iterations

The CCVB method is a simple low-scaling approximation [64, 65, 213] to exponentially scaling spin-coupled valence bond theory that can separate a system of $2n$ electrons into fragments with spin purity, provided that UHF can also reach the dissociation limit. One price to be paid for these advantages is a challenging orbital optimization problem: the CCVB orbitals have no invariances to rotations within the active space, in contrast to CASSCF. Hence a good initial guess is very important. Sano orbitals [146] have been commonly as a starting guess for valence bond methods [63, 213] such as CCVB due to their resemblance to antibonding orbitals. For simple alkanes, we examine how many iterations are needed

to converge a CCVB calculation with Sano and compare with AB2 shown in Fig. 2.7 with increasing the molecule size and the basis set size (using the Dunning cc-pVXZ sequence of basis sets [246]). Since the double-*zeta* basis set does not involve many high energy orbitals, both methods converge almost at the same speed. Upon increasing the size of the basis set, overly-contracted Sano orbitals deviate more from the optimal antibonding orbitals, and therefore require far more iterations to converge. For this reason, we recommend using AB2 orbitals as a starting guess for valence bond methods instead of the Sano orbitals.

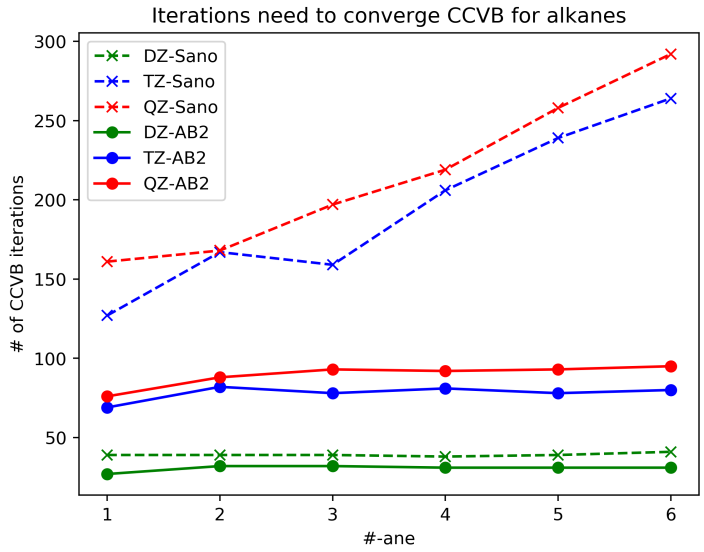


Figure 2.7: Number of iterations needed to converge CCVB calculations on alkanes of increasing size, with increasing ζ of the basis set. This shows a relatively constant number of iterations needed for AB2 regardless of system size, while the number of iterations rise unfavorably for the Sano guess in large basis sets. Geometric direct minimization (GDM) [63, 248] is used to determine the steps.

CAS methods

The relative fraction of correlation energy recovered using AB2, Sano, FNO or other choices for antibonding orbitals to complete an active space can help us discern which ones are most appropriate to use for configuration interaction with fixed orbitals, as well as for a CASSCF initial guess. As a simple example, we stretch the C-C bond in C_2H_4 while keeping the geometry of the methylene groups fixed at those of the equilibrium ground state geometry of ethene. Looking at Fig. 2.8, we see that canonical virtual orbitals capture less and less correlation as the def2 basis set is improved from SVP to TZVPP to QZVPP. We also observe that the gap between Sano and AB2 orbitals increases with increasing the size of the basis set. Finally, we can see that FNO and AB2 orbitals perform almost identically and are the

best choices, with AB2 having lower compute costs. Nonetheless, it is encouraging that both follow the CASSCF energies closely.

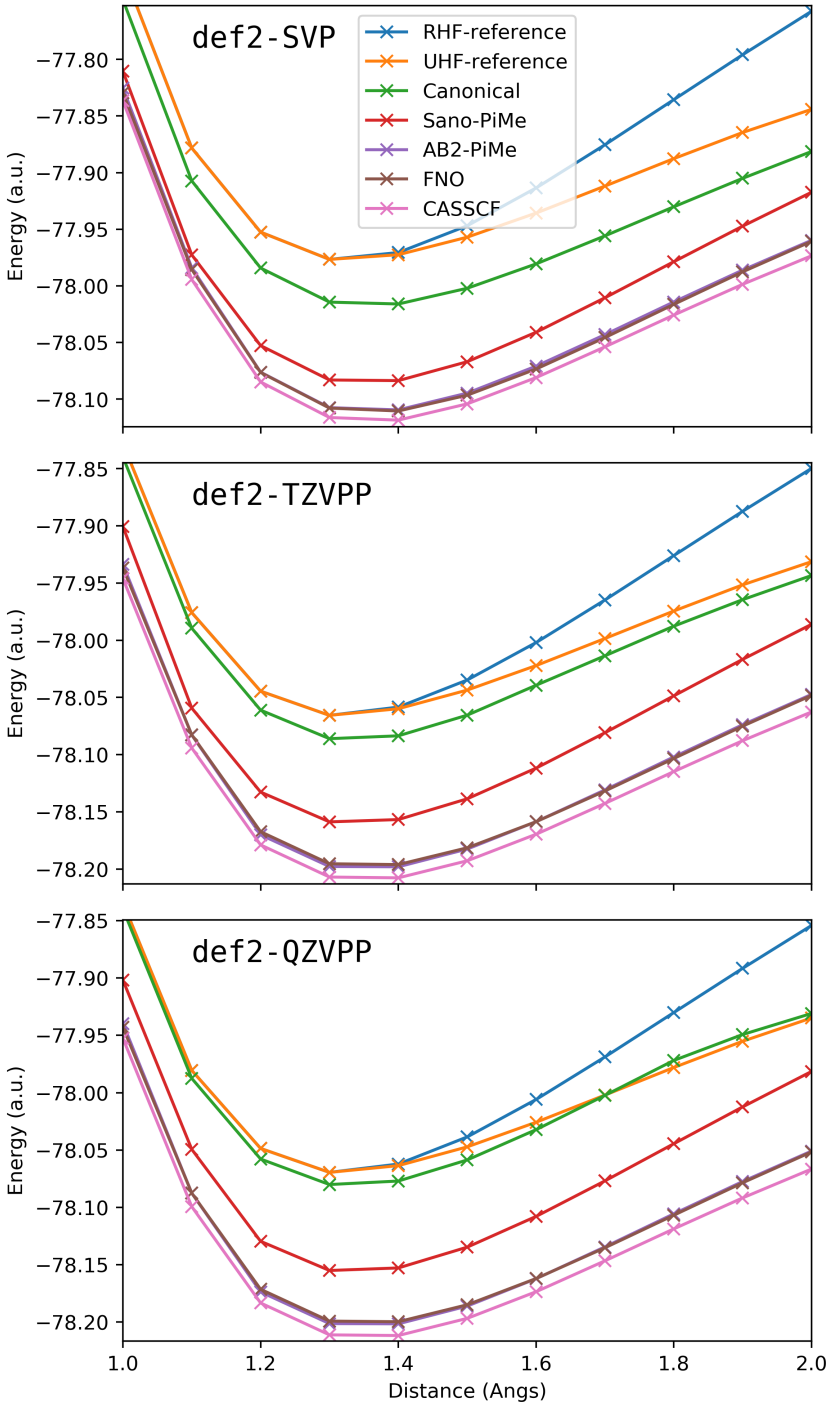


Figure 2.8: CAS-CI (12e,12o) for C_2H_4 using canonical, Sano, AB2, and frozen natural orbitals in three different basis sets. Restricted HF (RHF) and Unrestricted HF (UHF) curves without any correlation correction are shown for comparison.

Since the AB2 and FNO orbitals seem to capture quite a lot of the static correlation, we sought to compare them to CASSCF orbitals. In Fig. 2.9 we are comparing the smallest singular value of the overlap matrix between the CASSCF orbitals and those of canonical, Sano, AB2, and FNO, at the optimized geometry of C_2H_4. Once again, the canonical orbitals become dramatically worse with increasing the basis set size. Sano and FNO both become very slightly worse with increasing the size of the basis set, namely by increasing *zeta*, while AB2 seems to be nearly basis set-independent.

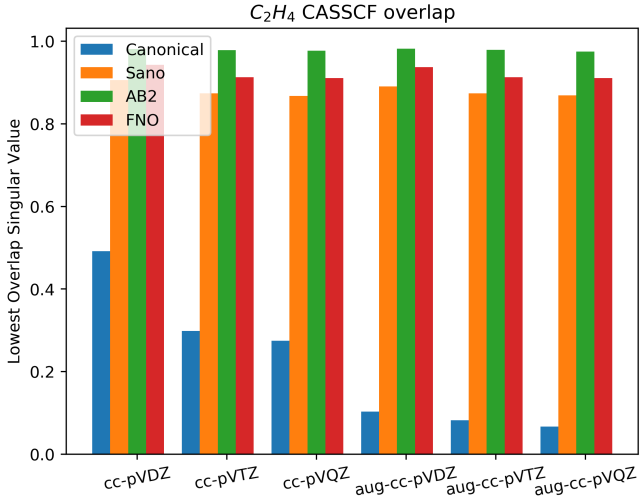


Figure 2.9: The smallest singular value from the overlap of CASSCF (12e,12o) orbitals with those from Sano, AB2, FNO and canonical orbitals. Canonical orbitals with the lowest energy and FNOs with the highest occupancy were selected. Canonical orbitals are differ strongly from optimized CASSCF orbitals while AB2 orbitals have the highest agreement.

Excited States

Since the AB2 orbitals seem to be good guesses for GVB methods, and yield orbitals close to converged CASSCF orbitals, this led us to believe that they could also provide a good description of valence excited states. State-specific methods, such as orbital-optimized DFT (OO-DFT) [249] need a suitable starting guess, as convergence is typically to the nearest stationary point [250], so we used Sano and AB2 guesses for the $\pi \rightarrow \pi^*$ excitation in methanal (H_2CO). For our purposes we employed the square gradient minimization method [250] which looks for saddle points in the orbital Hilbert space to converge restricted open-shell Kohn-Sham (ROKS) [249, 251, 252]. In Fig. 2.10 we compare the overlap of the π^* orbital from converged singlet open shell HF calculations with Sano and AB2 orbitals. For this excitation, AB2 orbitals overlap the optimized orbital by at least 0.9, and vary minimally with the size of the basis set. We note here that aside from the double-*zeta* case, converging

the excited state starting from the Sano orbital sometimes lands on a Rydberg excited state, while AB2 landed on the correct π^* state in all cases.

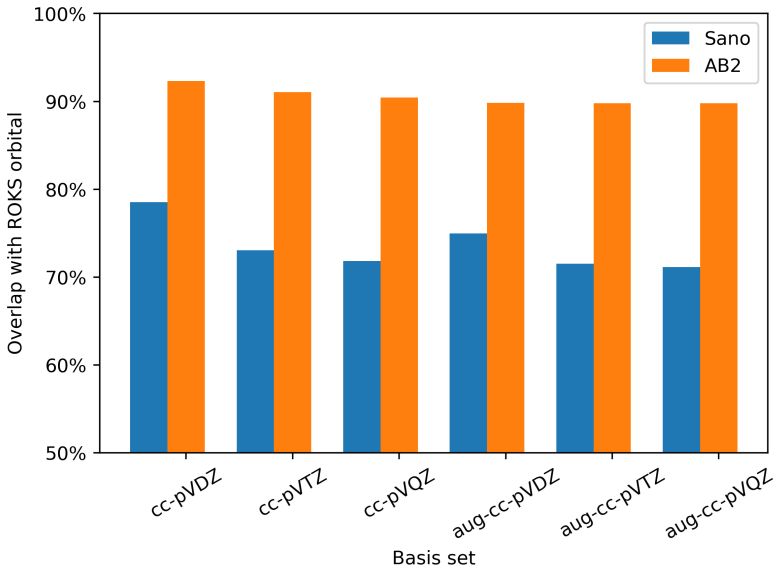


Figure 2.10: The overlap of the converged ROKS-HF antibonding orbital with the Sano and AB2 initial guesses in H_2CO for the $\pi \rightarrow \pi^*$ excitation. The π^* orbital is well described by AB2 regardless of basis set size.

Population Analysis

Antibonding orbitals belong to the valence space, and contribute to making a minimal basis that can be used to gain insight into chemistry, for instance via population analysis to assign effective charges on each atom. The population analysis we present here is constructed from the union of the occupied space and the antibonding orbitals without dependence on the basis set used. To study our atomic charge predictions and compare it to some other methods in the literature, we look into fluoro- and chloro-substituted methanes which have been studied theoretically [253–255] and experimentally [256, 257]. These simple systems are nonetheless interesting because they manifest the effect of substituting electron withdrawing halogen atoms of different sizes and electronegativities for hydrogen in methane. How consistent or inconsistent are different atomic population analysis schemes as descriptors of these chemical substitutions?

In Fig. 2.11 we examine the effect of progressive substitution of hydrogen by chlorine and fluorine in the methane molecule on the computed net charge at the C atom. We consider some commonly used methods, specifically charges on electrostatic potential grid (ChElPG) [128], iterative Hirshfeld (Iter-Hirsh) [129, 130], intrinsic atomic orbitals (IAO) [122, 258], and

the method presented in this work, molecular atomic orbitals (MAO). Most obviously, the charge transferred upon halogen substitution will depend strongly on the electronegativity difference between X and H. Furthermore, while halogens are more electronegative than hydrogen (or carbon), the electron donating capacity of C is not unlimited, and so we expect the first halogen substituted to pull away a greater fraction of an electron from C compared to the next, and so forth. Such a change will also have some dependence on the X vs H electronegativity difference. With these preambles aside, atomic charges are not observables and therefore no single answer should be viewed as strictly correct. Nevertheless, we can examine the results of each population analysis for signs of *incorrectness* relative to physical intuition.

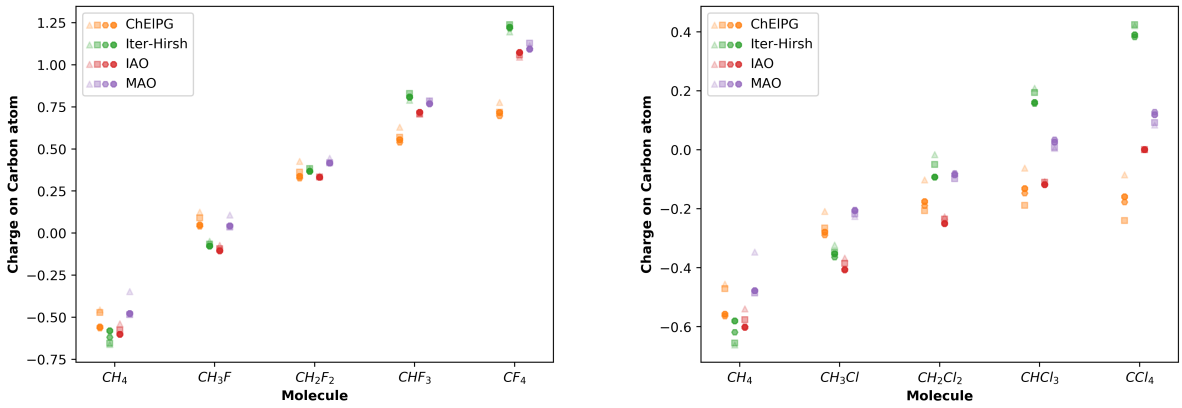


Figure 2.11: The charge on the carbon atom for successive chlorination and fluorination of methane predicted using four different population analysis methods (see text for the names). The triangle, square, hexagon, and octagon correspond to charges using def2-SV(P), def2-SVPD, def2-TZVPD, and def2-QZVPD, respectively.

For instance, while all methods agree that the C–H bonds of CH_4 are polarized $C^{\delta-}H^{\delta+}$, and all likewise agree that the C–F bonds of CF_4 are polarized $C^{\delta+}F^{\delta-}$, different methods predict different polarities for the C–Cl bond in CCl_4. Perhaps the most counterintuitive result is that the population on C becomes more negative via ChElPG upon going from $CHCl_3$ to CCl_4 despite the higher electronegativity of Cl vs H. At the other extreme, iterative Hirshfeld suggests that the change in C population with successive halogenation is linear, as if the electron-donating capacity of C does not saturate. This is especially striking for chlorination, and we suspect, as unreasonable as the ChElPG result for CCl_4. By contrast, we find no obvious fault with the MAO values, or with the IAO values for these interesting test cases, although we prefer the slight negative charge for Cl in CCl_4 predicted by MAO (recalling the electronegativity of chlorine is 3.5 versus carbon at 2.5 on the Pauling scale) relative to the slight positive charge predicted by IAO.

Finally we examine an unusual linear molecule, which is the result of insertion of Be into HCl, yielding H−Be−Cl [259, 260]. While H−Cl is polarized as $H^{\delta+}Cl^{\delta-}$, Be has lower electronegativity than H, and so there will be substantial charge transfer. Indeed the ionic limit would be $H^{-}Be^{2+}Cl^{-}$. What then, is the actual charge distribution when we consider the covalent character of the molecular orbitals? The calculated populations are shown in Fig. 2.12, and it is immediately evident that predicted charges on Be vary widely. The least polar picture comes from ChElPG and MAO, with q(Be)≈+0.5, while the IAO scheme suggests q(Be)≈+1.35. How should we understand this dramatic difference and suggest which might be more correct?

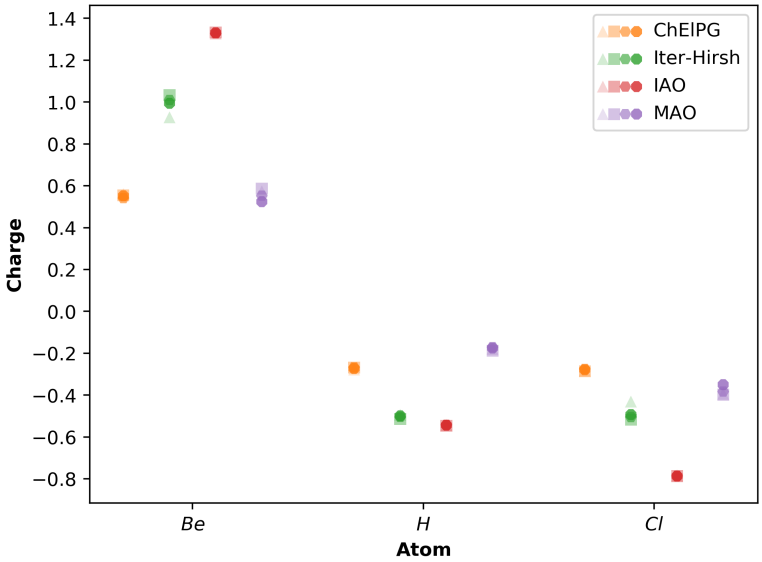


Figure 2.12: The charges on each atom in the BeHCl molecule predicted by the four methods mentioned in the text. The triangle, square, hexagon, and octagon correspond to charges using def2-SV(P), def2-SVPD, def2-TZVPD, and def2-QZVPD, respectively.

From the MAO perspective, there are two σ bonds involving Be, one with H and one with Cl. Each is made from *sp* hybrid orbitals on Be, meaning that the p orbitals of Be are at play in this σ bonding, as shown in Fig. 2.13. These bonds are both polarized away from Be, as expected. The origin of the much larger IAO charge can now be understood. The IAO reference minimal basis set, known as 'MINAO' [122], does not include 2p orbitals for Be, and therefore we are instead seeing essentially only the Be(2s) charge via the IAO approach! Iterative Hirshfeld evidently struggles with a similar issue, leading to similar overestimation of Be charge. Overall, this case nicely illustrates the advantages of the MAO population scheme that is based entirely on the system at hand, rather than some reference atomic orbitals or states.

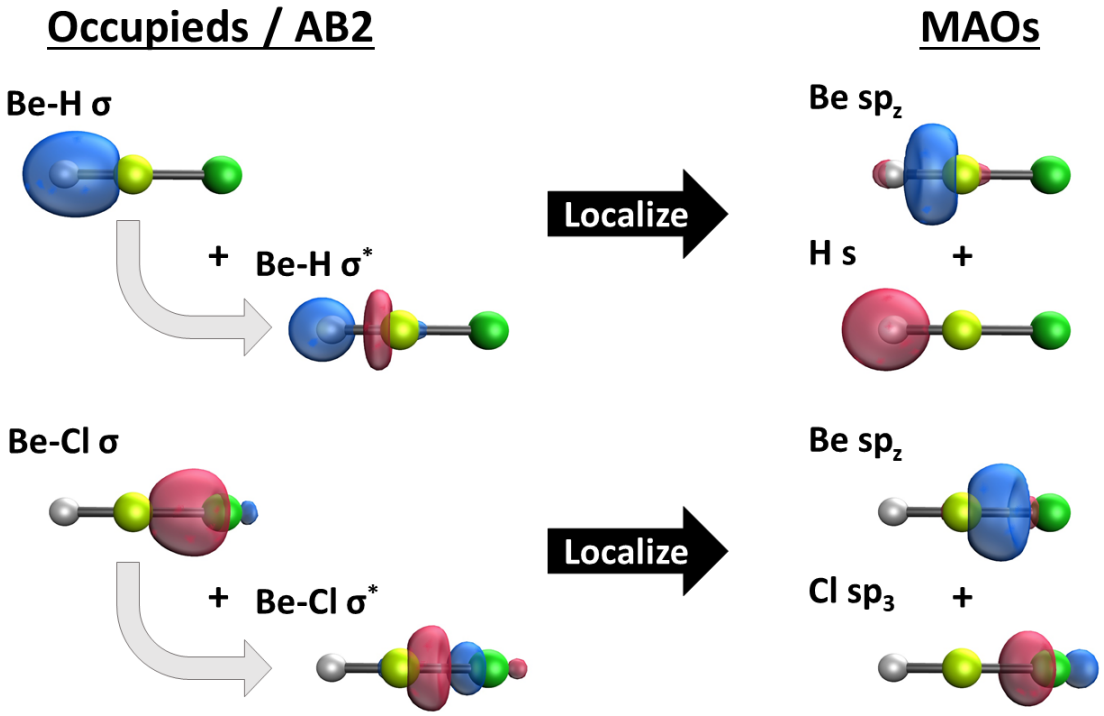


Figure 2.13: The union of the Boys localized bonding orbitals and their AB2 counterparts in BeHCl forms a complete valence space. Performing a Boys localization on this set of valence orbitals leads to molecule-adapted atomic orbitals that are intuitive and centered on atoms, which can then be used for population analysis.

Next we study the hypervalent molecule, SF_6, which has O_h symmetry, and whose chemical bonding has long been of interest [261]. While empty 3d functions on sulfur are needed to form 6 equivalent sp^3d^2 hybrids, the energetic cost of promoting electrons to the 3d shell is too high for d-orbital participation in the bonding to be chemically important [67, 262–265]. Rather, the bonding may be thought of as resonance between Lewis structures with 4 covalent S-F bonds, and 2 F^- anions, with a formal charge of +2 on S [266]. Using Boys localization produces 6 equivalent σ_{SF} orbitals, as shown at the left of Fig. 2.14. As expected, these σ_{SF} bonds are strongly polarized towards the more electronegative fluorine atom. The AB2 antibonding orbitals, also shown on the left of Fig. 2.14 are fascinating because contrary to simple chemical expectations, they are not strongly polarized towards the sulfur atom.

Localizing the union of bonding and antibonding sets produces atom-centered MAOs, where the sulfur valence orbitals are a set of 6 equivalent orbitals that are strongly polarized towards the fluorine atoms, as shown in the third image of Fig. 2.14. These 6 functions are linearly independent (though non-orthogonal), and combine with the conventional hybrid orbital on each F to form the strongly polarized σ_{SF} orbitals. This hypervalent bonding

problem is treated very naturally by the MAO analysis, while a minimal basis is insufficient to describe the occupied space since one in principle needs a set of 6 sp orbitals on sulfur. As measures of polarity, the calculated charges on S are +2.9 (IAO), +2.1 (Iter-Hirsh), and +1.6 (MAO) in the def2-QZVPPD basis set. The IAO charge may overestimate the polarity due to its minimal basis on S, while the MAO charge surely underestimates it, because the MAO orbital assigned to S is actually quite strongly polarized towards F. So while the MAOs themselves provide interesting chemical insight, they cannot resolve intrinsic limitations of population analysis.

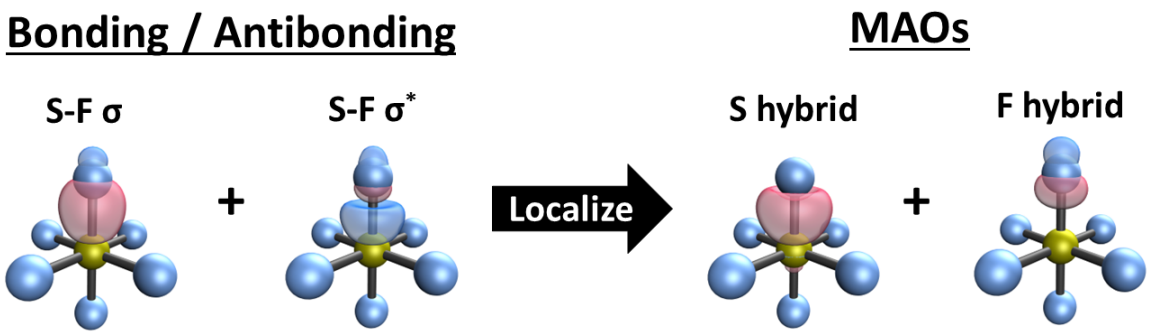


Figure 2.14: The union of the localized σ_{SF} bonding orbitals and the corresponding AB2 σ^*_{SF} antibonding orbitals in SF_6 forms a complete valence space describing the SF bonds (and excluding the F lone pairs). Localizing the set of valence orbitals leads to 6 molecule-adapted atomic orbitals on S (one is illustrated), showing no visual signs of d orbital participation.

There are some limitations associated with reducing the union of the localized occupied orbitals and the AB2 antibonding orbitals to a set of MAOs that should be mentioned. First, some conjugated π systems, such as benzene and $C_5H_5^-$, present a multiple minimum solution problem for orbital localization methods. Since our method relies on the localization procedure heavily, we expect there will be inconsistencies in these systems. For example, in benzene, there are different sets of solutions for the localized π orbitals, nominally corresponding to the two different Kekule structures. Using the Boys localized orbitals yields populations that reflect D_{6H} symmetry, while the Pipek-Mezey scheme gives alternating charges on successive carbons going around the ring. There is a second class of molecules that are inaccessible in our method. These are anions where the natural valence minimal basis is too small to provide an antibonding orbital for each bonding orbital. One such example is $C_5H_5^-$, the cyclopentadienyl anion. Forming the set of AB2 valence orbitals and taking the union with the occupied space leads to a set of orbitals that cannot be localized to atoms. Broadly, we can say that neutral species with a single Lewis structure are well-handled by the approach described here; as well as some more complex bonding situations like SF_6 discussed above.

2.4 Conclusion

We presented a relatively cheap, non-iterative procedure to produce a set of antibonding orbitals that vary minimally with the size of the atomic orbital basis set from which they are constructed. Specifically, antibonding second order (AB2) orbitals show far less variation with basis than the Sano orbitals which are sometimes used as valence antibonding orbitals. We showed that use of AB2 rather than Sano orbitals as initial guesses provides improved convergence for valence bond methods (specifically CCVB), as well as for CASSCF. The AB2 orbitals were successfully used as guesses for state-specific ROKS calculations of excited states, where they better resemble the converged orbitals than does the corresponding Sano orbital guess. We have shown how these AB2 orbitals can be used with the localized occupied orbitals to construct an effective minimal basis that can be used for population analysis among other things. Population analysis on the substituted fluormethane and chloromethane sequence shows the method is stable and consistent with other common methods that accord with chemical intuition. For the insertion of Be into HCl, the resulting charges show some advantages. Overall, the AB2 antibonding orbitals are relatively efficient to compute and quite useful for a variety of applications in quantum chemistry.

Chapter 3

Oxidation State Localized Orbitals

3.1 Introduction

The oxidation state (OS) [134] is a venerable concept reaching back to the early days of chemistry where the "oxydationsstufe" was introduced to rationalize the products obtained from reactions with oxygen. The electron-gathering tendency of oxygen is captured via its normal OS of -2 in compounds with ionic interactions, which is but one of the generally accepted counting rules to assign the OS. After a thorough revision of the concept, the IUPAC defines the OS of an atom as the charge of this atom after ionic approximation of its hetero-nuclear bonds [132]. They further recommend that this is accomplished by writing the Lewis structure of the compound of interest, and partitioning the electron pairs such that each shared electron pair is given to the more electronegative of the two associated atoms [133]. The IUPAC procedure is simple and generally effective, and for these reasons should be the first resort in assigning OSs in new compounds of interest.

While the OS is a chemical concept of enduring value, it must be stressed that nevertheless, the OS of an atom is not itself a precisely defined observable. It *may* correlate with observables such as x-ray absorption spectral shifts, but this requires calibration. Ultimately, the validity of the OS depends on the extent of ionicity in the bonding. Thus the OS becomes less well-defined as the chemical bonding approaches the covalent limit of electron pair sharing. Other situations such as ligand non-innocence [267] also can defeat normal OS conventions. Indeed the IUAPC report on OSs in chemistry states that there are "limits, beyond which OS ceases to be well-defined or becomes ambiguous". This situation is no different than other valuable chemical concepts such as aromaticity [268–270], and should not be viewed as a reason to discard the OS as something that cannot be measured. Instead, it is a reason to have tools that go beyond electron counting to assess the electron distribution in interesting and challenging borderline cases.

Electronic structure calculations directly yield the electron density, and therefore offer an ideal starting point for probing the borderline cases. Thus the assignment of OSs in molecular systems has drawn continuing attention in recent years [271–279]. Beyond the electron

density itself, there is particular interest in the development and application of specific schemes to extract OSs from electronic structure calculations, going beyond the (simple but clearly not satisfactory) use of partial atomic charges or atomic spin densities [280–283]. Most electronic structure approaches to OS assignment are predicated on assigning each electron pair (or individual electrons in case of open-shell systems) to one atom or ligand within the system based on some strategy that generalizes simple counting approaches such as the IUPAC definition. While we will concentrate on molecular systems in this work, it must be mentioned that precisely the same issue exists for oxidation state assignments in solid state materials [284–286].

Some years ago, Ramos-Cordoba et al. introduced a general OS elucidation method applicable to any molecular system and wavefunction (single-determinant or correlated) [282] that relies on Mayer's effective fragment orbitals (EFOs) and their occupations [287, 288]. The EFOs are obtained by diagonalization of the fragment's density matrix, according to some atom-in-molecule definition. For instance, in the case of QTAIM they lie spatially within the fragment's boundary, so they represent natural domain orbitals. The spin-resolved EFOs are obtained independently for each user-defined fragment (typically the transition metal (TM) and its ligands). They are sorted by decreasing occupation number, and electrons (or electron pairs for closed-shell systems) are assigned to them until one reaches the total number of electrons. The fragment OS is then obtained by subtraction from the corresponding nuclear charges. This effective oxidation states (EOS) analysis, also provides a measure to quantify the extent to which the OS assignment is clear-cut, based on the difference in occupancy between the last occupied and first unoccupied EFOs. EOS analysis has been successfully applied to a wide range of systems [289]. The method notably deviates from the IUPAC approach [132, 133] because individual bonds are never explicitly considered. That permits EOS analysis to formally consider more than one Lewis structure at a time (i.e. treat multireference wavefunctions) on an equal footing.

Single-determinant wavefunctions are invariant to unitary transformations within the occupied molecular orbitals. While the canonical orbitals are typically delocalized (because they are appropriate for ionization), this invariance can be exploited to generate a set of localized orbitals (LO) based on some criterion [162]. This is directly possible within Kohn-Sham density functional theory (DFT), which is the dominant electronic structure approach [95]. The LO representation often produces orbitals that resemble the individual bonds in the dominant Lewis structure, and it is then natural to apply the ionic approximation to each LO individually, following the IUPAC definition more closely. However, since there is no unique way to define localization, there is a slew of different localization schemes to produce localized orbitals, namely Boys [290] Pipek-Mezey (PM) [143], Edminston-Ruedenberg (ER) [144], or more recent realizations based on Cholesky decomposition of the density matrix [158] and the fourth moment [160] or Knizia's intrinsic bond orbitals (IBOs) [122], to name a few.

This avenue has been explored by a number of methods [280, 281, 283]. Thom et al. [280] first coupled orbital localization with population analysis in the localized orbitals bonding analysis (LOBA) to assign the electrons associated with each LO. The LOBA method starts

with orbital localization by a chosen scheme, and then obtains the atomic populations from each localized orbital. Using either PM or ER localization together with Löwdin population analysis produced quite robust results [273, 280]. In the original paper, the OS assignment focused on the TM of the complex. A threshold of 60% in the atomic population was used to decide whether the electron pair is assigned to the TM or not. Recently, some of us [291] described an extension to the method, loosening the weight of the aforementioned threshold in the OS determination, to allow the possibility of covalent assignment (split between two atoms/fragments), and introducing a confidence measure for the assignment (either ionic or covalent) of each electron pair.

In that work [291] we observed that for some of the most challenging systems such as TM-carbenes, the LOBA method struggled to reach the accepted OS. Careful inspection of the localized orbitals indicated that the first step of the procedure, namely the orbital localization, was not always producing orbitals one could easily relate to a Lewis structure. The LOs often involve several atomic centers with appreciable contributions, which hinder the process of OS assignment. Moreover, using a different localization scheme could also lead to different OS assignments in some controversial cases. We concluded that a different orbital localization scheme, tailored for the purpose of OS assignment, was necessary to make progress in such cases.

A maximally robust procedure to assign OSs should rely on separating the localized orbitals into fragments, for which the degree of locality of core or valence orbitals *within* each fragment has no special relevance. Indeed there has been much development of specialized methods that aim to specifically localize orbitals onto fragments [258, 292–299] rather than maximizing a global measure of localization. Such methods have considerable value in energy decomposition analysis of intermolecular interactions [185, 300], as well as for fragment methods and embedding [258, 297, 301]. In our context there is a different need for fragment localization. For instance, if two fragments A and B, each formally bearing n_{A} and n_{B} electron pairs, are linked via a single bond with ionic character, then the OS should solely depend on a single localized orbital involving both A and B, leading either to A^{+}-B^{-} or A^{-}-B^{+}. On the other hand, since standard *system-optimal* orbital localization schemes do not make a distinction between the contact atoms of the A-B bond and the remaining atoms of A and B, a potentially better localization of the critical A-B bonding orbital may be sacrificed for better overall localization of all $n_{\mathrm{A}} + n_{\mathrm{B}}$ orbitals.

In light of the above considerations, there are several new components that are presented here to enable assignment of OSs. First, we put forward a robust fragment-based orbital localization scheme. For a given fragment, the resulting oxidation state localized orbitals (OSLOs) comprise a full set of orbitals spanning the occupied space that are ordered by spatial locality in the fragment. Second, to further characterize each orbital, given a set of fragment populations, we introduce a Fragment Orbital Localization Index (FOLI) which measures the population of each OSLO on a per-fragment basis. Third, to obtain the fragment populations, a more robust Hilbert-space based population analysis based on Knizia's intrinsic atomic orbitals (IAOs) [122] is also introduced. Fourth, we use the above components to develop an iterative algorithm to best select a *subset* of the OSLOs for each fragment

to span the full occupied space. The oxidation state of a given fragment is then determined by its number of assigned OSLOs relative to its total nuclear charge. Finally, with the new procedure in hand, we turn to exploration of a variety of interesting borderline cases, with focus on examples where LOBA was previously demonstrated to have some issues [291].

3.2 Methods

Oxidation States from Localized Orbitals (OSLO)

Starting with a single-determinant wavefunction built of n_{occ} spin orbitals, the fragment localization procedure is based on minimizing the radial spread functional from a given reference point for fragment F, $\mathbf{R}_F$. For a TM atom, $\mathbf{R}_F$ will be the atomic position; for a ligand, $\mathbf{R}_F$ will be its center of charge or charge. The minimization can be easily achieved in the molecular orbital (MO) basis by building a spread matrix, $\mathbf{L}^F$, with elements

$$L_{ij}^F = \int \psi_i(\mathbf{r})(\mathbf{r} - \mathbf{R}_F)^2 \psi_j(\mathbf{r}) d\mathbf{r}, \tag{3.1}$$

where ψ_i is the i^{th} occupied MO. Eq. 3.1 simplifies to

$$L_{ij}^F = \int \psi_i(\mathbf{r})\mathbf{r}^2\psi_j(\mathbf{r})d\mathbf{r} - 2\mathbf{R}_F \cdot \int \psi_i(\mathbf{r})\mathbf{r}\psi_j(\mathbf{r})d\mathbf{r} + \mathbf{R}_F^2\delta_{ij}, \tag{3.2}$$

where the first term contains the isotropic quadrupole moment matrix elements, the second term involves the dipole matrix elements, and the third term is a merely a constant diagonal offset. The required matrix elements are readily available in many quantum chemistry software packages. Diagonalization of the matrix $\mathbf{L}^F$

$$\mathbf{L}^F\mathbf{U}^F = \mathbf{U}^F\mathbf{\Lambda}^F \tag{3.3}$$

yields eigenvalues $\lambda_i^F = \Lambda_{ii}^F$ and a corresponding set of n_{occ} localized orbitals centered around $\mathbf{R}_F$

$$\phi_i^{loc,F}(\mathbf{r}) = \sum_k U_{ki}^F \psi_k(\mathbf{r}), \tag{3.4}$$

with their (squared) spreads given by the λ_i^X values. When $\mathbf{R}_F$ corresponds to an atomic position, the localized orbitals obtained by this procedure reproduce the shell structure of the atom, with core orbitals having the smaller spread values.

The target is to define the OS of M user-defined fragments of a molecular system, such as the metal(s) and ligands of a TM complex. We localize around each fragment's center of nuclear charge and get n_{occ} OSLOs for each of them (Mn_{occ} altogether). Note that in cases like polydentate or hapto ligands the ligand's centroid may be far from the ligand's nuclei, or even coincide with that of the metal (e.g. TM-porphyrin compounds). Minimizing the spread function is simple, non-iterative (no multiple minimum problem [240]), independent of any

population assignment, and appealing. However, the OSLOs most strongly associated with a fragment cannot always be chosen based on the smallest spread. For instance, in the case of a TM center with some coordination sphere, OSLOs dominated by ligand contributions can exhibit lower orbital spreads than the most diffuse TM orbitals (e.g. a 4s-type orbital for a 3d metal). Similarly, when $\mathbf{R}_F$ is the center of a ligand, some compact ligand-centered OSLOs have a significant contribution from the neighboring TM center.

We therefore need a complementary measure to identify those OSLOs that are most localized on a fragment. Using Pipek's delocalization measure, [241] defined in terms of *fragment populations*, $N_F^i = \sum_{A \in F} n_A^i$, rather than atomic populations, n_A^i, is a suitable starting point:

$$D_i = \{\sum_F (N_F^i)^2\}^{-1}, \tag{3.5}$$

When an orbital is localized on a single fragment, then $D_i = 1$. If the i^{th} orbital is perfectly delocalized across two fragments F and F' then $N_F^i = N_{F'}^i = 1/2$ and $D_i = 2$, and so on.

Out of the OSLOs generated from fragment F with low delocalization measure, we are interested in those that are also highly localized on fragment F. For this purpose, we introduce the fragment orbital localization index (FOLI)

$$D_i^F = \sqrt{\frac{D_i}{N_F^i}}, \tag{3.6}$$

The FOLI, D_i^F, is 1 when orbital i is perfectly localized on that fragment ($D_i = N_F^i = 1$). The FOLI $D_i^F \to 2$ when the orbital is perfectly delocalized over two fragments ($D_i = 2, N_F^i = N_{F'}^i = 0.5$). The FOLI gradually increases for OSLOs that are more delocalized and less centered on fragment F. Note that while a FOLI value of $D_i^F = 1$ means perfect fragment localization, higher FOLI values can result from different instances of delocalization. For example a FOLI value of 2 can also arise from 3 fragments with $D_i = 2.339$ via $N_F = 0.584$, $N_{F'} = N_{F''} = 0.208$.

How should one select the n most fragment-localized OSLOs from amongst the redundant set of $M \cdot n$ candidates? One could select the n_{occ} OSLOs with the smallest FOLI values and assign them to their parent fragments. However, we have observed that this procedure sometimes leads to linear dependencies among the selected OSLOs. We instead prefer the iterative scheme depicted in Fig. 3.1. On the first iteration, the best localized orbital (in the sense of smallest FOLI value) is selected and projected out from the occupied space for the next iteration. The 2nd iteration begins by constructing a new set of $M \cdot (n_{\text{occ}} - 1)$ localized orbitals, followed by selecting and removing the best localized orbital. Iterations continue until a total of n_{occ} optimal fragment-localized orbitals are selected. In the case of an unrestricted Slater determinant, the procedure is carried out for the α and β occupied spaces separately, and individual α and β OSLOs are thus produced and assigned to each fragment.

The basic algorithm is modified by introducing a tolerance (typically 10^{-3}) associated with the lowest FOLI values so that all OSLOs with D_i^F values within the tolerance are

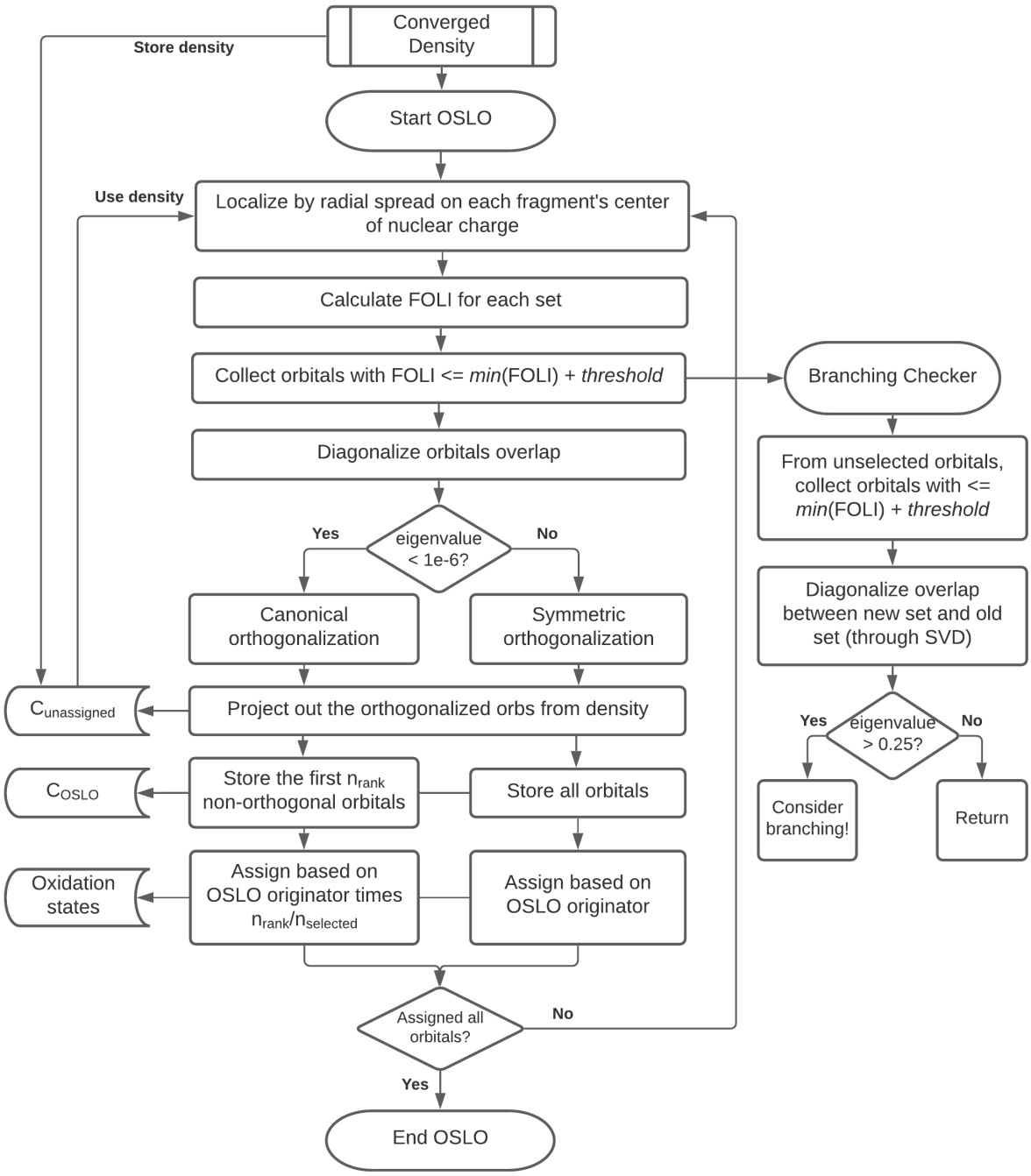


Figure 3.1: Flowchart of the iterative OSLO algorithm, where the most strongly fragment-localized orbitals (core and valence spectator orbitals) are projected out from the occupied space before the least fragment-localized orbitals that are most relevant to OS assignment are generated and inspected. This procedure has the desirable side-effect of improving the fragment-localization of orbitals that are not selected on later iterations.

selected in a given iteration. These orbitals are symmetrically orthogonalized and then projected out from the occupied space for the next iteration. This strategy avoids the problem that for symmetric systems, projecting out individual localized orbitals may result in a symmetry-broken density matrix for the next iteration.

There are a number of aspects of the procedure that are worth discussing in some detail. The localized orbitals obtained in the first few iterations are basically the atomic core orbitals of the fragment's atoms. As the iterative process advances, on-fragment localized valence orbitals are produced. They correspond to orbitals not particularly involved in the bonding between fragments, i.e. spectator orbitals. In the later iterations, the least-fragment localized valence orbitals are eventually selected. They correspond to bonds (or dative bonds) between fragments (e.g. TM-ligand orbitals). A nice side effect of the iterative procedure is that, by first removing the more fragment-localized orbitals from the occupied space, the relevant across-fragment orbitals are better localized on fragments (i.e. their FOLI values are *smaller* than those obtained in the first iteration using the whole occupied space). The final result thus depends to some extent on the order in which OSLOs are selected.

In borderline cases (where the FOLI-based selection is a close call) this may affect OS assignment. The algorithm allows the user to explore alternative outcomes in borderline cases by flagging when the OSLO selection procedure could *branch* into 2 (or more) paths. Consider a simple case with a single bond between fragments F and G. At some point in the iterative procedure, the corresponding bond localized orbital centered on F will be produced. At the same time, a similar bond localized orbital will be produced in the OSLOs associated with fragment G. The one with the smaller FOLI value is selected and projected out from the remaining occupied space. In the following iteration, this bond orbital will be absent from the new set of OSLOs obtained for both fragment F *and* G. If the $F-G$ bond is very non-polar (rare in TM complexes), the D_i^F and D_i^G values would be very similar, and one can argue that instead choosing the OSLO associated with G would produce a plausible alternative solution to selecting the one associated with F.

Our iterative algorithm automatically detects these (borderline) cases as follows. At each iteration, linear dependencies are checked between the OSLO that is selected and that with the *second* smallest FOLI value that is not selected (there may be multiple selected and non-selected localized orbitals if their FOLI values are within the tolerance). If near linear dependencies are indeed found *and* the difference in D_i^F values is small enough, our algorithm will print a diagnostic message. This allows the user to rerun the calculation, toggling a branching flag that selects the OSLO with the somewhat larger FOLI value, and proceed to obtain a second distinct solution.

To turn to OS assignment, we recall that each of the selected localized orbitals was generated from a fragment center with a low FOLI value. This makes it natural to assign the fragments' oxidation states based on the originator fragment. That is the procedure followed in the results reported here, and represents a "winner-takes-all" approach to the OS, similar to the IUPAC rules. However, once the optimal set of orthogonal fragment-localized orbitals are obtained, each orbital's allegiance may be reassigned according to the fragments' populations, either in a "winner-takes-all" fashion again, or alternatively by

allowing covalent assignments in non-polar cases, as described elsewhere [291].

An atomic partitioning scheme is necessary to evaluate the FOLI values. In this work, we use two very different partitioning approaches to demonstrate that different reasonable choices in fact work very similarly. First, we use the so-called Topological Fuzzy Voronoi Cells (TFVC) atomic definition [302], a real-space scheme that is used in the effective oxidation states (EOS) approach [282]. Second, we introduce a Hilbert-space based procedure based on Knizia's intrinsic atomic orbitals (IAOs) [122], where the reference minimal basis is obtained on-the-fly at the chosen level of theory. This IAO-AutoSAD procedure is described below.

The IAO-AutoSAD Reference Minimal Basis

Hilbert-space methods to assign atomic or fragment populations so that the results do not artificially depend on the underlying AO basis set often rely on using a minimal basis to *exactly* span the occupied space [120, 121, 201]. Amongst many such possibilities, Intrinsic Atomic Orbitals (IAOs) are perhaps the simplest, and have been shown to be robust for population analysis [122, 258, 303]. The basic idea of IAOs is to rely on a projection onto a reference minimal basis to facilitate atom-tagging.

Starting from a converged SCF solution, one projects the occupied MO coefficients, $\mathbf{C}_{\text{occ}}$, into the small reference minimal basis set and back to the big one as follows:

$$\tilde{\mathbf{C}}_{\text{occ}} = \text{ortho}(\mathbf{R}_{\text{ls}}\mathbf{R}_{\text{sl}}\mathbf{C}_{\text{occ}}), \tag{3.7}$$

$\mathbf{R}_{\text{sl}} = \mathbf{s}^{-1}\mathbf{S}_{\text{sl}}$ projects from the big basis into the small basis, given that $\mathbf{S}$ and $\mathbf{s}$ are the overlap matrices in the large and small basis sets, and $\mathbf{S}_{\text{sl}}$ is the matrix of overlaps between functions in the small and large basis sets. In the same notation, $\mathbf{R}_{\text{ls}} = \mathbf{S}^{-1}\mathbf{S}_{\text{ls}}$ projects from the small basis into the big basis. After symmetric orthogonalization to restore orthonormality, the so-called de-polarized orbitals are gathered in the matrix $\tilde{\mathbf{C}}_{\text{occ}}$.

The rectangular transformation matrix, $\mathbf{A}$, from the large basis, $\{\omega_\mu\}$, to the minimal IAO basis, $\chi_\alpha^{\text{IAO}} = \sum_\mu \omega_\mu A_{\mu\alpha}$, is produced by the following double projection step:

$$\mathbf{A}_{\text{ls}} = \text{ortho}(\mathbf{P}\mathbf{S}\tilde{\mathbf{P}}\mathbf{S}_{\text{ls}} + \mathbf{Q}\mathbf{S}\tilde{\mathbf{Q}})\mathbf{S}_{\text{ls}}) \tag{3.8}$$

$\mathbf{P} = \mathbf{C}_{\text{occ}}\mathbf{C}_{\text{occ}}^\dagger$ and $\tilde{\mathbf{P}} = \tilde{\mathbf{C}}_{\text{occ}}\tilde{\mathbf{C}}_{\text{occ}}^\dagger$ are the density matrices (i.e. occupied projectors) associated with the original occupied MOs and the de-polarized occupied MOs, respectively. Their orthogonal complements are $\mathbf{Q} = \mathbf{S}^{-1} - \mathbf{P}$ and $\tilde{\mathbf{Q}} = \mathbf{S}^{-1} - \tilde{\mathbf{P}}$. Once the orthogonal IAOs are available, the IAO atomic population of atom B is obtained as

$$N_B = \sum_{\beta\in B} \left(\mathbf{P}^{\text{IAO}}\right)_{\beta\beta} = \sum_{\beta\in B} \left(\mathbf{A}^\dagger\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{A}\right)_{\beta\beta}. \tag{3.9}$$

The reference minimal basis originally used [122] for construction of the IAOs is the co-called "MinAO" set, which is the standard cc-pVTZ AO basis manually truncated to a minimal basis. When the molecular calculation uses effective core potentials (ECPs),

"MinAO-PP" was employed, which is cc-pVTZ-PP truncated to a minimal basis (i.e. excluding core AOs). However, this reference minimal basis fails to be valid for ECPs of other sizes (i.e. larger or smaller core). These limitations that result from the MinAO reference minimal basis are particularly relevant when dealing with transition metals and heavier elements.

A universally applicable reference minimal basis is made on-the-fly from appropriate free atom density matrices obtained with the same functional and basis set as the molecular calculation. This is done using Q-Chem's [237] so-called AutoSAD functionality, which is normally employed to construct superposition of atomic density (SAD) initial guesses for DFT calculations at the target level of theory. Sphericalization is necessary to ensure proper shell structure in the reference minimal basis, since many atoms have partly occupied degenerate orbitals. For simplicity as well as to avoid ambiguity, we use the ground state of the neutral atom, although for some atoms, a case can be made for using different spin or charge states. For open shell atoms, the unrestricted SCF equations are solved, and the resulting α and β density matrices are spin-averaged as well as sphericalized.

The IAO-AutoSAD procedure first solves the following generalized eigenvalue problem separately for each free atom, A:

$$\mathbf{P}^A \mathbf{c}_\alpha^A = \mathbf{S}^A \mathbf{c}_\alpha^A \lambda_\alpha \tag{3.10}$$

Each matrix is defined in the full basis of the free atom (rank n_A), and $\mathbf{P}^A$ is the sphericalized and spin-averaged density matrix. The reference minimal basis set on atom A is defined by choosing m_A orbitals, corresponding to the fully and fractionally occupied atomic orbitals (with $\lambda_\alpha \geq \frac{1}{14}$, such that the f shell is selected even for a cerium atom with a single f electron). The set of selected column vectors $\{\mathbf{c}_\alpha^A\}$ defines an $n_A \times m_A$ transformation to the MBS, $\mathbf{C}_{\text{MBS}}^A$. The reference minimal basis has rank $M = \sum_A m_A$ with functions defined by the direct sum of the atomic transformations:

$$\mathbf{T} = \bigoplus_A \mathbf{C}_{\text{MBS}}^A \tag{3.11}$$

Given the $N \times M$ transformation from the AO basis to the reference minimal basis, $\mathbf{T}$, all quantities needed to evaluate the IAOs with this MBS are available. For instance, referring back to Eqs. 3.7 and 3.8, we see that $\mathbf{s} = \mathbf{T}^\dagger \mathbf{S} \mathbf{T}$ and $\mathbf{S}_{\text{sl}} = \mathbf{T}^\dagger \mathbf{S}$.

One must be aware that IAO orbitals and atomic charges do depend on the underlying choice of reference minimal basis. Fortunately, our IAO-AutoSAD procedure shows results that are generally very similar to using MinAO for problems where the latter can be applied (i.e. no pseudopotentials). Some examples are shown in Fig. 3.2, for a wide range of AO basis sets. This is encouraging, and sets the stage for the results we report for the OSLO procedure in the following section. However, IAO charges do exhibit some dependence on the reference minimal basis. For instance, the O atomic charge in H_2O changes by $0.25e^-$ when using STO-3G as the reference minimal basis instead of MinAO or AutoSAD. For IAO-AutoSAD applied to transition metals, the choice of atomic state also has an impact. These issues may deserve further study in the future, as IAOs become more widely used.

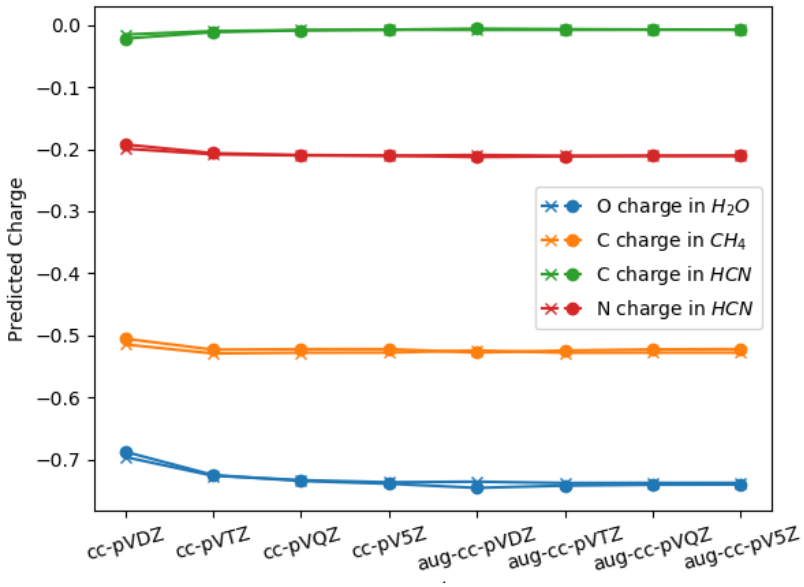


Figure 3.2: IAO atomic charges for H_2O, CH_4, and HCN where the crosses are charges from calculations using the AutoSAD reference minimal basis, while the circles used the MinAO basis [122]. All calculations are done with Hartree-Fock wavefunctions, and it is evident that the two sets of results are nearly indistinguishable.

Implementation and Computational Details

We have completed two independent implementations of the OSLO method, which serves as validation that both are correct, and also provides the opportunity to employ two different approaches to evaluate the fragment populations. Adopting the IAO-AutoSAD Hilbert space approach to charges described above, one of our implementations of the OSLO method is within the Q-Chem program package [237], and consequently uses OSLO fragment populations and FOLI values that are obtained analytically. Our second implementation uses the Topological Fuzzy Voronoi Cells (TFVC) real-space atomic definition [302], within the APOST-3D [304] package. The numerical evaluation of the TFVC charges used the atom-centered Becke multicenter quadrature scheme [305] with 40×146 grid points (per atom).

Geometry optimizations were performed by using the ωB97X-V density functional [97] with the def2-TZVP basis set (all electron for light atoms (through Kr) and with def2-ECP for heavier atoms) [247]. The ωB97X-V functional performs very well for both main group [95, 306] and transition metal compounds [307, 308]. Vibrational frequency calculations, to confirm minima on the potential energy surface, were computed at the same level of theory. Wavefunctions, energies and orbital localizations were also evaluated at the same level. All DFT calculations were performed with the Q-Chem package [237], while the OSLO analysis

was performed with both the Hilbert space and real-space implementations described above.

3.3 Results and Discussion

We evaluate the performance of the OSLO approach for a number of challenging systems, including high-valent oxides, TMs with non-innocent ligands, sulfur dioxide adducts with different bonding patterns, a Zn-based porphyrinic system and TM carbenes of different types (Schrock, Fischer and Grubbs 1st and 2nd generation). We apply both IAO-AutoSAD and TFVC atomic population schemes, to test the robustness of the OSLO procedure to the definition of fragment charges. The full set of results obtained are summarized in Table 3.1. The OS of the TM and the relevant ligand are reported, together with the FOLI values of the last selected OSLO, which is the least localized one (i.e. largest FOLI value among all selected OSLOs). In fact, once $n_{\mathrm{occ}} - 1$ localized orbitals have been projected out from the density matrix, there is only one (localized) orbital left. This last OSLO is associated with the fragment with the smallest FOLI value.

In many cases the last FOLI value is close to its smallest value of 1, indicating very good orbital localization and, consequently, a clear-cut OS assignment. In cases where the last FOLI value is larger than 1, it is very instructive to examine the Δ-FOLI value (the difference between the smallest and second smallest FOLI values) to see how clear-cut the OS assignment is: the larger the better. Cases with Δ-FOLI > 1 suggest clear-cut ionic character. Overall, the formal OS assignments using both fragment charge schemes agree in almost all cases (30 out of 33). The very few discordant cases have associated Δ-FOLI values below 0.2, which is probably smaller than can be meaningfully associated with application of the ionic approximation (this will be discussed more later).

We will discuss a few of the more clear-cut cases only briefly. For the high-valent oxides ranging from TiO_2 to IrO_4^+, the OS obtained with OSLO are in full agreement with LOBA, EOS and also with IUPAC's ionic approximation [291]. Clear-cut formal oxo (O^{2-}) ligands are obtained in all cases, resulting in OS as high as Ir(+9).

The PtO_4^{2+} system deserves particular attention. We treat the case of the spin-restricted solution, which is stable in orbital space, and leads to the optimized T_d geometry. There are lower energy spin-polarized solutions but we do not consider them here. The valence MO diagram can be found elsewhere [309]; there are eight π and four σ-type doubly-occupied MOs. In a T_d environment, the $5d$ orbitals of Pt split into E and T_2 symmetries, while the 6s orbital is A_1. The four symmetry-equivalent O atoms lead to σ- and π-type symmetry-adapted orbitals, with symmetries $\Gamma_\sigma = A_1 + T_2$ and $\Gamma_\pi = E + T_1 + T_2$. Towards the end of the iterative procedure, the E type OSLOs on Pt are very close in FOLI with the eight Γ_π of the O centers, due to the significant covalent character of the Pt-O bonds. When using TFVC, the FOLI value of O atoms (1.67) is smaller than that of Pt's E OSLOs (1.84), so they are selected and projected out of the P matrix for the next iteration. Since there is only one set of E orbitals in the occupied space, the aforementioned E-type OSLOs on Pt

Table 3.1: Summary of the OSLO results for the molecular systems studied, obtained with the IAO-AutoSAD and TFVC (in parenthesis) population analysis. OS for TM and selected ligand (L) in bold. tBu = tert-butyl, Cp = cyclopentadienyl, Ar = 2,6-diisopropylphenyl, Ar^1 = 2,6-dimethylphenyl, Cy = cyclohexyl, IMes = 1,3-Dimesitylimidazol-2-ylidene. $^{(a)}$ IAO-AutoSAD alternative solution. $^{(b)}$ TFVC results using tolerance value of 10^{-4}.

Complex	M OS	L OS	Δ-FOLI	Last FOLI
$[Ti\mathbf{O}_2]$	+4 (+4)	-2 (-2)	3.453 (3.367)	1.321 (1.372)
$[V\mathbf{O}_4]^{3-}$	+5 (+5)	-2 (-2)	1.548 (1.748)	1.466 (1.461)
$[Fe\mathbf{O}_4]^{2-}$	+6 (+6)	-2 (-2)	1.598 (1.682)	1.569 (1.623)
$[Re\mathbf{O}_4]^{-}$	+7 (+7)	-2 (-2)	1.806 (1.829)	1.470 (1.480)
$[Os\mathbf{O}_4]$	+8 (+8)	-2 (-2)	1.415 (1.363)	1.592 (1.609)
$[Ir\mathbf{O}_4]^{+}$	+9 (+9)	-2 (-2)	1.529 (1.084)	1.705 (1.742)
$[Pt\mathbf{O}_4]^{2+(a)}$	+10 (+10)	-2 (-2)	1.023 (0.707)	1.859 (1.904)
$Fe\mathbf{Cp}_2$	+2 (+2)	-1 (-1)	1.800 (2.343)	1.313 (1.437)
Zn(**porphyrin**)	+2 (+2)	-2 (-2)	0.958 (1.470)	1.509 (1.319)
$[Ni(\mathbf{S}_2\mathbf{C}_2\mathbf{Me}_2)_2]^0$	+2 (+2)	-1 (-1)	0.000 (0.000)	2.000 (2.245)
$[Ni(\mathbf{S}_2\mathbf{C}_2\mathbf{Me}_2)_2]^{1-}$ α	+2 (+3)	-1.5 (-2)	0.603 (0.913)	1.634 (1.656)
$[Ni(\mathbf{S}_2\mathbf{C}_2\mathbf{Me}_2)_2]^{1-}$ β			0.000 (0.110)	2.000 (1.901)
$[Ni(\mathbf{S}_2\mathbf{C}_2\mathbf{Me}_2)_2]^{2-}$	+2 (+2)	-2 (-2)	1.085 (1.374)	1.482 (1.509)
$[Cu(\mathbf{CF}_3)_4]^{1-}$	+3 (+3)	-1 (-1)	0.373 (0.728)	1.516 (1.531)
$[Cu(\mathbf{CF}_3)_4]^{2-}$ α	+2 (+2)	-1 (-1)	4.823 (4.845)	1.075 (1.152)
$[Cu(\mathbf{CF}_3)_4]^{2-}$ β		-1 (-1)	2.528 (2.867)	1.267 (1.270)
$[Cu(\mathbf{CF}_3)_4]^{3-}$	+1 (+1)	-1 (-1)	4.383 (4.581)	1.084 (1.145)
$Rh(\mathbf{SO}_2)Cl(PH_3)_2$ (L-type)	+1 (+1)	0 (0)	1.421 (1.209)	1.402 (1.509)
$Rh(\mathbf{SO}_2)Cl(CO)(PH_3)_2$ (Z-type)	+1 (+1)	0 (0)	1.064 (1.688)	1.606 (1.516)
$Ru(\mathbf{SO}_2)Cl(NO)(PH_3)_2$ (π-type) $^{(b)}$	0 (0)	0 (0)	0.514 (0.339)	2.432 (2.550)
$[Fe(CN)_5\mathbf{NO}]^{2-}$	+2 (+2)	+1 (+1)	0.981 (0.802)	1.573 (1.827)
$[Fe(CN)_5\mathbf{NO}]^{3-}$ α	+2 (+2)	0 (0)	0.839 (0.674)	1.688 (1.987)
$[Fe(CN)_5\mathbf{NO}]^{3-}$ β			2.638 (2.162)	1.375 (1.436)
$(CO)_5W{=}\mathbf{CHN(CH_3)_2}$ (1) (Fischer)	0 (0)	0 (0)	3.069 (2.205)	1.920 (3.196)
$(CO)_5W{=}\mathbf{CHOCH_3}$ (2) (Fischer)	0 (0)	0 (0)	1.860 (1.148)	2.037 (3.311)
$(CO)_5W{=}\mathbf{CF_2}$ (3) (Fischer)	0 (0)	0 (0)	1.567 (0.893)	2.017 (3.232)
$(CO)_5W{=}\mathbf{CH_2}$ (4) (Fischer)	0 (+2)	0 (-2)	0.612 (0.194)	2.279 (3.142)
$NAr(OtBu)_2W{=}\mathbf{CHtBu}$ (5) (Schrock)	+6 (+6)	-2 (-2)	0.283 (0.659)	1.924 (1.933)
$NAr(OtBu)_2W{=}\mathbf{CH_2}$ (6) (Schrock)	+6 (+6)	-2 (-2)	0.455 (0.806)	1.845 (1.908)
$NAr^1(OtBu)_2Mo{=}\mathbf{CHCMe_2Ph}$ (7) (Schrock)	+6 (+6)	-2 (-2)	0.253 (0.565)	1.916 (1.917)
$NAr^1(OtBu)_2Mo{=}\mathbf{CH_2}$ (8) (Schrock)	+6 (+6)	-2 (-2)	0.203 (0.487)	1.956 (2.018)
$NAr^1(OtBu)_2Mo{=}\mathbf{CHPh}$ (9) (Schrock)	+6 (+6)	-2 (-2)	0.153 (0.430)	1.986 (2.042)
$PCy_3Cl_2Os{=}\mathbf{CH_2}$ (10) (Grubbs)	+2 (+4)	0 (-2)	0.048 (0.044)	2.089 (2.278)
$H_2IMesCl_2Os{=}\mathbf{CH_2}$ (11) (Grubbs)	+4 (+4)	-2 (-2)	0.192 (0.259)	2.375 (2.709)
$PCy_3Cl_2Ru{=}\mathbf{CH_2}$ (12) (Grubbs)	+2 (+2)	0 (0)	0.222 (0.150)	1.964 (2.199)
$H_2IMesCl_2Ru{=}\mathbf{CH_2}$ (13) (Grubbs)	+2 (+2)	0 (0)	0.276 (0.306)	2.013 (2.149)
$(PH_3)_2Cl_2Ru{=}\mathbf{CH_2}$ (14) (Grubbs)	+2 (+2)	0 (0)	0.241 (0.090)	1.961 (2.153)

are now absent, and the OSLOs with smaller FOLI value become those corresponding to the Γ_σ of the O centers, leading to a fairly clear Pt(+10) assignment (see Table 3.1).

When using IAO-AutoSAD, however, at the same step of the process the situation is reversed. The FOLI value of the Pt E OSLOs is smaller (1.50) than that of the eight Γ_π (1.69). When the former are selected and projected out from the P-matrix, the eight π-type OSLOs of the O centers become rank deficient for the next iterations. This leads to two undesirable outcomes. First, the localized orbitals on the O are mixed-up by the canonical orthogonalization process. Second, a last OSLO of A_1 symmetry delocalized over the four O centers remains left in the last iteration, leading to a huge FOLI value (4.63) and a split of the electron pair among the four equivalent O centers. By applying the branching option on the IAO-AutoSAD calculation, the Pt(+10) picture obtained with TFVC is recovered, with similar Δ-FOLI value. Selecting OSLOs with a higher FOLI value at a given step of the iterative process ends up providing a final solution where the *sum* of the FOLI values of the selected OSLOs is smaller (36.34 vs 33.22).

The Zn-porphyrin system is potentially challenging for OSLO because, due to its symmetry, the center of charge of the porphyrin ligand *exactly* coincides with the position of the Zn nucleus. Nevertheless, as a result of using the FOLI values, the OSLO procedure performs smoothly, yielding the expected Zn (+2) OS with Δ-FOLI value close to or even larger (TFVC) than 1. The OSLO results for the nitroprusside anion ($[Fe(CN)_5NO]^{2-}$) and its reduced form are also very clear, leading to a formal Fe(+2) species and a ligand-based reduction, in line with the well-known non-innocent nature of the nitrosyl ligand. The IAO-AutoSAD population yields larger Δ-FOLI values as compared to TFVC.

Ferrocene (see Fig. 3.3) is a nice example because it shows the utility of the OSLOs themselves. The cyclopentadienyl anion OSLOs are shown in panels (a) through (c), and the fragment localized π orbitals are particularly pleasing because they resemble the delocalized π orbitals of the isolated anion. In other words, this shows the advantage of *fragment* localization over *global* localization (see also the recent treatment via intrinsic fragment orbitals [258]). The 3 occupied Fe (3d) orbitals emerge as expected, and the OS assignment is very clear based on the small FOLI value of the last orbital as well as the large Δ-FOLI gap.

The redox series of nickel diothiolate complexes, $[Ni(S_2C_2Me_2)_2]^{n-}$ with $n = 0, 1, 2$, is particularly interesting. The $n = 0$ complex is a closed shell singlet. Fig. 3.4 gathers the most relevant OSLOs and the corresponding FOLI values. For Ni, four well-localized d-type orbitals (Fig. 3.4a) are obtained, leading to a Ni(+2) OS. Then, for each thiolate ligand, one finds two S lone pairs and two σ-type orbitals associated with the two S-Ni σ bonds (see Fig. 3.4b). Since each thiolate is a fragment, these two sets of orbitals are not localized into individual S lone pair and S-Ni bonds, but form two in-phase (+,+) and out-of-phase (+,-) localized orbitals within the fragment. The $\sigma(+,-)$ OSLO exhibits a relatively large FOLI value (~ 1.7), indicating some partial contribution from the Ni center. The $\sigma(+,+)$ orbital shows a minor Ni contribution, leading to a smaller FOLI value (~ 1.3). In addition, each thiolate exhibits a well localized π-type orbital on the two sp^2 carbon atoms, with FOLI ~ 1. The last OSLO, in Fig. 3.4b (bottom right), corresponds to a π-type orbital delocalized over

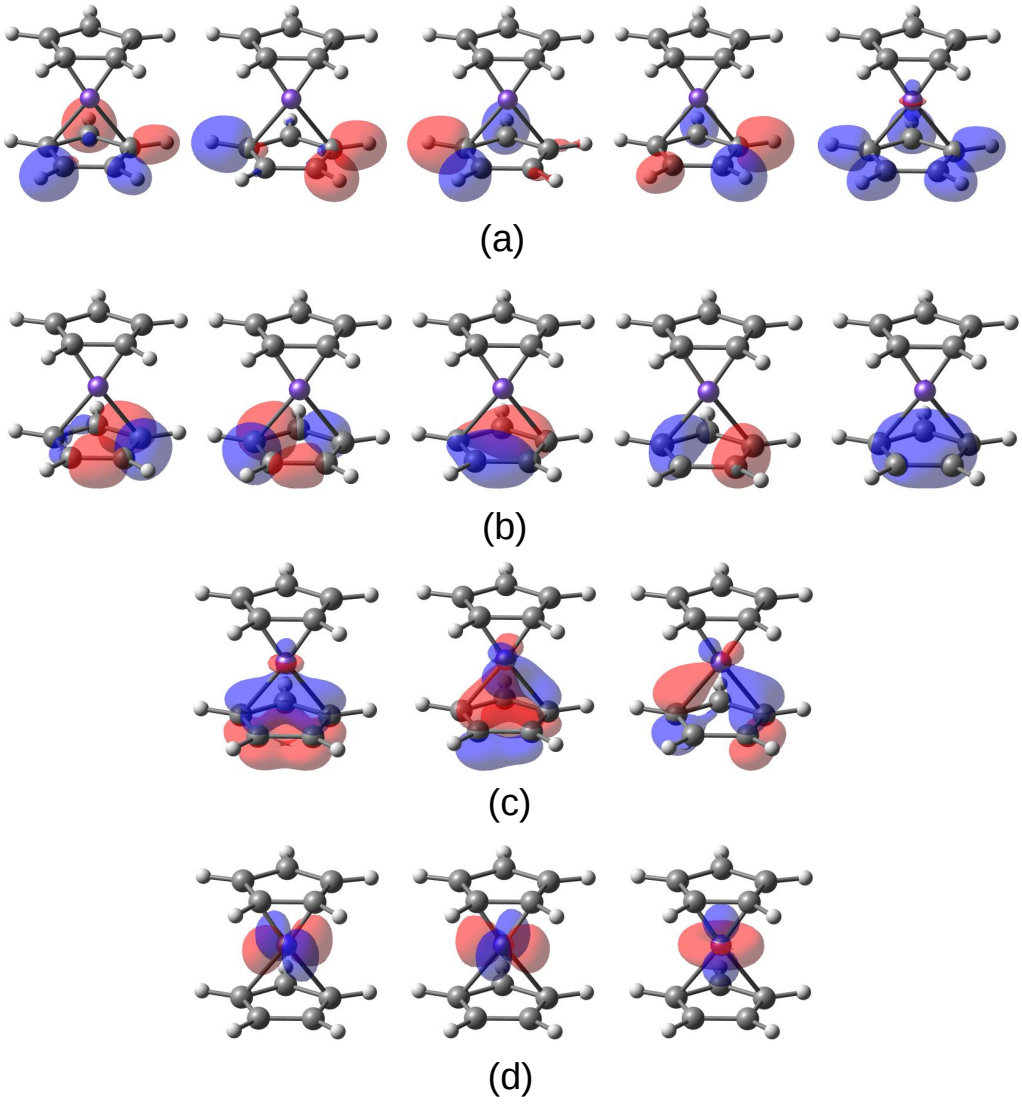


Figure 3.3: Valence OSLOs for the $FeCp_2$ complex as produced by the algorithm shown in Fig. 3.1. The lower cyclopentadienyl ligand's σ_{C-H} OSLOs are shown in panel (a), its σ_{C-C} OSLOs are shown in panel (b), and its π OSLOs in panel (c). The 3 d-type OSLOs on Fe are shown in panel (d). The isosurface value is 0.075 a.u.

the two thiolate ligands, consistent with the FOLI value (~ 2). Moreover, the Δ-FOLI value is *exactly* zero for both population schemes. This indicates a formal split of the electron pair between the two ligands (in other words, a covalent assignment), leading to two thiolate (-1) moieties to accompany the Ni (+2) center. We can envisage similar situations with the OSLO procedure when dealing with mixed-valence compounds. Visual inspection of the critical localized orbital/s will confirm or deny mixed valence or covalent character suggested by very small Δ-FOLI values.

The two-electron reduction of $[Ni(S_2C_2Me_2)_2]$ leads to the S = 0 closed-shell species $[Ni(S_2C_2Me_2)_2]^{2-}$. The OSLO procedure yields essentially the same valence localized orbitals as in the previously discussed oxidized form (i.e. four d-type localized orbitals on Ni, two lone pairs and two σ type S-Ni orbitals), except that the last delocalized orbital is replaced by two well-localized π-type orbitals, one on each thiolate ligand, as shown in Fig. 3.4c. The Δ-FOLI value is larger than 1, clearly pointing to ligand-based reduction, and in turn, Ni (+2) and two thiolate (-2) moieties.

One-electron reduction to $[Ni(S_2C_2Me_2)_2]^{1-}$ is more tricky. The system is an open-shell doublet ($S = \frac{1}{2}$), and the α and β parts are treated separately. The α part is rather clear cut, yielding similar localized orbitals as in the fully reduced $n = 2$ species, with Δ-FOLI > 0.6. For the β part, using IAO-AutoSAD populations leads to localized orbitals comparable to those of the oxidized form ($n = 0$): four d orbitals on Ni and a last π orbital delocalized over the two thiolate moieties, with FOLI ~ 2 and Δ-FOLI= 0. These OLSOs suggest a mixed-valence situation with Ni(+2) and two partially-reduced thiolates (-1.5), from equal sharing of the last beta electron between the thiolates. This result, as well as those for the closed-shell species, agrees with the experimental evidence [310, 311] and with the EOS scheme [291].

However, the results for the β part of the $S = \frac{1}{2}$ species are somewhat different when using TFVC populations. In the iterative process, two equivalent π-type orbitals centered on each thiolate with significant contribution from the Ni center (see Fig. 3.4d) are selected over the d-type orbital centered on Ni (which also exhibits significant mixing with the ligands). The FOLI values are 1.901 and 2.011, respectively, indicating a large degree of delocalization of these last orbitals. As a consequence, the picture obtained is a Ni(+3) with two fully reduced thiolate (-2) ligands, with a Δ-FOLI value of merely 0.11. This very small Δ-FOLI value argues for equal sharing of the last electron pair.

In Naumann's ion [312], $[Cu(CF_3)_4]^-$, Cu OS and role of the CF_3 ligands has been debated for more than 25 years [313–318]. Based on the DFT frontier molecular orbitals, Snyder considered that the metal center is best described as Cu (+1) (i.e. d^{10}), instead of a d^8 Cu (+3) as would follow if all CF_3 ligands were formally anionic (-1) [313]. According to Snyder, the anion features an "inverted" ligand field [318], where the lowest unoccupied molecular orbital (LUMO) exhibits dominant ligand character. This assignment has been questioned by others, who support the latter, more conventional, view of four CF_3^- ligands [314, 315]. Most recently, the Lancaster group have put forward further experimental and computational arguments in favor of d^{10} [316, 317]. Other arguments supporting the Cu (+1) and Cu (+3) picture have also been given [318].

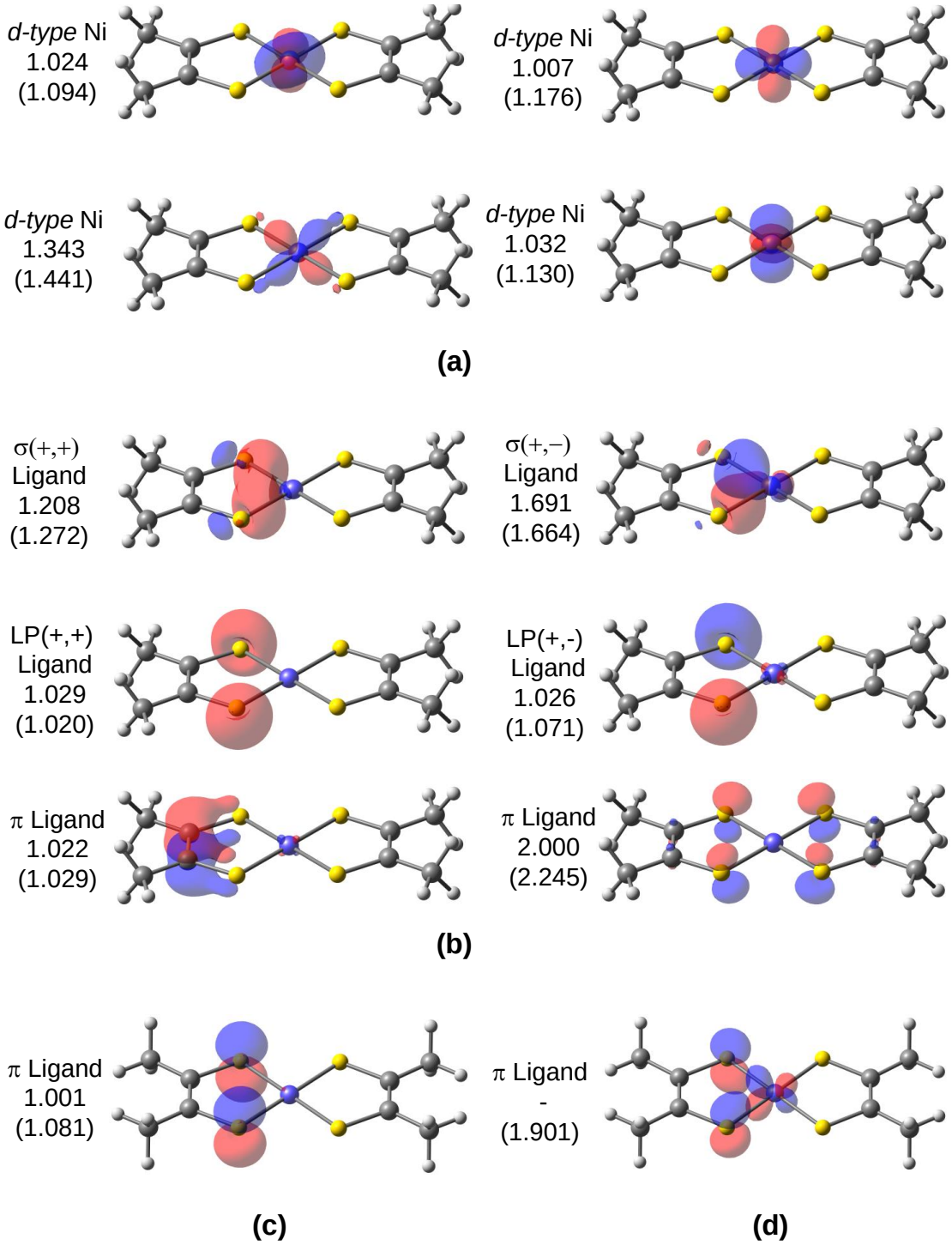


Figure 3.4: Valence LOs of the $[Ni(S_2C_2Me_2)_2]^0$ system with IAO-AutoSAD and TFVC (in parenthesis) FOLI values. *d*-type orbitals on Ni (a), ligand's σ, lone pair (LP) and π obitals (b). Last localized ligand π orbital for $[Ni(S_2C_2Me_2)_2]^{2-}$ (c). Last localized ligand π orbital from the beta density of $[Ni(S_2C_2Me_2)_2]^-$ using TFVC (d). The isosurface value is 0.075 a.u.

These conflicting views are rooted in the relatively non-polar character of the Cu-C bond. In recent work, some of us showed that both the LOBA and EOS schemes give OS assignments consistent with a formal Cu (+3) species [291]. While the LOBA results were rather clear, the R (%) = 51.7 value of the EOS assignment indicated a very close call. The OSLO procedure (with either population analysis) points towards formal anionic CF_3 (-1) ligands and hence a Cu (+3) species, as shown on Table 3.1. One can identify four well-localized d-type orbitals centered on Cu (Fig. 3.5a), while the σ-type interaction between Cu and each CF_3 is captured by four equivalent ligand-centered orbitals with non-negligible contribution from the Cu (Fig. 3.5b). Notice also the mixing of p-type orbitals from the F atoms of the CF_3 moiety. These localized orbitals provide a much clearer picture as compared to those obtained with PM localization for the same wavefunction [291]. Encouragingly. the FOLI values of the last orbitals are virtually the same with both atomic population schemes ($\sim$ 1.5). The Δ-FOLI value for the assignment is somewhat smaller using IAO-AutoSAD (0.373) as compared to TFVC (0.728), but the same picture emerges in both cases. Finally, OSLO results for the one- and two-electron reduction of $[Cu(CF_3)_4]^-$ indicate that both processes are metal based, as expected given the Cu (+3) assignment of the anion.

In TM carbenes a double bond is formed between the TM and the carbene. The σ bond is understood as originating from σ-donation of a sp^2 lone pair on the carbon atom to the TM. The nature of the π-type interaction is much more system-dependent, leading to two well-established situations. In the so-called Fischer carbenes, the π electrons formally sit on the TM d-type orbital, which back-donates to a formally neutral carbene moiety. In Schrock-type carbenes, the π electrons are formally associated with the carbene moiety, which becomes anionic (-2). Previous experience indicated that OS assignment in TM-carbene complexes is challenging. Often, EOS analysis yields low R(%) values rather close to 50, driven by nearly equal populations of the π-type EFOs on the TM and the carbene moiety. With LOBA, Pipek-Mezey localized orbitals do not readily correspond with the σ and π bonds [291]. We studied a set of fourteen TM carbenes [319]. The set includes four conventional W-based Fischer carbenes (1–4), five Schrock W- and Mo-based catalysts (5–9) and five Ru- and Os-based 1st and 2nd generation Grubbs catalysts (10–14).

Referring again to Table 3.1, the OSLO procedure combined with IAO-AutoSAD populations correctly identifies all prototypical Fischer and Schrock carbenes, while all Grubbs catalysts but 10 are pictured as formal neutral Fisher-type carbenes. Notice the Δ-FOLI values are mostly below 0.3 (especially for nominal Schrock and Grubbs carbenes), with FOLI values around 2.0, indicating significant delocalization of the last orbital. Such values could support a covalent division of charge, and, at the very least, call for inspection of the relevant σ-type and π-type TM-carbene OSLOs. To this end, Fig. 3.6 shows two examples of OSLOs involving the TM and the carbene unit. The FOLI values of the σ-type bonds (left) are noticeably smaller than those of the π-type bonds (right), which exhibit a very similar contribution from both fragments. Yet, one can see that the OSLO procedure produces nice, chemically interpretable localized orbitals for the σ and π bonding.

For the two cases shown in Fig. 3.6, the IAO-AutoSAD and TFVC fragment charges lead to different OS assignments, despite yielding almost identical sets of localized orbitals.

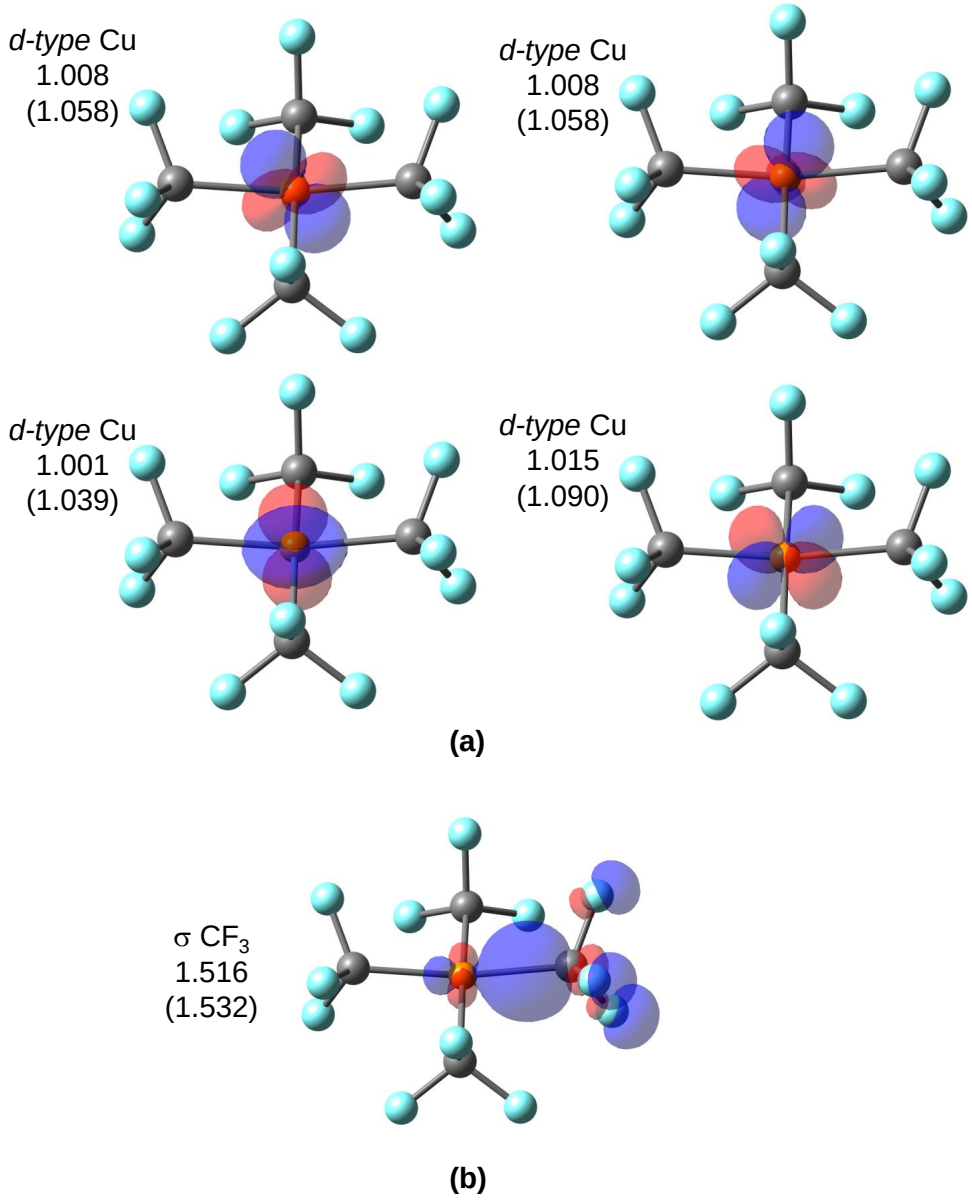


Figure 3.5: Selected LOs of $[Cu(CF_3)_4]^-$ with IAO-AutoSAD and TFVC (in parenthesis) FOLI values for Cu (a) and the CF_3 ligand (b). The relatively non-polar character of the σ Cu-CF_3 interaction is clearly evident. However the FOLI value (~ 1.5) as well as visual inspection indicates that this orbital has greater $CF_3{}^-$ character than Cu (3d) character so that in a winner-take-all assignment, the ligands emerge as $CF_3{}^-$ and the metal adopts a Cu (+3) OS. The isosurface value for the plots is 0.075 a.u.

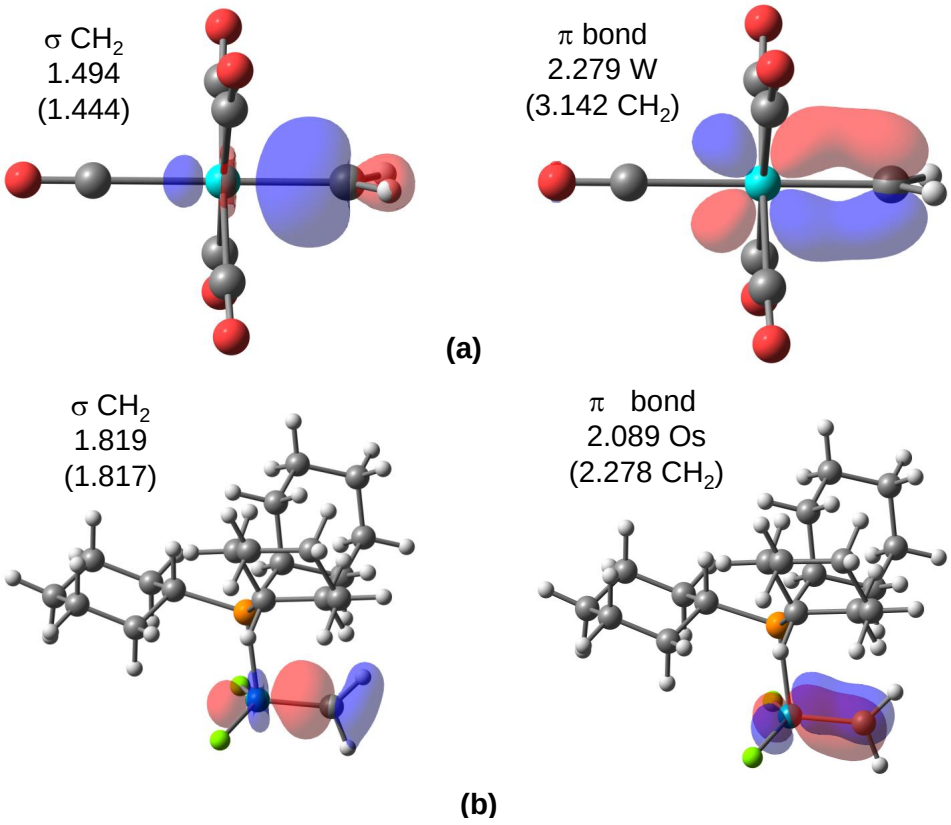


Figure 3.6: σ and π TM-carbene OSLOs for (a) the Fischer-type $(CO)_5W{=}\mathbf{CH}_2$ complex (species 4), and (b) the Grubbs-type $PCy_3Cl_2Os{=}\mathbf{CH}_2$ complex (species 10). The FOLI values for each orbital are shown using IAO-AutoSAD (with the corresponding TFVC values in parentheses). The large FOLI values (> 2) for the π-type TM-carbene interaction, as well as visual inspection show the shared electron character of this interaction. The isosurface value is 0.075.

The σ-type OSLO belongs to the carbene, with FOLI values of $\sim$ 1.5 and $\sim$ 1.8 for the Fischer-type $(CO)_5W{=}\mathbf{CH}_2$ complex (4), and the Grubbs-type $PCy_3Cl_2Os{=}\mathbf{CH}_2$ complex (10), respectively. The π-type OSLO is the origin of the discrepancy. Since this is the last selected orbital in the OSLO procedure, its allegiance is based on the FOLI values for each fragment. In the case of 10, TFVC provides FOLI values of 2.28 for the carbene and 2.32 for the TM, while for IAO-AutoSAD the values are 2.14 and 2.09, respectively. Consequently, ionic assignment leads to a neutral CH_2 according to IAO-AutoSAD and to an anionic CH_2 (-2) according to TFVC. Notice that Δ-FOLI values are below 0.05 in both cases, the smallest seen in this study. The genuinely covalent nature of this π bond precludes meaningful classification of this system as Fischer or Schrock: instead the electron pair is shared.

On the other hand, the different OS assignment for 4 is rather unexpected, as both population methods produce virtually the same set of OSLOs. With IAO-AutoSAD, (4) is quite clearly a neutral Fischer-type carbene with a Δ-FOLI value of 0.61. However, when using TFVC, the assignment is not only reversed, with a small Δ-FOLI (0.19), but also the FOLI values of the last π-type OSLO are significantly higher (3.14 and 3.34 for the carbene and TM metal, respectively). The fragment TFVC populations on the carbene and the W atom are 0.76 e^- and 0.66 e^-, respectively. Hence, the remaining 0.58 e^- belongs to the spectator CO ligands, which explains the large FOLI value obtained. By contrast, with IAO-AutoSAD, the population of the carbene and W are 0.64 and 1.03 e^-, so the electron pair is more clearly on W (although covalent character is visually evident in Fig. 3.6).

As a last example, let us consider the species described in IUPAC's technical report illustrating three different bonding modes of the SO_2 ligand [132, 133]. Karen showed that when acting as a Z-type ligand (i.e. as a Lewis acid), the electronegative-acceptor caveat had to be applied to the ionic approximation so that the SO_2 ligand remains neutral. With EOS analysis the expected neutral SO_2 ligand was recovered in all three cases [289]. The OSLO results of Table 3.1 also clearly identify a neutral SO_2 moiety for both the L-type and Z-type configurations, with large Δ-FOLI values of over 1.0. In the case of the π-type bonding configuration, the SO_2 is once again clearly identified as neutral, but there is a close-call situation involving Ru and a non-innocent nitrosyl ligand trans to the SO_2, that calls for visual inspection of the OSLOs. The last three OSLOs (from IAO-AutoSAD results) belong to Ru, with FOLI values of 1.44, 1.88 and 2.43, and are depicted in Fig. 3.7. The admixture of contributions from the SO_2 and NO ligands is indeed significant for the last OLSO (Fig. 3.7b), but the Δ-FOLI value of 0.51 suggests that an ionic assignment to Ru remains justified. Overall, this results in a Ru (0), SO_2 (0) and NO (+1) OS assignment.

When using TFVC populations with the default tolerance of 10^{-2}, the last two localized orbitals, one centered on Ru and another on NO, have FOLI values within the tolerance and are therefore selected together in the last step of the iterative procedure, leading to a different OS assignment. However, the linear-dependency check indicates significant overlap between these two orbitals. As a consequence, their shape substantially changes after orthogonalization, leading to a pair of localized orbitals very similar to those of Fig. 3.7. In this case, being a non-symmetric system, a tighter tolerance of 10^{-4} affords the selection of the last orbitals in the iterative process one by one, readily producing the same results obtained with IAO-AutoSAD.

3.4 Conclusions

The purpose of this work was to report a new oxidation state localized orbital (OSLO) scheme that performs orbital localization based on *molecular fragments*, after a DFT calculation with chosen total charge and spin state. The user should select a fragmentation of the target complex, such as separation into one (or more) metal centers and individual ligands. After an iterative process to select the most strongly fragment-localized OSLOs, each molecular

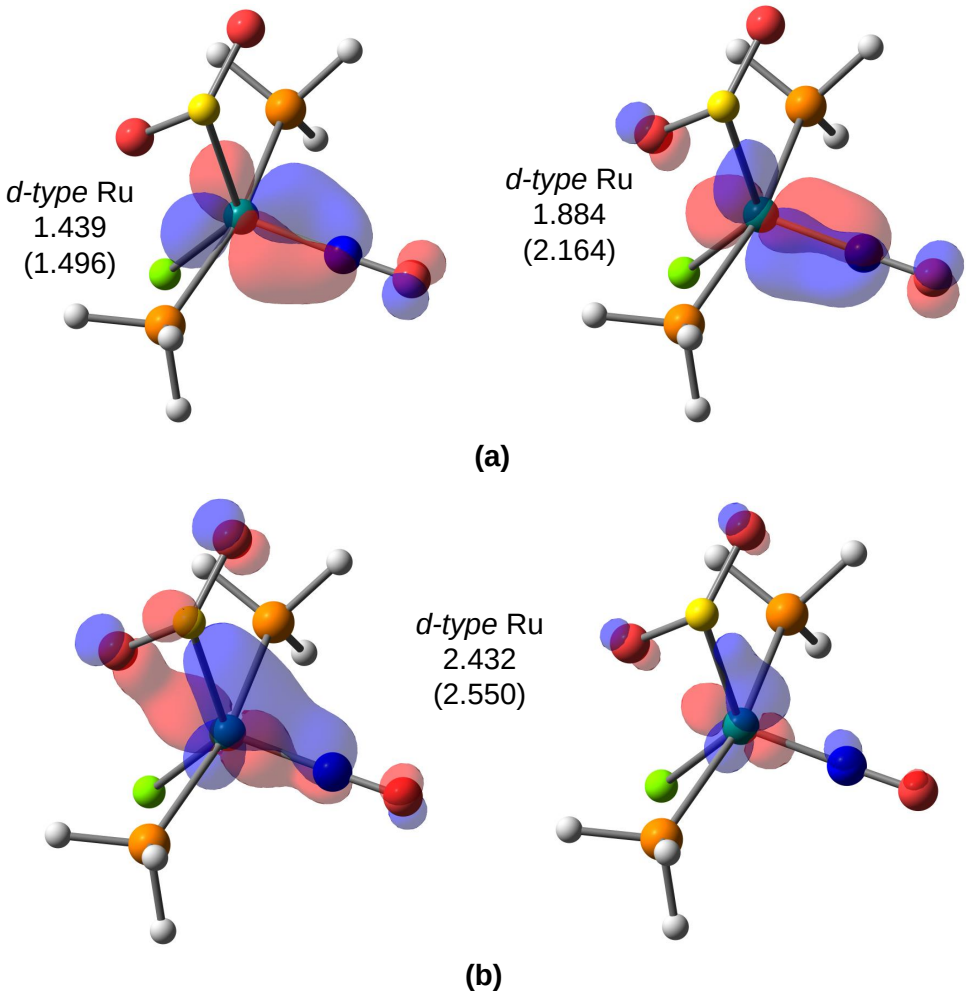


Figure 3.7: Selected Ru-centered OSLOs for $Ru(SO_2)Cl(NO)(PH_3)_2$ (π-type), with IAO-AutoSAD and TFVC (in parenthesis) FOLI values. The isosurface is 0.075 a.u. for panel (a) which shows the 3rd and 2nd last OSLOs to be selected. In panel (b), which shows the last OSLO selected, there are also significant ligand contributions as evident from the larger FOLI value, and visual inspection of the orbital isosurface of 0.075 a.u. (left), which can be clarified by choosing a larger value of 0.125 a.u. for the isosurface (right).

fragment is associated with a set of localized orbitals derived from a simple orbital spread criterion. This association, in turn, determines the fragment's formal charge or oxidation state (OS) in a natural manner.

We introduced a new index, namely the fragment orbital localization index (FOLI), to quantify the degree of locality of each OSLO (or any input orbital) on each fragment.

As examples, the lowest possible FOLI value of 1 corresponds to complete localization on that fragment, while perfect delocalization between 2 fragments yields a FOLI value of 2. Evaluation of FOLIs requires fragment populations. Two distinct population schemes have been tested for this purpose, namely a real-space approach (TFVC) and a new version of the intrinsic atomic orbitals (IAO-AutoSAD) that uses on-the-fly evaluation of the reference minimal basis based on superposition of atomic densities (SAD).

The OSLO iterative procedure selects the orbital with lowest FOLI value on each iteration, so that the last OSLO produced has the largest FOLI value (and is least strongly fragment-localized) among the whole set. The Δ-FOLI value for the last localized orbital measures the gap with the second smallest FOLI value among the fragments. Δ-FOLI measures the reliability of the OS assignment, such that a Δ-FOLI value larger than 0.5 usually indicates a clear OS assignment. Smaller values suggest increasingly covalent character in the least localized OSLO.

Numerical tests of the new scheme shows that the OSLOs are in much better agreement with the expected Lewis structure than those obtained with other *global* localization schemes such as Pipek-Mezey (apart from straightforward cases). As a result, previously identified limitations of the localized orbital bonding analysis (LOBA) procedure for OS assignment that originate in the use of global orbital localization methods are overcome with the OSLO approach. Transition metal carbenes are one such class of examples.

The OSLOs themselves carry significant chemical information, and their visualization helps to clarify borderline OS assignments. One such example discussed here is the $Cu(CF_3)_4^-$ anion, where the OSLO corresponding to the $\sigma(Cu-CF_3)$ interaction exhibits some covalent character, but supports a conventional d^8 Cu configuration rather than d^{10}. Another example is the Grubbs catalyst, $PCy_3Cl_2Os{=}CH_2$, where the OSLO corresponding to the Os-carbene π bond is almost perfectly covalent, thus rendering the conventional Fischer and Schrock classifications inapplicable.

We find the IAO-AutoSAD population scheme performs well in combination with OSLO, outperforming the TFVC scheme that is conventionally used in the framework of effective oxidation state (EOS) analysis. IAO-AutoSAD represents a promising all-round general, fast, analytical, basis-set independent Hilbert-space based atomic population scheme.

Chapter 4

Force Decomposition Analysis

4.1 Introduction

Intermolecular interactions are important for understanding chemistry as they affect structures, properties, and reactivity of chemical systems. Examples include red- or blue-shifts in vibrational frequencies when forming hydrogen bonds [320–326], wavelength tuning of organic chromophores by the solvation or protein environment [327–333], and modulation of the catalytic performance of molecular CO2RR catalysts through interactions with ligands in complexes' second coordination sphere [334–341].

Decomposing the non-covalent interactions has been increasingly important to understanding the origins of these interactions as well as the development of classical force fields for the simulation of chemical and biochemical systems [163, 342–349]. Moreover, to obtain statistical mechanical ensembles of a condensed phase chemical system, molecular dynamics simulations are required, for which accurate and efficient evaluation of intermolecular forces is important for systems where quantum chemical calculations are impractical.

Many energy decomposition analysis (EDA) methods have been proposed for separating different physical contributions to the non-covalent interaction energy, which are reviewed elsewhere [170, 174, 185, 300, 350–352]. In this work, we are using the absolutely localized molecular orbitals EDA (ALMO-EDA) [183–185, 353, 354], which divides the interaction energy into frozen (interaction between unrelaxed monomers), polarization (energy lowering due to intra-fragment relaxation of monomer wavefunctions), and charge transfer (energy lowering due to electron delocalization between fragments) contributions. The adiabatic EDA [355] optimizes the geometry successively on each of the intermediate potential energy surfaces (frozen, polarized, and fully relaxed), in essence attributing geometric changes and shifts in other molecular properties upon the formation of intermolecular complexes to the same terms as in ALMO-EDA. The adiabatic EDA has been successfully used to understand geometric changes arising from intermolecular interactions in a wide variety of systems [325, 355–360]. For example, the ∠N-B-H angle in the ammonia-borane complex only bends when charge is allowed to flow from the ammonia to the borane molecule [355, 360]. The adiabatic

EDA thus attributes changes in *observables* to the different EDA contributions, which can be crucial for connecting to experimental results.

Many fixed-charge and polarizable force fields have been developed over the years for condensed-phase molecular simulations [345–347, 349, 361–367]. Recently, the T. Head-Gordon lab developed the MB-UCB many-body force field for water-water and water-ion interactions [347, 349], which employs terms that resemble those produced by ALMO-EDA of quantum mechanical calculations. For example, the polarization energy in the second-generation ALMO-EDA allows electrons to move to the space of dipolar and quadrupolar density response to an external electric field, while MB-UCB uses distributed multipole analysis of classical anisotropic dipolar polarization to evaluate the polarization energy. Impressively, the two terms co-validate each other for a wide variety of water dimer geometries [347]. Similarly, the other terms are also related to each other as will be described below. While the total interaction energy and its breakdown given by MB-UCB and ALMO-EDA are in very good agreement, the forces remain to be compared to ensure the quality of dynamics driven by MB-UCB within a large ensemble of configurations.

In this work, we decompose the forces of an intermolecular interaction into constituent terms that directly correspond to those within the ALMO-EDA. After presenting the relevant theory, we demonstrate the usefulness of this decomposition for understanding chemistry with proof-of-concept examples of water interacting with sodium and chloride ions as well as the water dimer. We transform these forces to the internal coordinates using Wilson's B-matrix [368], allowing us to see forces that are more intuitive and relatable to vibrational spectroscopy. For example, the H-O-H bending in the water molecule is used to understand the molecular environment [369]. We then study CO_2 adsorption and activation on Au and Ag anions and compare the two systems' forces within internal coordinates. The force decomposition is also applied to validate the forces produced by the MB-UCB force field. The force decomposition results, based on high-quality DFT calculations of the forces, may also be useful for future force-field training. This work builds on the adiabatic EDA to advance the idea of "property decomposition" analysis, in which not only the interaction energy is broken down, but also other derivatives of the energy, which are molecular properties. This general approach can be extended to the effect of intermolecular interactions on other properties of interest, such as the hessian matrix, NMR chemical shieldings, dipole moments or polarizabilities, etc. The observable changes in properties associated with non-covalent interactions can then be attributed to the different physical effects at play.

4.2 Theory

Energy Decomposition Analysis

In the ALMO-EDA, the total binding energy of an intermolecular complex is broken down into four components coming from the successive removal of constraints to minimize the

energy of the supersystem:

$$\Delta E_{\text{bind}} = \Delta E_{\text{gd}} + \Delta E_{\text{frz}} + \Delta E_{\text{pol}} + \Delta E_{\text{ct}} \tag{4.1}$$

The geometry distortion term (ΔE_{gd}) refers to the energy consumed for each fragment to change its geometry from the equilibrium structure in isolation to that in the complex. The frozen (FRZ) interaction energy, ΔE_{frz}, is defined by the energy of the frozen wavefunction [370], relative to that of the isolated non-interacting fragments (in distorted geometries). It corresponds to the energy change upon moving the isolated fragments into their positions in the complex while keeping their own electronic structure unchanged. The frozen wavefunction is the antisymmetric product of the isolated fragment wavefunctions, whose associated one-particle density matrix (1PDM), $\mathbf{P}_{\text{frz}}$, is given by

$$\mathbf{P}_{\text{frz}} = (\mathbf{C}_o)_{\text{frz}} \boldsymbol{\sigma}_{\text{frz}}^{-1} (\mathbf{C}_o)_{\text{frz}}^T, \tag{4.2}$$

where $(\mathbf{C}_o)_{\text{frz}}$ is the direct sum of the (occupied) AO-to-MO coefficient matrices of the isolated fragments and $\boldsymbol{\sigma}_{\text{frz}}$ is the overlap matrix of the orbitals coming from the $(\mathbf{C}_o)_{\text{frz}}$ matrix. This 1PDM definition gives us the frozen interaction energy definition:

$$\Delta E_{\text{frz}} = E[\mathbf{P}_{\text{frz}}] - \sum_A E[\mathbf{P}_A]. \tag{4.3}$$

The next contribution, the polarization energy (ΔE_{pol}), arises from allowing the occupied orbitals on each fragment to mix with the virtuals *only* on the same fragment. Minimizing the energy subject to this constraint, also known as the SCF-MI procedure [292, 294, 295, 371, 372], leaves the AO-to-MO coefficient matrix still block-diagonal and the corresponding MOs "absolutely localized" on each fragment while also polarized in the presence of each other. The resulting electronic wavefunction is referred to as the polarized state, whose 1PDM is denoted as $\mathbf{P}_{\text{pol}}$. The polarization energy is then defined as the energy lowering of the polarized state relative to the frozen state:

$$\Delta E_{\text{pol}} = E[\mathbf{P}_{\text{pol}}] - E[\mathbf{P}_{\text{frz}}]. \tag{4.4}$$

Since the polarization density comes from variationally minimizing the energy, ΔE_{pol} is negative semi-definite. Lastly, by removing the ALMO constraint, we obtain the fully relaxed state for the intermolecular complex. The energy lowering due to electron delocalization is defined as the charge transfer term (ΔE_{ct}):

$$\Delta E_{\text{ct}} = E[\mathbf{P}_{\text{full}}] - E[\mathbf{P}_{\text{pol}}], \tag{4.5}$$

where $\mathbf{P}_{\text{full}}$ is the 1PDM for the fully relaxed state.

Force Decomposition Analysis

Within a variational EDA scheme like the ALMO-EDA, the analytic nuclear forces associated with each of the intermediate (including the initial and final) states can be obtained. Following the derivations in our previous work [355], namely the adiabatic EDA scheme where the nuclear forces were used to optimize the complex geometry on the frozen, polarized, and fully relaxed surfaces, here we introduce the ALMO-based force decomposition analysis (FDA) method, where the nuclear derivatives of the frozen (ΔE_{frz}), polarization (ΔE_{pol}), and charge transfer (ΔE_{ct}) components of the interaction energy, as well as that of the classical electrostatics component of the frozen interaction ($\Delta E_{\text{cls-elec}}$), are obtained. Note that many of the derivations here can be applied to other variational EDA schemes.

Just as ALMO-EDA decomposes an interaction energy, the ALMO-based FDA decomposes *intermolecular forces* into frozen, polarization, and charge transfer components:

$$\Delta\mathbf{F}_{\text{int}} = \Delta\mathbf{F}_{\text{frz}} + \Delta\mathbf{F}_{\text{pol}} + \Delta\mathbf{F}_{\text{ct}} \tag{4.6}$$

The frozen component of the intermolecular forces ($\Delta\mathbf{F}_{\text{frz}}$) can be obtained by differentiating Eq. (4.3) with respect to the nuclear coordinates:

$$\Delta\mathbf{F}_{\text{frz}} = \mathbf{F}_{\text{frz}}[\mathbf{P}_{\text{frz}}] - \sum_{A} \mathbf{F}_{A}[\mathbf{P}_{A}] \tag{4.7}$$

where $\mathbf{F}_{\text{frz}}$ denotes the forces on the frozen PES, and the term being subtracted on the right-hand side is the collection of isolated fragment forces that arises from the distortion of fragment structures within the complex. Note that these isolated fragment forces can be compared to the nuclear derivatives of the bonded terms in molecular mechanical force fields, while in this work we focus on the intermolecular force components. With superscripts "x" signifying derivatives with respect to the x-th nuclear coordinate (and superscripts "$\mathbf{S}$" and "$\mathbf{\Delta}_A$" in the same fashion), based on the derivation in our previous work [355], the x-th component of forces on the frozen surface, $(F_{\text{frz}})_x$, is given by

$$(F_{\text{frz}})_x = -\left[V_{\text{nn}}^x + \mathbf{P}_{\text{frz}}\cdot\mathbf{h}^x + \frac{1}{2}\mathbf{P}_{\text{frz}}\cdot\mathbf{II}^x\cdot\mathbf{P}_{\text{frz}} + E_{\text{xc}}^x + E_{\text{frz}}^{\mathbf{S}}\cdot\mathbf{S}^x + \sum_{\text{A}}\left(E_{\text{frz}}^{\mathbf{\Delta}_{\text{A}}}\cdot\mathbf{\Delta}_{\text{A}}^x\right)\right], \tag{4.8}$$

where V_{nn} is the nuclear-nuclear coulomb repulsion potential, $\mathbf{h}$ is the core Hamiltonian (kinetic energy and nuclei-electron attractions), II is the AO two-electron integrals, E_{xc} is the Kohn-Sham (KS) exchange-correlation energy, $\mathbf{S}$ is the AO overlap matrix, and $\mathbf{\Delta}_{\text{A}}$ is the matrix of occupied-virtual orbital rotations (variational parameters) within a given fragment. The first four terms are identical to those in the KS-DFT nuclear forces, while the last two terms require treatments that are specific to the frozen PES. The reader is referred to Ref. [355] for details.

Similarly, the polarization contribution to the intermolecular forces can be evaluated by differentiating Eq. (4.4):

$$\Delta\mathbf{F}_{\text{pol}} = \mathbf{F}_{\text{pol}}[\mathbf{P}_{\text{pol}}] - \mathbf{F}_{\text{frz}}[\mathbf{P}_{\text{frz}}] \tag{4.9}$$

The derivation of $\mathbf{F}_{\text{pol}}$ depends on the definition of fragment polarization subspaces in the SCF-MI calculation [373], i.e., the degrees of freedom for each fragment's occupied-virtual mixing to occur. In the simplest case where the full AO space of each fragment is active in the polarization (SCF-MI) calculation as in the 1st-generation ALMO-EDA [183], $\mathbf{F}_{\text{pol}}$ has a similar expression to Eq. (4.8) except that the last term vanishes due to the stationary condition of SCF-MI ($E_{\text{pol}}^{\mathbf{\Delta}_A} = \mathbf{0}$):

$$(F_{\text{pol}})_x = -\left[V_{\text{nn}}^x + \mathbf{P}_{\text{pol}} \cdot \mathbf{h}^x + \frac{1}{2}\mathbf{P}_{\text{pol}} \cdot \mathbf{II}^x \cdot \mathbf{P}_{\text{pol}} + E_{\text{xc}}^x + E_{\text{pol}}^{\mathbf{S}} \cdot \mathbf{S}^x\right] \tag{4.10}$$

Note that in this simplest case, $E_{\text{pol}}^{\mathbf{S}}$ has an identical form to that in the standard SCF energy gradient [355]. Finally, the charge-transfer contribution to the intermolecular forces can be obtained by differentiating Eq. (4.5):

$$\Delta\mathbf{F}_{\text{ct}} = \mathbf{F}_{\text{full}}[\mathbf{P}_{\text{full}}] - \mathbf{F}_{\text{pol}}[\mathbf{P}_{\text{pol}}] \tag{4.11}$$

where $\mathbf{F}_{\text{full}}$ stands for the standard KS-DFT forces for the fully relaxed complex.

The frozen interaction term in ALMO-EDA comprises contributions from permanent electrostatics, Pauli repulsion, and dispersion [374]. To improve the interpretability of FDA results and to facilitate comparison with terms in polarizable force fields, here we introduce how one can evaluate forces arising from "quasi-classical" electrostatics ($\Delta\mathbf{F}_{\text{cls-elec}}$), i.e., coulomb interactions between charge distributions (nuclei and electrons) of different fragments, which can be employed to benchmark forces arising from permanent charge and multipole interactions in a force field. The remainder of $\Delta\mathbf{F}_{\text{frz}}$ then incorporates contributions from the non-electrostatic components of the frozen interaction (Pauli repulsion and dispersion), which we refer to as the van der Waals (vdW) contribution since it corresponds roughly to the sum of attractive and repulsive vdW potential in a force field:

$$\Delta\mathbf{F}_{\text{frz}} = \Delta\mathbf{F}_{\text{cls-elec}} + \Delta\mathbf{F}_{\text{vdw}} \tag{4.12}$$

The quasi-classical electrostatic interaction among N fragments can be expressed in the following compact form:

$$\Delta E_{\text{cls-elec}} = \frac{1}{2}\sum_{A\neq B}^{N}\left[\mathbf{P}_A \cdot \mathbf{V}_{\text{ee+en}}^B + V_{\text{nn}}^{AB}\right] \tag{4.13}$$

where $\mathbf{P}_A$ is the AO-basis 1PDM of isolated fragment A, $\mathbf{V}_{\text{ee+en}}^B$ is the coulomb potential (nuclear and electronic) arising from fragment B, also in the AO basis, and V_{nn}^{AB} is the nuclear-nuclear repulsion potential between fragments A and B. Differentiating Eq. (4.13) yields

$$(\Delta F_{\text{cls-elec}})_x = -\frac{1}{2}\sum_{A\neq B}\left[(\mathbf{P}_A)^x \cdot \mathbf{V}_{\text{ee+en}}^B + \mathbf{P}_A \cdot (\mathbf{V}_{\text{ee+en}}^B)^x + (V_{\text{nn}}^{AB})^x\right], \tag{4.14}$$

where the derivative of isolated fragment density $\mathbf{P}_A$ can be further expanded based on its dependence on fragment A's AO overlap matrix ($\mathbf{S}_A$) and occupied-virtual orbital rotation ($\mathbf{\Delta}_A$):

$$(\mathbf{P}_A)^x = \mathbf{P}_A^{\mathbf{S}_A} \cdot (\mathbf{S}_A)^x + \mathbf{P}_A^{\mathbf{\Delta}_A} \cdot (\mathbf{\Delta}_A)^x \tag{4.15}$$

Note that the detailed forms of $\mathbf{P}_A^{\mathbf{S}_A}$ and $\mathbf{P}_A^{\mathbf{\Delta}_A}$ have been derived in our previous work [355, 375]. For completeness here we show the mathematical details regarding these two derivatives in SI Sec. A1.

The MB-UCB Force Field

The MB-UCB force field [347, 349] was developed based on the principles of the many-body expansion combined with ALMO-EDA variational energy decomposition analysis for each of the terms of the total intermolecular energy

$$E_{\text{inter}} = E_{\text{elec}} + E_{\text{pol}} + E_{\text{CT}} + E_{\text{disp}} + E_{\text{Pauli}} \tag{4.16}$$

This advanced non-reactive force field introduced anisotropic atomic polarizability of the water molecule [376], as well as explicit treatment of charge transfer and charge penetration contributions for both water and aqueous alkali metal and halogen ions [347, 349].

The permanent electrostatics for the MB-UCB force field uses atom centered point multipoles

$$E_{\text{elec}} = \sum_{i<j} \mathbf{M}_i^T \mathbf{T}_{ij} \mathbf{M}_j \tag{4.17}$$

where $\boldsymbol{M}_i^T$ is the multipole coefficient vector and T_{ij} is the multipole interaction tensor that consists of appropriate associated derivatives of $1/r_{ij}$. The monopole-monopole term is modified to describe charge penetration (CP) via separation of the atomic charge into a core nuclear charge, Z_i and smeared electron cloud charge $Z - q_i$. Hence the modified charge-charge electrostatic interactions between two atoms A and B with atomic charges q_A and q_B are expressed as

$$\begin{aligned} E_{\text{elec}}^{q-q} =& \frac{Z_{\text{A}} Z_{\text{B}}}{r} - \frac{Z_{\text{A}} \left(Z_{\text{B}} - q_{\text{B}}\right)}{r} f_{\text{damp}} - \frac{Z_{\text{B}} \left(Z_{\text{A}} - q_{\text{A}}\right)}{r} f_{\text{damp}} \\ &+ \frac{\left(Z_{\text{A}} - q_{\text{A}}\right)\left(Z_{\text{B}} - q_{\text{B}}\right)}{r} f_{\text{damp}}^{\text{overlap}} \end{aligned} \tag{4.18}$$

The two damping functions,

$$\begin{aligned} f_{damp} &= (1 - \exp(-\alpha r)) \\ f_{damp}^{overlap} &= (1 - exp(-\beta_A r))(1 - exp(-\beta_B r)) \end{aligned} \tag{4.19}$$

require two parameters, α and β, to control the damping of core–electron and electron–electron interactions, respectively, in order for the charge penetration effects to vanish rapidly and to

recover the classical Coulombic multipolar interactions at longer distances. We use the CP model parameterization due to Piquemal and co-workers [377].

Many-body polarization is explicitly incorporated by point induced dipoles, μ_{ind}, at each atomic center [361]

$$\boldsymbol{\mu}_i^{\mathrm{ind}} = \alpha_i \left[\sum_j \mathbf{T}_{ij} \mathbf{M}_j - \sum_{j \neq i} \mathbf{T}_{ij}^{\mathrm{d-d}} \boldsymbol{\mu}_j^{\mathrm{ind}} \right] \tag{4.20}$$

where α_i is the atomic polarizability and $\mathbf{T}_{ij}\mathbf{M}_j$ formulates the permanent electric field. T_{ij}^{d-d} is the dipole-dipole interaction tensor in which the off-diagonal blocks of T_{d-d} are Thole damped [378] Cartesian interaction tensors between induced dipoles of two polarizable sites i and j. Unlike other polarizable force fields such as AMOEBA and AMOEBA+ that use rotationally invariant isotropic atomic polarizabilities [346, 366, 372], MB-UCB uses a rank two anisotropic atomic polarizablity tensor. The polarization energy can expressed in terms of induced dipoles as

$$E_{\mathrm{pol}} = -\frac{1}{2} \sum_i \boldsymbol{\mu}_i^{\mathrm{ind}} \mathbf{E}_i \tag{4.21}$$

and the induced dipoles at each multipole site are obtained by solving Equation 4.20 self-consistently [379, 380].

MB-UCB uses an empirical many-body function similar to the polarization energy induced multipoles to incorporate the many-body charge transfer energy [182].

$$\begin{aligned} E_{\mathrm{CT-ind}} &= -\frac{1}{2} \sum_i \mu_i^{\mathrm{ct-ind}} \mathbf{E}_i^{\mathrm{ct}} \\ \boldsymbol{\mu}_i^{\mathrm{ct-ind}} &= \alpha_i^{\mathrm{ct}} \left[\sum_j \mathbf{T}_{ij}^{\mathrm{ct}} \mathbf{M}_j - \sum_{j \neq i} \mathbf{T}_{ij}^{\mathrm{ct[d-d]}} \boldsymbol{\mu}_j^{\mathrm{ct-ind}} \right] \end{aligned} \tag{4.22}$$

where α_i^{ct} controls the charge transfer energy between two multipole sites through a response to the permanent electrostatics, and the multipole interaction matrix (T^{ct}) elements are damped with an exponential damping function.

$$\mathbf{T}_\zeta^{\mathrm{ct}} = -\left[1 - d \exp\left(-bu^3\right)\right] \frac{r_\zeta}{r_{ij}^3}, \quad \zeta = x, y, z \qquad u = \frac{r_{ij}}{\left(\alpha_i^{\mathrm{ct}} \alpha_j^{\mathrm{ct}}\right)^{1/6}} \tag{4.23}$$

The three parameters α_{ct}, b and d are responsible for the fast exponential decay of the charge transfer energy, which should be more short-ranged than polarization [347, 349].

The remaining energy terms are Pauli repulsion and dispersion, and are modeled in MB-UCB as a van der Waals interaction using a buffered 14-7 pairwise-additive function proposed by Halgren [381] and utilized in all AMOEBA force fields [361, 366, 372].

$$E_{\mathrm{vdW}} = \sum_{i<j} \epsilon_{ij} \left(\frac{1+\delta}{\sigma_{ij}+\delta} \right)^7 \left(\frac{1+\gamma}{\sigma_{ij}^7+\gamma} - 2 \right) \tag{4.24}$$

where ϵ defines the energy scale, $\sigma = R_0/r$ is the distance between two atoms, and R_0 is the distance corresponding to the minimum energy. Like AMOEBA [166], we set the two constants δ and γ to 0.12 and 0.07, respectively. Given the total functional forms of the energy terms of MB-UCB, the corresponding force terms are easily defined through the usual chain rule formulations and easily compared to the FDA analysis proposed here.

Computational details

The force decomposition analysis method discussed here was implemented in a developer version of Q-Chem 5 [237]. The geometries used for molecular calculations were optimized and run at the ωB97X-D [96]/def2-TZVPPD [382] level of theory, with exception to the gold/silver CO_2 complex, where ωB97X-V [97] density functional instead along with the appropriate def2-ECP [383]. To diagnose the atomic forces of MB-UCB, we use the same level of theory used in the original paper [347], namely ωB97X-V [97]/def2-QZVPPD [382]. DFT numerical integration was performed on (99,590) grid for XC functional and SG-1 [384] for non-local correlation. All geometries are included in the Supplementary Material (SI).

Fifty water dimer geometries were used to compare the atomic force contributions between FDA and MB-UCB. The geometries were taken from the iAMOEBA training data set [364], where pairs of molecules were randomly picked from AMOEBA liquid water simulation between 257.15 - 373.15 K such that it is representative of a wide range of the phase space. The geometries are provided in the supplementary information.

Forces are turned into internal coordinates by a linear transformation using the pseudo-inverse of Wilson's B matrix [368, 385, 386]. The B matrix was generated using Q-Chem 5 [237] with a minor modification of connecting intra-fragment atoms first and then minimally connecting the inter-fragment atoms to avoid non-intuitive bonds and angles. Details are included in SI Sec. A2

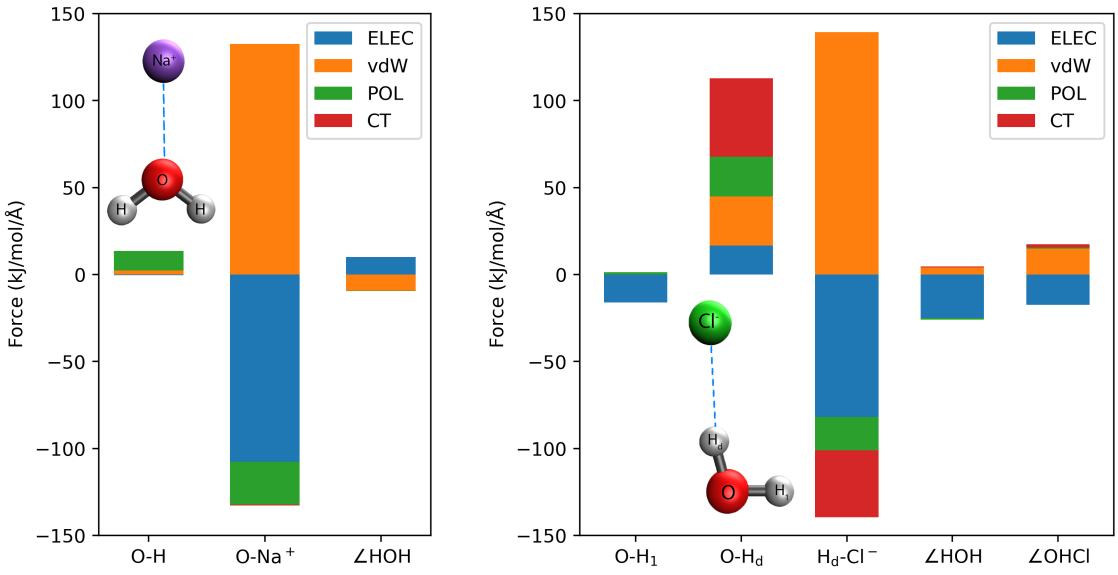
4.3 Results and discussion

$H_2O\cdots Na^+$, $Cl^-\cdots HOH$ and the water dimer

First we look at the water molecule interacting with an innocent cation, Na^+, and a simple anion, Cl^-. Aside from the importance of these examples in understanding water-ion interactions, they will illustrate the nature of the FDA information, as well as its representation in internal coordinates. We use geometries where the position of the ion is optimized relative to a water molecule constrained to its isolated geometry. This is a convenient choice because $\mathbf{F}_{\text{tot}} = \Delta\mathbf{F}_{\text{int}}$ since $\mathbf{F}_A = \mathbf{0}$ for $A = H_2O$. As a result, the interfragment degrees of freedom will have zero net force, as well as zero resultant force in the FDA in order to see how the FDA components cancel each other out. By contrast, there will be nonzero $\Delta\mathbf{F}_{\text{int}}$ for intramolecular degrees of freedom, which will indicate how such internal coordinates will

deform in a fully optimized complex. The FDA will reveal which component contributions are primarily responsible for such changes.

At equilibrium, the $Na^+ - O$ vector is aligned with the water dipole vector, optimizing the charge-dipole interaction. With C_{2v} symmetry, there are only three non-redundant internal coordinates, namely O–H, O–Na, and ∠HOH. The FDA is shown in the left panel of Fig. 4.1. By far the most interesting result is $\Delta\mathbf{F}_{\text{int}}(O - Na^+)$, which is overall zero, as a result of a strong force of extension due to Pauli repulsion (the van der Waals term in Fig. 4.1) being compensated by an equally strong force of contraction due to electrostatic attraction. The electrostatic attraction force is about 80% due to the permanent electrostatics, and only 20% due to polarization (of water by Na^+). There is negligible contribution from charge transfer, emphasizing the innocent nature of Na^+ as an ineffective Lewis acid. The O–H bonds are remote from Na^+, so the forces distorting the optimal monomer geometry in the complex are small. The largest formation force is $\Delta\mathbf{F}_{\text{int}}(O - H)$, which has a small force of extension driven by polarization. We can understand this effect as a result of promotion of a fraction of an electron from the σ_{OH} orbitals to antibonding orbitals, presumably with σ^*_{OH} character.



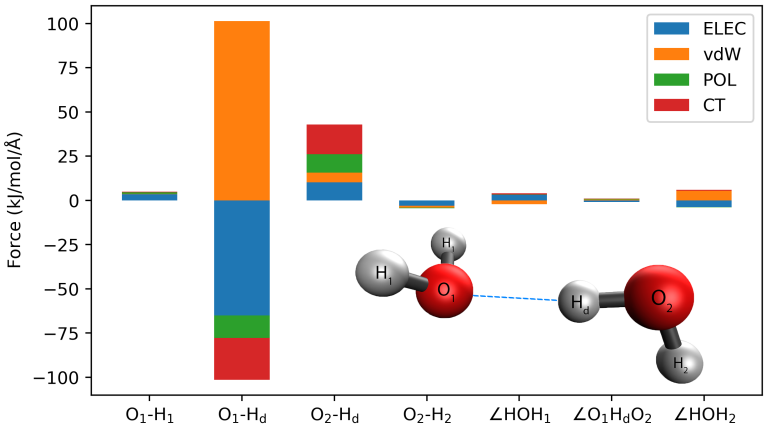
ELEC	-0.3	-107.6	10.0
vdW	2.4	132.5	-9.3
POL	11.1	-24.6	-0.2
CT	0.2	-0.6	0.0
Total	13.3	-0.3	0.6

ELEC	-16.0	16.6	-81.9	-25.0	-17.5
vdW	0.1	28.3	139.4	4.0	14.9
POL	1.3	22.7	-19.0	-0.9	0.6
CT	0.1	45.1	-38.5	0.8	2.0
Total	-14.6	112.8	-0.1	-21.1	0.0

Figure 4.1: FDA results in the internal coordinates for H_2O interacting with Na^+ (left) and Cl^- (right). These forces are evaluated at the PES minima, keeping the water molecule at its isolated geometry. The tables at the bottom replicate the data with numbers for easier comparison.

FDA for the hydrogen bonded complex formed between water and the chloride anion, as given in the right panel of Fig. 4.1, presents an interesting contrast with the $H_2O\cdots Na^+$ case, which was previously analyzed by the ALMO-EDA [375, 387]. Despite on-going debate [195, 388–391], it is fairly well-established that hydrogen-bonds involve significant contributions from permanent and induced electrostatics, and charge transfer [184, 355, 391], in competition with Pauli repulsion. Focusing first on the inter-fragment H_d-Cl^- force, which is optimized to zero, we see three forces of contraction (permanent electrostatics > charge transfer > polarization) balanced by the extension force due to Pauli repulsion. From a force equilibrium perspective, this very nicely illustrates the "driving forces" that give rise to the hydrogen bond. The other interfragment coordinate, $\angle OH_dCl$, is optimized as a balance between permanent electrostatics (attempting to shrink the angle), and Pauli repulsion (attempting to enlarge the angle). This competition is controlled by the frozen part of the interaction energy, as previously noted for the water dimer [355]. Within the water molecule, there is a strong force of extension along the OH_d bond. Its primary origin is charge trans-

fer, followed by Pauli repulsion, induced electrostatics and permanent electrostatics. Both CT and polarization can be readily understood in terms of partial occupation of the $\sigma^*_{OH_d}$ orbitals.



	O_1-H_1	O_1-H_d	O_2-H_d	O_2-H_2	$\angle HOH_1$	$\angle O_1H_dO_2$	$\angle HOH_2$
ELEC	3.5	-65.1	10.1	-3.0	3.1	-0.9	-3.7
vdW	0.3	101.3	5.6	-1.1	-2.1	0.5	5.5
POL	0.8	-12.6	10.3	-0.3	0.1	0.4	-0.2
CT	0.4	-23.6	16.8	-0.0	0.8	0.1	0.5
Total	5.0	-0.0	42.9	-4.5	1.9	0.1	2.2

Figure 4.2: FDA results in the internal coordinates for the water dimer at the minimum-energy distance, with the monomers fixed in their isolated geometries. The table at the bottom replicates the data with numbers for easier comparison.

Next, we examine FDA for the hydrogen-bonded water dimer, as shown in Fig. 4.2; this system has also been carefully analyzed previously by the ALMO-EDA [195, 375]. The zero net force on the hydrogen-bond coordinate, O_1H_d, shows the FDA view of this characteristic hydrogen-bond interaction. Classical electrostatics dominates the forces seeking to further shorten the hydrogen bond, consistent with force-field viewpoints. Intermolecular charge transfer is the second strongest force of contraction, followed by polarization. Pauli repulsion provides an exactly balancing force of extension. Within the proton-donor water molecule (which of course is the electron pair acceptor), the intramolecular O_2H_d bond constrained to the geometry of the free water molecule experiences a force of extension to which all components contribute with the same sign. Similar to the water-chloride complex, charge transfer and polarization partially occupy the antibonding $\sigma^*_{OH_d}$ orbitals and are the leading drivers of O_2H_d bond elongation.

Assessing force components of an advanced water force field

Generating the force vectors corresponding to each energy term in the ALMO-EDA yields a greatly augmented set of data at each geometry. Such data can in principle be employed to

aid in the development or validation of advanced force fields, perhaps in conjunction with powerful existing tools such as Force Balance [372, 392] or machine learning [393, 394]. To illustrate the use of FDA data, we assess the forces that are obtained from a recently reported force field for water, MB-UCB [347]. Although the energies of MB-UCB have already shown to be in excellent agreement with ALMO-EDA [347], no comparable assessment of the decomposed forces has yet been done, although total and force components of the complete energy derivative have been assessed for other force fields such as iAMOEBA [364] and AMOEBA14 [366] to which we compare below. The term-by-term force contributions from the FDA against MB-UCB has been assessed for a set of 50 water dimer geometries extracted from finite temperature MD trajectory as described in the computational details section. For these snapshots, the ALMO-EDA energy components and the corresponding contributions for MB-UCB, shown in SI Sec. A3, show excellent agreement, as expected based on our previous work [347].

We begin our assessment by comparing the FDA and MB-UCB force components on the center of mass (CoM) of each water molecule in the data set broken down by interaction. The COM forces are a sum of all atomic forces on a molecule, also referred to as molecular forces or net forces [364, 395]. The results are shown as correlation plots in Fig. 4.3 in which the RMS error in the total CoM force is $\sim$8 kJ/mol/Å. This is a reasonably small error when considering the fact that a DFT geometry optimization is considered converged at a maximum force of $\sim$1-2 kJ/mol/Å, and is comparable to the $\sim$10 kJ/mol/Å RMS error in AMOEBA and iAMOEBA forces versus ab initio forces reported for water clusters [364, 366]. Perhaps the most important point that emerges from Fig. 4.3 is the fact that the RMSD in each non-bonded

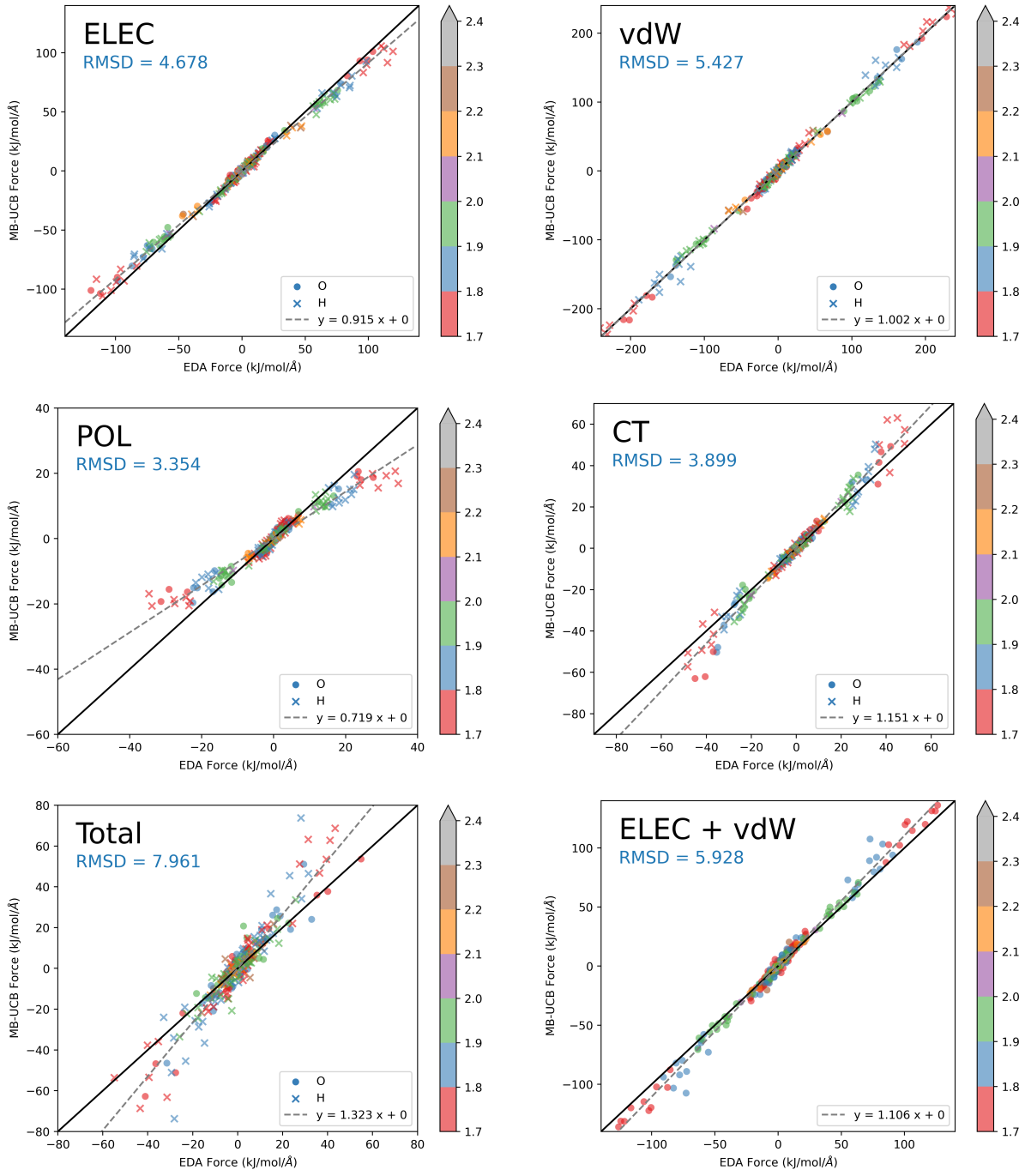


Figure 4.3: *Correlation between FDA decomposed forces and the corresponding MB-UCB forces on the centers of mass of each water molecule for a sample of 50 water dimer geometries.* The force decompositions considered are electrostatics (ELEC), van der Waals (vdW), polarization (POL), and charge transfer (CT), and the total intermolecular interaction (TOT). The color bins indicate the distance to the closest atom of the other fragment, i.e., small numbers indicate the dimer is in the compressed region. The equilibrium water dimer closest contact atoms sit at 1.9 Å, which corresponds to the data points colored in blue. The dashed line corresponds to a least squares fit of the errors, where the line fit equation is shown in the legend of each plot.

contribution to the MB-UCB CoM force is smaller than the RMSD in the MB-UCB total force. Even the very large ELEC and vdW forces exhibit RMSD values of only 4.7 and 5.4 kJ/mol/Å, respectively. The largest deviations are associated with the largest forces, as expected, where short-range damping, such as Thole damping of polarization [378], are likely to exert an influence. Overall, we can conclude that the decomposed contributions to the CoM forces via MB-UCB are as good or better behaved than the total MB-UCB CoM forces.

A more stringent FDA test is to assess the errors in the Cartesian forces on each atom, for which a correlation plot between the FDA and MB-UCB decompositions is shown in Fig. 4.4. The overall RMSD value is increased by only ∼10% for the atomic forces vs the CoM forces, rising to ∼8.8 kJ/mol/Å, which is encouragingly good performance when compared to iAMOEBA or AMOEBA, in which atomic forces showed RMS errors more than twice as large as CoM forces [364]. Although the RMS of the ELEC atomic forces increases relative to the COM electrostatic forces, they are still comparable to the total atomic force errors.

However, the vdW term shows a significantly larger atomic force error compared to the total or COM force error. It is pertinent to mention that MB-UCB situates the vdW centers for the hydrogen atoms at a fixed fraction (0.91) of the OH bond length, rather than at the atomic centers themselves. Hence the virtual site forces must be redistributed over the particles with mass in a consistent way, which only guarantees that the total force is preserved, and may explain some of the vdW force deviations observed. Even so, there is some error cancellation between ELEC and vdW atomic forces as seen in the ELEC+vdW plot in Fig. 4.4, which was not the case in the CoM force components. Finally, the more challenging nature of the atomic force components (and the total atomic force) is also evident in the fact that the largest errors no longer occur predominantly at the largest absolute force values. Fig. 4.4 shows RMS errors when either MB-UCB predicts near zero atomic forces compared to finite FDA forces (such as for POL or CT) or that finite MB-UCB atomic forces are found when FDA forces are near zero (for example, vdW).

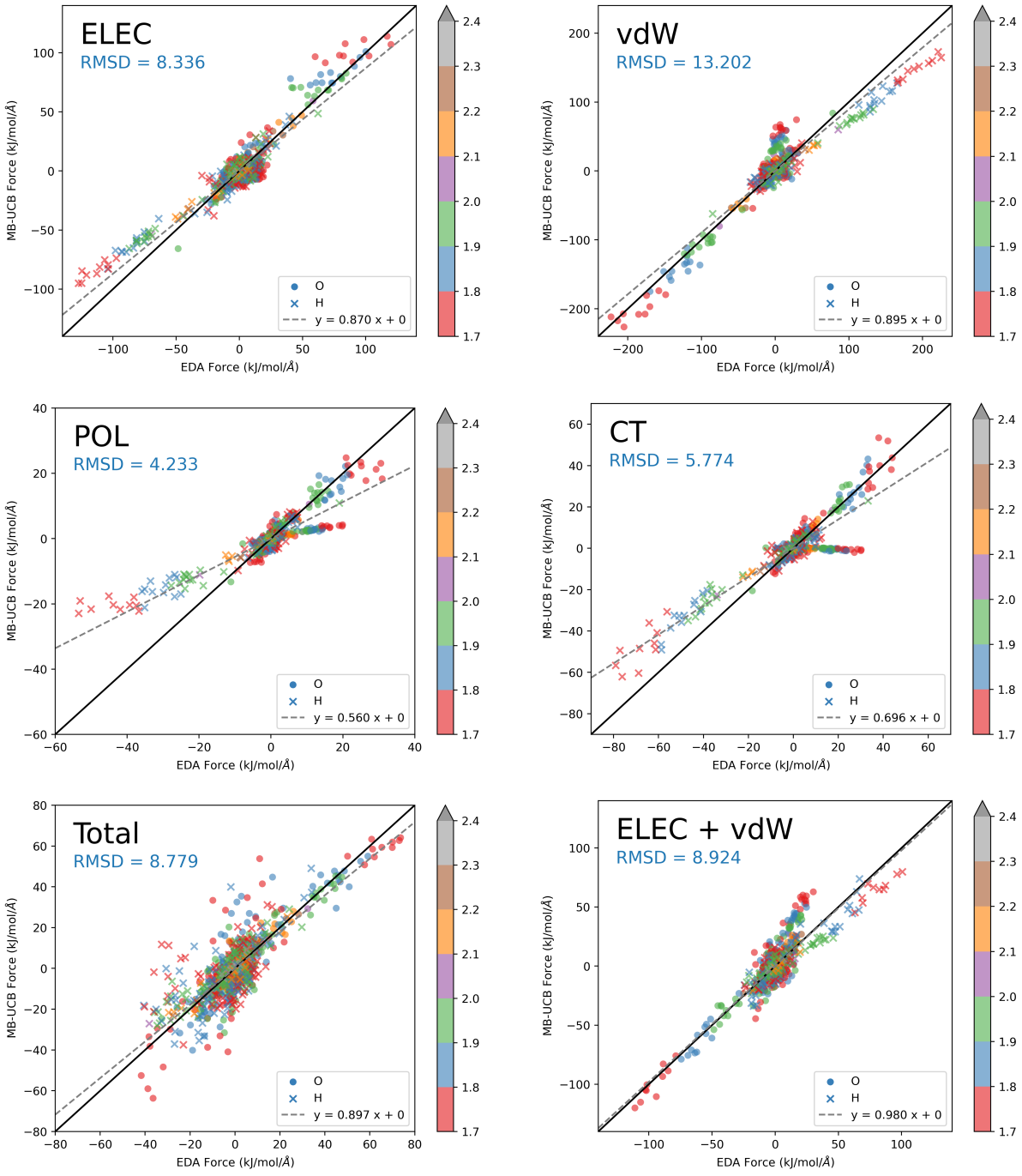


Figure 4.4: *Correlation between FDA decomposed component forces and the corresponding MB-UCB forces on the atomic centers of each water molecule for a sample of 50 water dimer geometries.* Other details are as defined in Fig. 4.3.

One more way to compare MB-UCB forces against the FDA results is to separately evaluate the RMS deviations in the total CoM force (left panel) and the atomic forces (right panel) as a function of the closest intermolecular distance in Fig. 4.5. These plots make it clear that errors decay rapidly as a function of intermolecular separation. The plots also serve to emphasize the fact that the quality of the individual MB-UCB decomposed CoM force is statistically better than the MB-UCB total force. On the other hand, at intermolecular distances of 2 Å and shorter, errors in the MB-UCB vdW force contribution are larger than the total MB-UCB RMSD; in other words, there is partial error compensation with the ELEC term in particular. It is encouraging that the errors associated with the MB-UCB description of charge transfer and polarization contributions remain relatively low even in the compressed region for both the atomic force and CoM force, although their magnitude increases with the reduction of the closest contact distance. The CT term shows more scope for improvement, which is likely to be a result of the less physically appropriate form that was employed within MB-UCB [347].

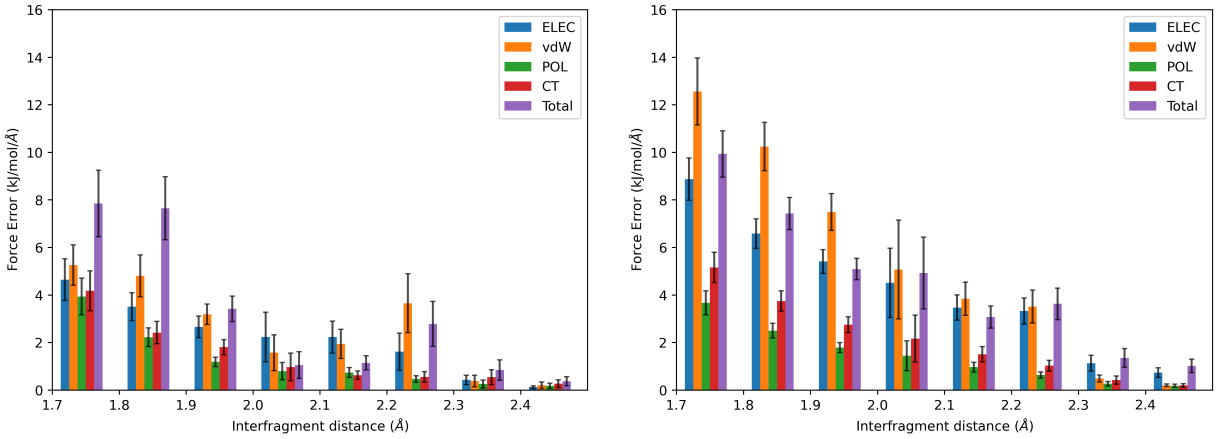


Figure 4.5: *Mean absolute deviations in the total CoM force (left panel) and the atomic forces (right panel) as a function of the closest intermolecular distance.* The mean errors in MB-UCB forces plotted against the closest contact between the two water molecules broken down into the non-bonded components of interaction. The left panel applies to the CoM force on each water molecule, while the right panel applies to the atomic forces. Values plotted are the RMSDs for all data points within each 0.1 Å bin of closest intermolecular distance. The error bars indicate 95% confidence interval.

Interaction of CO_2 with Au^- and Ag^-

The reduction of CO_2 to $CO_2^{\bullet-}$ is the first step on the pathway towards conversion of CO_2 into fuels, of which the simplest is 2-electron reduction to CO_2 [396]. The reverse reaction, CO oxidation to CO_2, is also well-studied [397]. At the level of model systems, negatively

charged gold oxide clusters have been shown to react with CO to yield CO_2 [398], via reactions as simple as $AuO^- + CO \longrightarrow Au^- + CO_2$. The exit channel complex, $[Au \cdots CO_2]^-$, has been studied as part of that reaction [398], as well as characterized by separate experiments and computations [399–401]. Remarkably, as shown in Fig. 4.6(a) and (b), there are two local minima in the exit channel: a strongly bound chemisorbed structure, which exhibits significant activation (i.e., reduction) of the CO_2 ($\angle OCO = 143°$), and a physisorbed complex where CO_2 is not activated ($\angle OCO = 172°$).

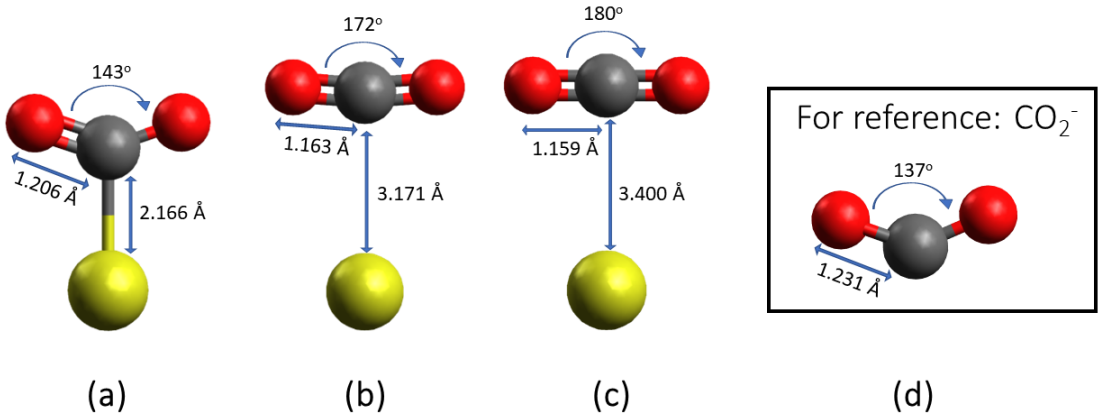


Figure 4.6: A diagram showing the different configurations of the Au^-CO_2 with the labeled bond distances and angles; due to C_{2v} symmetry, there are only 3 non-redundant internal coordinates, which are the Au−C distance, the C−O bond length, and the CO_2 bending coordinate. (a) the chemisorbed species at $R(\mathrm{AuC}) \sim 2.2$ Å, (b) the physisorbed species at $R(\mathrm{AuC}) \sim 3.2$ Å, (c) a constrained geometry (R(AuC) optimized with CO_2 fixed at its optimal isolated geometry) exhibiting a minimum at $R(\mathrm{AuC}) \sim 3.4$ Å, (d) the charged anion of $CO_2{}^{\bullet -}$.

On the other hand, the silver anion was reported experimentally to exhibit only the physisorbed species [401], perhaps reflecting the smaller size of the gold atom compared to silver due to relativistic contraction. We show a fully relaxed potential energy scan along the M−C distance with the energy decomposition analysis results in Fig. 4.7. General agreement with the experimental facts is evident in the PES scan. The size effect is already clear with Ag showing more repulsive van der Waals interactions (sum of Pauli repulsion and dispersion) at 4 Å and stronger electrostatic attraction than Au. Despite charge transfer being a dominant contribution to the interaction, CT is very comparable for Au and Ag at shorter M−C distances, although Ag's CT is stronger than Au's at longer distances due to size. With an ionization energy of only 126 kJ/mol [402], Ag^- is a stronger electron donor (Lewis base) than Au^-, whose ionization energy is 223 kJ/mol [403].

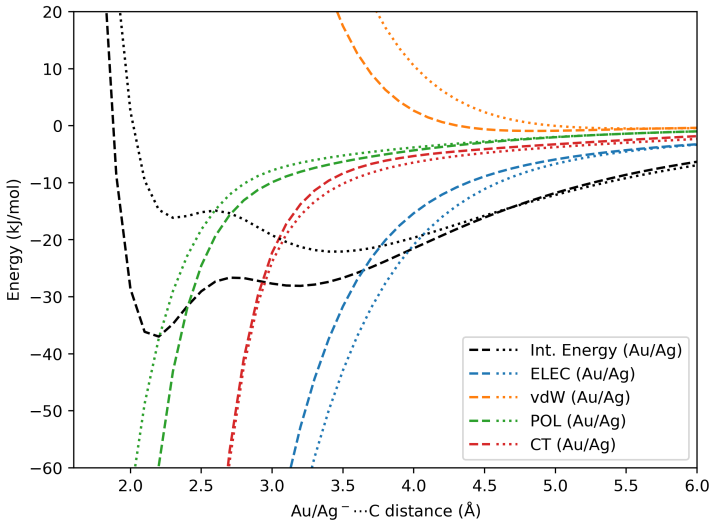
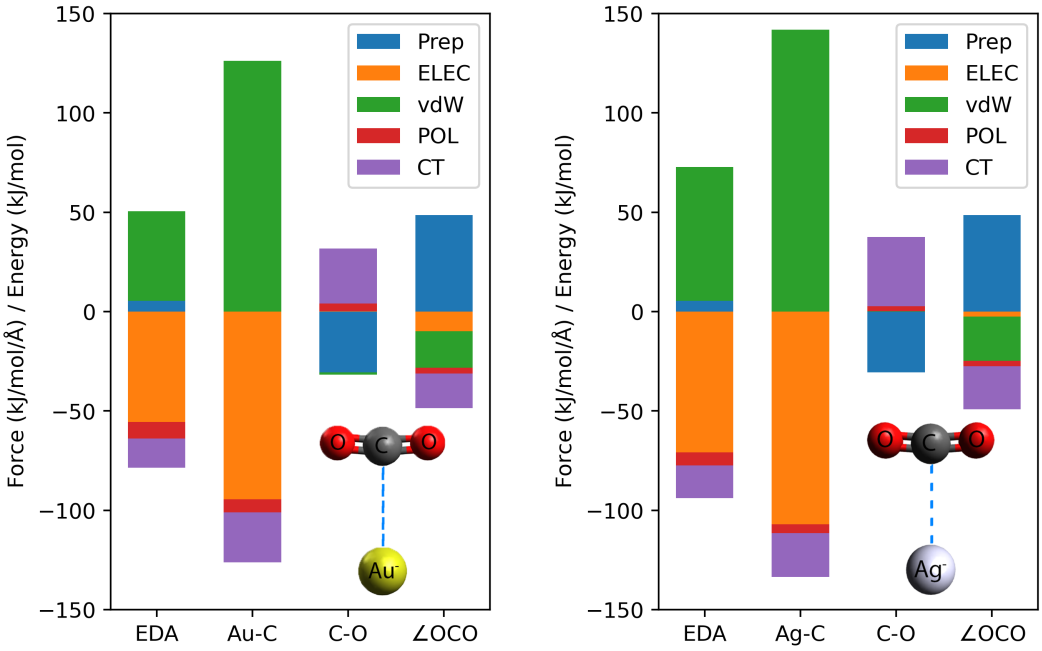


Figure 4.7: Fully relaxed potential energy surface scans (kJ/mol) for $Au^{-}CO_2$ (black dashes) and $Ag^{-}CO_2$ (black dots) with the EDA components (dashes for $Au^{-}CO_2$ and dots for $Ag^{-}CO_2$).

Next, we look into the forces for both the physisorbed and chemisorbed species. For easier comparison, we take the geometries of two minima for the gold complex and use these same geometries for silver. Since the CO_2 molecule in these geometries is distorted compared to the isolated molecule, we will refer to that difference in energy as a preparation (Prep) energy. Similarly, we refer to the forces that arise from the preparation energy (which will cause the CO_2 to relax back to its isolated geometry) as the preparation force. The FDA results at the physisorbed geometry are shown in Fig. 4.8, while results at the chemisorbed geometry are shown in Fig. 4.9.



Prep	5.5	0.0	-30.5	48.6
ELEC	-55.6	-94.5	0.2	-9.8
vdW	44.9	126.1	-1.2	-18.4
POL	-8.3	-6.6	3.9	-2.9
CT	-14.6	-25.0	27.8	-17.4
Total	-28.1	0.0	0.0	-0.0

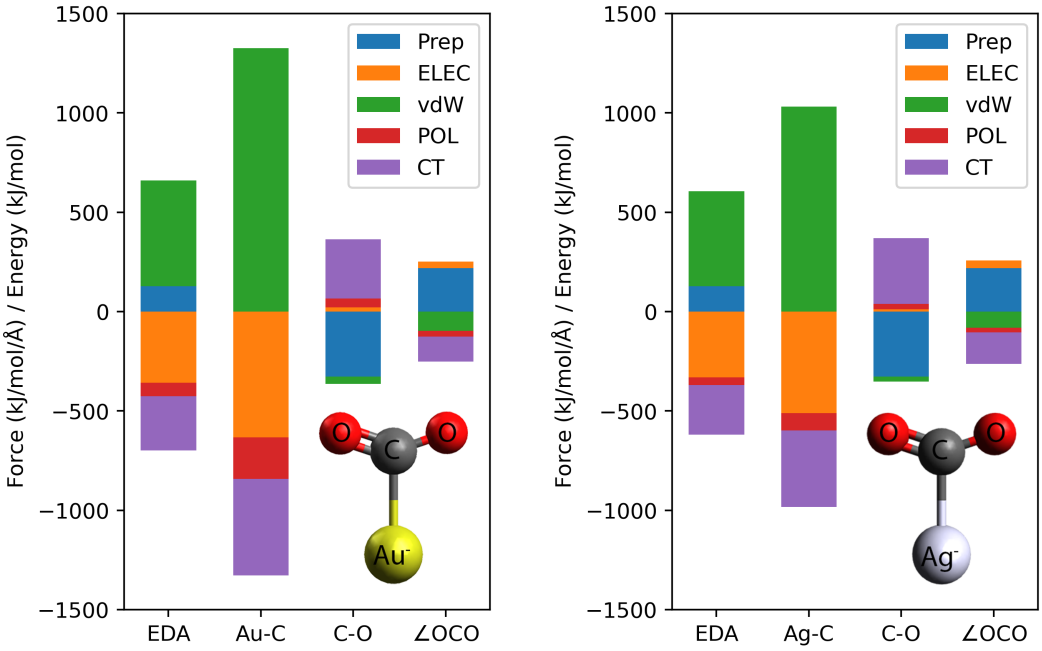
Prep	5.5	0.0	-30.5	48.6
ELEC	-70.8	-107.0	0.1	-2.5
vdW	67.3	141.9	0.1	-22.3
POL	-6.6	-4.6	2.7	-2.7
CT	-16.4	-22.1	34.8	-21.5
Total	-21.0	8.3	7.0	-0.5

Figure 4.8: Comparison of the EDA components (in kJ/mol) and the decomposed forces (in kJ/mol/Å) in internal coordinates for physisorbed $Au^- \cdots CO_2$ (left panel) and $Ag^- \cdots CO_2$ (right panel) complexes, both evaluated at the nuclear coordinates optimized for $Au^- \cdots CO_2$. The energies and forces are decomposed into preparation (Prep), electrostatics (ELEC), van der Waals (vdW), polarization (POL), and charge transfer (CT), and the total intermolecular interaction (Total). The table summarizes the same data with additional significant figures.

We first discuss the physisorption results shown in Fig. 4.8. The larger size of Ag^- versus Au^- results in a more attractive electrostatic interaction as well as stronger Pauli repulsion in the van der Waals term, with no significant difference in polarization and charge transfer terms. As we use the optimized $Au^- \cdots CO_2$ nuclear coordinates, the net force along each internal coordinate is exactly zero for the $Au^- \cdots CO_2$ complex. Thus inspection of the FDA reveals an exact force balance. Along the Au−C coordinate, van der Waals repulsion is primarily balanced by electrostatics, with small contributions from polarization and CT also favoring shorter bonds. The fact that those latter contributions are so small

indicates that CO_2 is scarcely activated, consistent with the near linearity of its optimized geometry. The preparation force favors removing the slight lengthening of the CO bond and very slight bending of the CO_2. In opposition, the drive for CO bond lengthening comes almost entirely from CT, while van der Waals and CT both favor increased angle bending. Comparing $Ag^- \cdots CO_2$ against $Au^- \cdots CO_2$ shows relatively subtle differences associated with the stronger Pauli repulsion forces in the Ag system, which favor longer Ag−C separation, and extension of the C−O distance as a result of its stronger CT (due to better donor-acceptor overlap, as well as Ag^- being a stronger Lewis base). Finally, as regards the physical driving forces behind the physisorbed complex, both EDA and FDA reveal it to be synergy between dispersion (as indicated by the net binding provided by ELEC+vdW), charge transfer, and polarization.

For the chemisorbed $Au^- - CO_2$ and $Ag^- - CO_2$ species shown in Fig. 4.9, at the coordinates of the optimized $Au^- - CO_2$ complex, there is a binding energy difference of 24 kJ/mol in favor of the Au complex. Note that the scale for Fig. 4.9 is 10 times larger than for the physisorbed structures given in Fig. 4.8. By far the dominant driving force behind the chemisorption geometry is charge transfer. The smaller gold anion exhibits stronger binding from electrostatics, polarization, and charge transfer as well as more repulsive van der Waals interaction compared to the more diffuse silver anion. Despite the lower ionization energy of Ag^- vs. Au^-, the compactness of the gold anion makes charge transfer in $Au^- - CO_2$ significantly more attractive than that in $Ag^- - CO_2$ in which the anion is more diffuse. Accordingly, FDA on $Ag^- - CO_2$ shows a net force for Ag−C elongation. Turning to FDA within the CO_2 subunit, CT (elongation) vs. Prep (contraction) determine the net force on the C−O bond. On the other hand, the net ∠OCO force displays an interesting synergy between the van der Waals repulsion and charge transfer (both favoring bending), versus the preparation force (favoring linearization). To sum up, the compactness of the gold versus silver anion trumps the stronger Lewis basicity of the silver anion in leading to much stronger chemisorption in the Au complex.



Prep	127.1	0.0	-326.4	218.9
ELEC	-358.7	-633.3	20.4	33.0
vdW	533.7	1326.7	-37.9	-96.9
POL	-67.1	-208.4	46.0	-28.0
CT	-272.3	-485.1	297.8	-127.0
Total	-37.2	-0.1	-0.1	0.0

Prep	127.1	0.0	-326.4	218.9
ELEC	-330.6	-510.4	12.4	37.6
vdW	479.0	1030.9	-25.5	-80.6
POL	-39.9	-88.3	26.9	-23.7
CT	-249.0	-383.8	330.1	-158.0
Total	-13.4	48.4	17.4	-5.8

Figure 4.9: Comparison of the EDA components (in kJ/mol) and the decomposed forces (in kJ/mol/Å) in internal coordinates for chemisorbed $Au^- - CO_2$ (left panel) and $Ag^- - CO_2$ (right panel) complexes, both evaluated at the nuclear coordinates optimized for $Au^- - CO_2$. The energies and forces are decomposed into preparation (Prep), electrostatics (ELEC), van der Waals (vdW), polarization (POL), and charge transfer (CT), and the total intermolecular interaction (Total). The table summarizes the same data with additional significant figures.

4.4 Conclusions

We have reported theory, implementation, and model applications of an extension to the adiabatic energy decomposition analysis [355] to perform force decomposition analysis of the force components obtained from an EDA method. In particular, the variational absolutely localized molecular orbital EDA (ALMO-EDA) approach [185] is used to analyze Kohn-Sham density functional theory calculations on molecular complexes by differentiating key intermediate energies associated with each non-bonded term. The result is a more information-rich

vector of how the different physical driving forces of intermolecular interactions affect each atomic or internal coordinate force within a complex. We expect that our FDA approach can be readily applied to other variational EDA schemes and extended to other molecular properties besides nuclear forces.

More specifically, the net force on either each atom or each internal coordinate of a molecular complex is decomposed into the following physically interpretable contributions.

1. A preparation force (Prep), which results from deforming a fragment optimized in isolation to its geometry in the complex. The Prep force will always favor restoring the fragment to its isolated geometry.

2. Forces associated with quasi-classical electrostatics (ELEC), and van der Waals (vdW) interactions (including attractive dispersion and repulsive Pauli interactions) are obtained which sum to the net force resulting from the frozen interaction energy [374] of the ALMO-EDA method. For strongly interacting complexes ELEC and vdW forces are strong and opposite in sign, and it can be advantageous to instead examine the frozen force.

3. Forces associated with the polarization (POL) of the complex [373], as described by the self-consistent field for molecular interactions (SCF-MI) approach [292, 294, 295, 371, 372] in the basis of fragment atomic orbitals.

4. Forces associated with charge delocalization or charge transfer [375, 404] between the fragments comprising the complex, which represent the final increment to obtain the total forces.

The model applications presented here are of some intrinsic interest, as well as serving to illustrate the future utility of the FDA for more advanced problems. We presented three sets of examples:

1. We used FDA to examine the Na^+H_2O and Cl^-H_2O complexes, keeping H_2O constrained to its free-molecule geometry. The resulting force balance along the intermolecular distance revealed a greater role for CT in the chloride complex. The net forces within the water molecule showed the role of different components on the intermolecular interaction in distorting the geometry.

2. To illustrate the potential value of FDA to the advanced force field development community, we assessed the fidelity of contributions to the MB-UCB water force field against the FDA components on snapshots of the water dimer. The results showed very good performance for the total atomic forces, and particularly good performance for the center of mass force decompositions, whose RMSD vs FDA components was smaller than the total RMSD.

3. The FDA was also employed to analyze the physisorbed and chemisorbed complexes formed between Au^- and CO_2, and to compare them against the corresponding Ag^-CO_2

complexes. The results showed that while Ag^- is a stronger electron donor than Au^-, the smaller size of Au^- is crucial to the stronger chemisorption of CO_2 to it.

Appendix A

Additional Information on the Force Decomposition Analysis

A.1 Derivation of classical electrostatics derivative

In the derivation below, we follow the steps and notation of the adiabatic energy decomposition analysis paper [355]. We skip steps that were derived explicitly and explained sufficiently. In passing, the notation below uses $\mu, \nu, \lambda, ..$ for AO basis functions, $p, q, r, ...$ for molecular orbitals, $i, j, k, ..$ for occupied orbitals, $a, b, c, ..$ for virtual orbitals, and $A, B, C, ..$ for fragments.

In the ALMO-EDA, the quasi-classical electrostatic energy uses the isolated densities of each fragment, A:

$$E_{cls.elec.} = \frac{1}{2} \sum_{A,B \neq A} \int \int \frac{\rho^{A,iso}(r_1)\rho^{B,iso}(r_2)}{|r_1 - r_2|} dr_1 dr_2 \tag{A.1}$$

where

$$\rho^{A,iso}(r) = \sum_{\mu\nu} P^{A,iso}_{\mu\nu} \phi_\mu(r)\phi_\nu(r) + \sum_{a \in A} Z_a \delta(r - R_a) \tag{A.2}$$

where $\mathbf{P}^{iso} = \mathbf{C}_{occ}\mathbf{C}^{\dagger}_{occ}$ are the isolated electronic densities, in terms of fragment MOs that

are non-orthogonal between fragments. In matrix form this interaction is simply:

$$E_{cls.elec.} = \frac{1}{2}\sum_{A,B\neq A} V_{ee}^{AB} + V_{Ne}^{AB} + V_{NN}^{AB} \tag{A.3a}$$

$$= \frac{1}{2}\sum_{A,B\neq A}\sum_{\mu\nu\in A}\big(P_{\mu\nu}^{A,iso}\sum_{\lambda\sigma\in B}(\mu\nu|\lambda\sigma)P_{\lambda\sigma}^{B,iso}\big) + \big(P_{\mu\nu}^{A,iso}(V_{Ne}^{B})_{\mu\nu}\big) + V_{NN}^{AB} \tag{A.3b}$$

$$= \frac{1}{2}\sum_{A,B\neq A}\sum_{\mu\nu\in A} P_{\mu\nu}^{A,iso}\big(\sum_{\lambda\sigma\in B}(\mu\nu|\lambda\sigma)P_{\lambda\sigma}^{B,iso} + (V_{Ne}^{B})_{\mu\nu}\big) + V_{NN}^{AB} \tag{A.3c}$$

$$= \frac{1}{2}\sum_{A}\mathbf{P}^{A,iso}\cdot\mathbf{V}^{A} + \frac{1}{2}\sum_{A,B\neq A} V_{NN}^{AB} \tag{A.3d}$$

where $\mathbf{V}^A$ is the electric field (Coulombic potential) experienced by fragment A from other fragments. The dot products are element-by-element multiplication.

The derivative with respect to the nuclear displacement of a nucleus on fragment A, in one of its components, call it x, is only non-zero if fragment A is involved in the interaction. The term V_{NN}^{AB} and its derivative are the usual ones in quantum chemistry and won't be discussed here. The force is the negative of the derivative of the nuclear displacement and it looks like:

$$\frac{\partial}{\partial x}E_{cls.elec} = \frac{1}{2}\big(\frac{\partial\mathbf{P}^{A,iso}}{\partial\mathbf{S}}\frac{\partial\mathbf{S}}{\partial x}\cdot V^{A} + \frac{\partial\mathbf{P}^{A,iso}}{\partial\mathbf{\Delta}}\frac{\partial\mathbf{\Delta}}{\partial x}\cdot V^{A} + \mathbf{P}^{A,iso}\cdot\mathbf{V}^{A^x}\big) + \frac{1}{2}\sum_{B\neq A} V_{NN}^{AB^x} \tag{A.4}$$

where the coulombic term derivative, $\mathbf{V}^{A^x}$, is the usual one in quantum chemistry and $\mathbf{\Delta}$ is the orbital rotation matrix. Below we deal with the classical density derivative, $\mathbf{P}^{A,iso^x}$, starting with the overlap chain,

$$\frac{\partial\mathbf{P}^{A,iso}}{\partial\mathbf{S}}\frac{\partial\mathbf{S}}{\partial x}\cdot\mathbf{V}^{B} = -\frac{1}{2}(\mathbf{P}\mathbf{S}^{x}\mathbf{P} + \mathbf{P}\mathbf{S}^{x}\mathbf{P})\cdot\mathbf{V}^{A} = -(\mathbf{P}\mathbf{V}^{A}\mathbf{P})\cdot\mathbf{S}^{x} \tag{A.5}$$

where the matrices were rearranged for easier evaluation with other terms as will be seen below.

Now for the orbital response chain where we have

$$\frac{\partial\mathbf{P}^{A,iso}}{\partial\mathbf{\Delta}}\frac{\partial\mathbf{\Delta}}{\partial x}\cdot V^{A} = \sum_{\mu\nu\in A}\sum_{pq}\frac{\partial P_{\mu\nu}^{A,iso}}{\partial\Delta_{pq}}\frac{\partial\Delta_{pq}}{\partial x}V_{\mu\nu}^{A} \tag{A.6a}$$

$$= \sum_{\mu\nu\in A}\sum_{ia}(C_{\mu i}C_{\nu a}\Delta_{ia}^{x} + C_{\mu a}C_{\nu i}\Delta_{ai}^{x})V_{\mu\nu}^{A} \tag{A.6b}$$

$$= \mathbf{V}^{A}\cdot\mathbf{\Delta}^{x} + c.c. \tag{A.6c}$$

$$= \mathbf{V}^{A}\cdot\Big(-E^{\mathbf{\Delta}\mathbf{\Delta}^{-1}}E^{\mathbf{\Delta}x}\Big) + c.c. \tag{A.6d}$$

$$= \mathbf{z}\cdot E^{\mathbf{\Delta}x} + c.c. \tag{A.6e}$$

where $\mathbf{z}$ is the usual $\mathbf{z}$-vector technique here defined as:

$$E^{\mathbf{\Delta\Delta}}\mathbf{z} = -\mathbf{V}^A, \tag{A.7}$$

where $E^{\mathbf{\Delta\Delta}}$ is the usual orbital hessian. We solve the $\mathbf{z}$ with the conjugate gradient method.

Now we only need to evaluate $\mathbf{z} \cdot E^{\mathbf{\Delta}x}$:

$$\mathbf{z} \cdot E^{\mathbf{\Delta}x} = \mathbf{z} \cdot \frac{\partial}{\partial \mathbf{\Delta}}\big(\frac{\partial E}{\partial \mathbf{h}}\mathbf{h}^x + \frac{\partial E}{\partial \mathbf{II}}\mathbf{II}^x + \frac{\partial E}{\partial \mathbf{S}}\mathbf{S}^x\big) \tag{A.8a}$$

$$= \mathbf{P}_z \cdot \mathbf{h}^x + \mathbf{P}_z \cdot \mathbf{II}^x \cdot \mathbf{P} - \mathbf{P}_z\mathbf{F}\mathbf{P} \cdot \mathbf{S}^x - \mathbf{P}(\mathbf{P}_z \cdot \mathbf{II})\mathbf{P} \cdot \mathbf{S}^x - \mathbf{P}\mathbf{F}\mathbf{P}_z \cdot \mathbf{S}^x \tag{A.8b}$$

where

$$\mathbf{P}_z = \mathbf{z} \cdot \frac{\partial \mathbf{P}}{\partial \mathbf{\Delta}} \implies P_{z\mu\nu} = z_{ia}(C_{\mu i}C_{\nu a} + C_{\mu a}C_{\nu i}) \tag{A.9}$$

Putting it together: The classical electrostatics force due to a change in coordinate, x, of a nucleus on fragment A is:

$$\begin{aligned} E^x_{cls.elec} =& \frac{1}{2}\big(\mathbf{P}_z \cdot \mathbf{h}^x + \mathbf{P}_z \cdot \mathbf{II}^x \cdot \mathbf{P}\big) - \big(\mathbf{P}_z\mathbf{F}\mathbf{P} + \mathbf{P}(\mathbf{P}_z \cdot \mathbf{II})\mathbf{P} + \mathbf{P}\mathbf{F}\mathbf{P}_z + \mathbf{P}\mathbf{V}^A\mathbf{P}\big) \cdot \mathbf{S}^x \\ &+ \frac{1}{2}\mathbf{P}^A \cdot \mathbf{V}^{Ax} + \frac{1}{2}\sum_{B\neq A} V^{ABx}_{NN} \end{aligned} \tag{A.10}$$

A.2 Details on Internal Coordinates Transformation

Wilson's B matrix linearizes the internal to Cartesian coordinate transformation, i.e.,

$$\mathbf{B} = \frac{\partial \mathbf{q}}{\partial \mathbf{x}}. \tag{A.11}$$

The internal coordinates used here are the ones used in Q-Chem, which are taken from standard procedure in the literature. The procedure is

1. Connect every two atoms with distance $<$ the sum of their van der Waals radii.
2. Connect Hydrogen atoms to atoms $<$ hydrogen-bond distance away.
3. Connect every set of connected atoms (called fragment) to all other fragments, starting with the shortest distance first. This is done using Kruskal's algorithm.

We modify the procedure by applying it separately on each fragment (user-defined fragments, as is required for ALMO-EDA calculations). Then, we repeat the last step to connect all fragments minimally. This produces internal coordinates that are more useful for understanding fragment-fragment interaction.

Finally, we use the pseudo-inverse of Wilson's B matrix as follows,

$$\mathbf{F_q} = \mathbf{B}^+\mathbf{F_x}, \tag{A.12}$$

where $\mathbf{F_q}$ and $\mathbf{F_x}$ are the forces in internal coordinates and Cartesian coordinates, respectively while $\mathbf{B}^+$ is the pseudo-inverse of $\mathbf{B}$ using the Armadillo package.

A.3 MB-UCB vs. ALMO-EDA energy comparison

Below, we show the comparison of the energies for the same 50 water dimers used for the force comparison shown in the main paper. These MB-UCB energies show excellent accuracy as was demonstrated in the original MB-UCB paper [347].

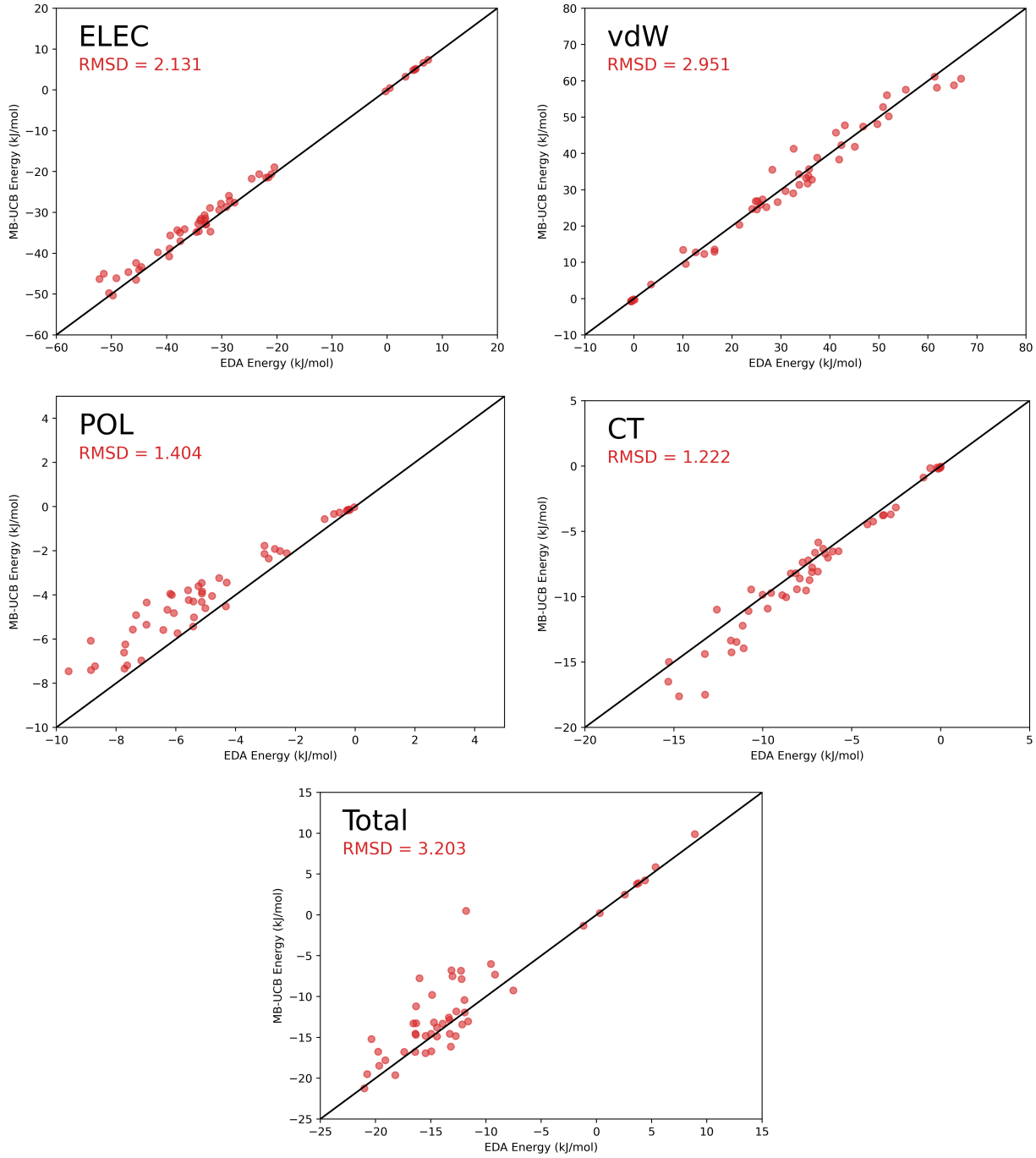


Figure A.1: Comparing the energies of MB-UCB with those of EDA shows very good agreement, as already established in the original MB-UCB paper.[347]

A.4 Forces in a smaller basis

The energy components of MB-UCB were compared against the second generation of ALMO-EDA [184], which uses a truncated virtual space for the description of polarization [373]. The truncated virtual space consists of fragment electric-field response functions (FERFs), which produces the correct basis set limit of the polarization energy, and thus the charge transfer energy. However, the polarization energy derivative using FERFs has not been implemented yet, and as a proxy to it, we repeat the procedure in a smaller basis, namely def2-TZVPD, as recommended in early EDA papers [183].

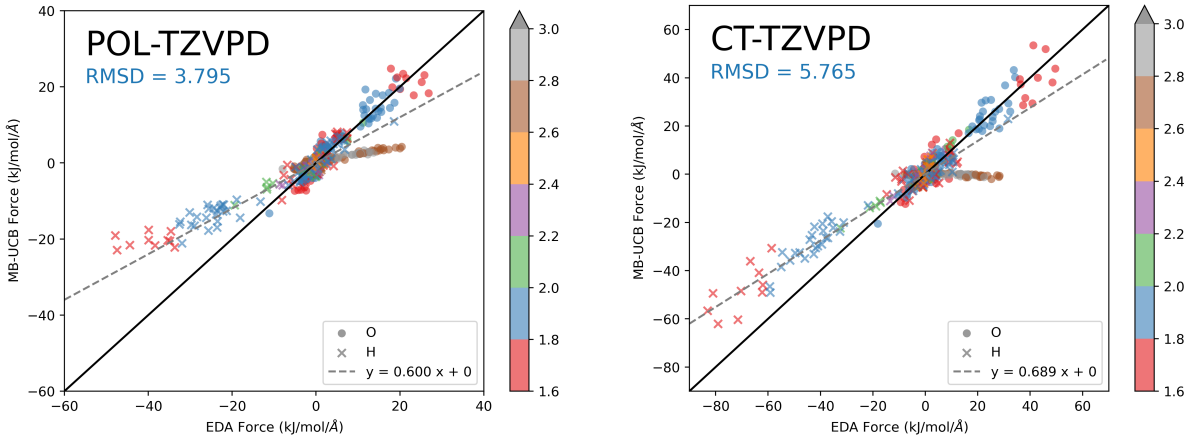


Figure A.2: Upon reducing the basis set used, MB-UCB forces perform significantly better. This may be because they were trained on ALMO-EDA-2.

A.5 Gold and silver complexes PES

Relative energies for these systems may depend quite a bit on the treatment of dynamic correlation, and we provide here a comparison of the performance of the density functional we use to CCSD(T). Here we do a constrained optimization using ωB98X-V, and then run HF, MP2, CCSD, and CCSD(T) on the same geometry for both gold and silver complexes. In the silver case, the restricted Hartree-Fock wavefunction is not a stable solution and we use the unrestricted solution instead. We see that the performance of ωB98X-V is on par with that of CCSD(T) for this system.

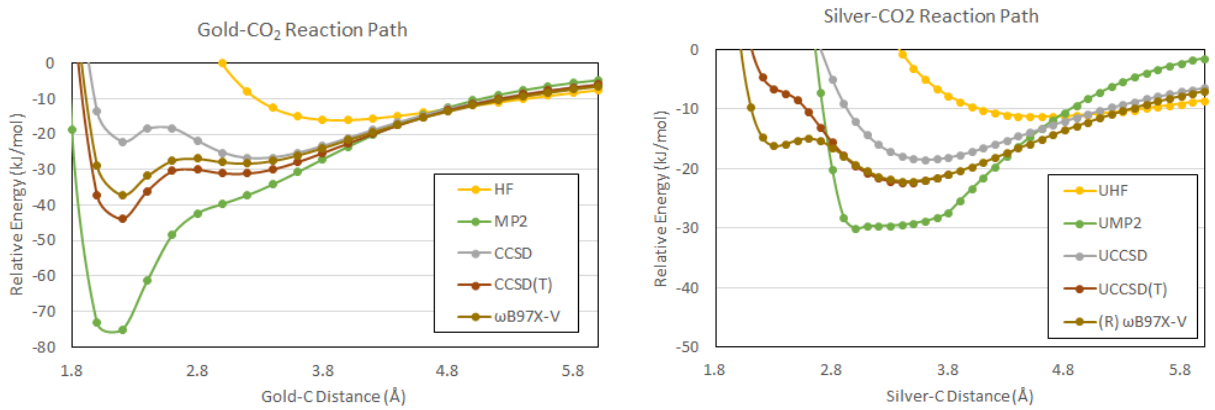


Figure A.3: Evaluating the energy on the intrinsic reaction coordinate using some wavefunction based methods to compare against our DFT answer.

Bibliography

(1) Dirac, P. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1929**, *123*, 714–733.

(2) Atkins, P. W.; Friedman, R. S., *Molecular Quantum Mechanics*; Molecular Quantum Mechanics v. 1; Oxford University Press: 1997.

(3) Baer, M., *Beyond Born-Oppenheimer*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.

(4) Szabo, A.; Ostlund, N. L., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: 1996.

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

(6) Jensen, F., *Introduction to Computational Chemistry*, Second; Wiley: 2017, p 624.

(7) Löwdin, P. O. Quantum theory of many-particle systems. III. Extension of the Hartree-Fock scheme to include degenerate systems and correlation effects. *Physical Review* **1955**, *97*, 1509–1520.

(8) Jensen, F. Atomic orbital basis sets. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 273–295.

(9) Mulliken, R. S. Criteria for the construction of good self-consistent-field molecular orbital wave functions, and the significance of LCAO-MO population analysis. *The Journal of Chemical Physics* **1962**, *36*, 3428–3439.

(10) Mulliken, R. S. Chemical bonding. *Annual review of physical chemistry* **1978**, *29*, 1–30.

(11) Helgaker, T.; Jørgensen, P.; Olsen, J., *Molecular Electronic-Structure Theory*; Wiley: 2000, pp 1–908.

(12) Helgaker, T.; Klopper, W.; Tew, D. P. Quantitative quantum chemistry. *Molecular Physics* **2008**, *106*, 2107–2143.

(13) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry* **2008**, *29*, 2044–2078.

(14) Lehtola, S. A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules. *International Journal of Quantum Chemistry* **2019**, *119*, e25968.

(15) Pople, J. A. Two-Dimensional Chart of Quantum Chemistry. *The Journal of Chemical Physics* **1965**, *43*, S229–S230.

(16) 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. *The Journal of Chemical Physics* **1989**, *90*, 5622–5629.

(17) Löwdin, P. O. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Physical Review* **1955**, *97*, 1474–1489.

(18) Slater, J. C. The theory of complex spectra. *Physical Review* **1929**, *34*, 1293–1322.

(19) Condon, E. U. The Theory of Complex Spectra. *Physical Review* **1930**, *36*, 1121–1133.

(20) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *The Journal of Chemical Physics* **1987**, *87*, 5968–5975.

(21) Scuseria, G. E.; Schaefer, H. F. Is coupled cluster singles and doubles (CCSD) more computationally intensive than quadratic configuration interaction (QCISD)? *The Journal of Chemical Physics* **1989**, *90*, 3700–3703.

(22) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical Review* **1934**, *46*, 618–622.

(23) Leininger, M. L.; Allen, W. D.; Schaefer, H. F.; Sherrill, C. D. Is Mo/ller–Plesset perturbation theory a convergent ab initio method? *The Journal of Chemical Physics* **2000**, *112*, 9213–9222.

(24) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. Scaled MP3 noncovalent interaction energies agree closely with accurate CCSD(T) benchmark data. *ChemPhysChem* **2009**, *10*, 282–289.

(25) Grimme, S. Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *Journal of Chemical Physics* **2003**, *118*, 9095–9102.

(26) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order moller-plesset correlation energy: An economical electronic structure method. *Journal of Chemical Physics* **2004**, *121*, 9793–9802.

(27) Surján, P. R.; Szabados, A. Damping of perturbation corrections in quasidegenerate situations. *Journal of Chemical Physics* **1996**, *104*, 3320–3324.

(28) Stück, D.; Head-Gordon, M. Regularized orbital-optimized second-order perturbation theory. *Journal of Chemical Physics* **2013**, *139*, 244109.

(29) 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. *Journal of Chemical Theory and Computation* **2018**, *14*, 5203–5219.

(30) Hylleraas, E. A. Über den Grundterm der Zweielektronenprobleme von H-, He, Li+, Be++ usw. *Zeitschrift für Physik* **1930**, *65*, 209–225.

(31) Hylleraas, E. A. In *Advances in Quantum Chemistry*; C; Academic Press: 1964; Vol. 1, pp 1–33.

(32) Sinanoğlu, O. Relation of Perturbation Theory to Variation Method. *The Journal of Chemical Physics* **1961**, *34*, 1237–1240.

(33) Pulay, P. Localizability of dynamic electron correlation. *Chemical Physics Letters* **1983**, *100*, 151–154.

(34) Saebø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annual Review of Physical Chemistry* **1993**, *44*, 213–236.

(35) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.

(36) Čižek, J.; Paldus, J. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methodst. *International Journal of Quantum Chemistry* **1971**, *5*, 359–379.

(37) Paldus, J.; Íek, J.; Shavitt, I. Correlation problems in atomic and molecular systems. IV. Extended coupled-pair many-electron theory and its application to the BH3 molecule. *Physical Review A* **1972**, *5*, 50–67.

(38) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.

(39) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics* **1980**, *48*, 157–173.

(40) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model. *Chemical Physics* **1982**, *71*, 41–49.

(41) Siegbahn, P. E.; Almlöf, J.; Heiberg, A.; Roos, B. O. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *The Journal of Chemical Physics* **1981**, *74*, 2384–2396.

(42) 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. *The Journal of Chemical Physics* **2007**, *126*, 164101.

(43) Sherrill, C. D.; Krylov, A. I.; Byrd, E. F. C.; Head-Gordon, M. Energies and analytic gradients for a coupled-cluster doubles model using variational Brueckner orbitals: Application to symmetry breaking in O4+. *The Journal of Chemical Physics* **1998**, *109*, 4171–4181.

(44) Krylov, A. I.; Sherrill, C. D.; Byrd, E. F. C.; Head-Gordon, M. Size-consistent wave functions for nondynamical correlation energy: The valence active space optimized orbital coupled-cluster doubles model. *The Journal of Chemical Physics* **1998**, *109*, 10669–10678.

(45) Sinanoğlu, O. Many-Electron Theory of Atoms and Molecules. I. Shells, Electron Pairs vs Many-Electron Correlations. *The Journal of Chemical Physics* **1962**, *36*, 706–717.

(46) Sinanoğlu, O. Many-Electron Theory of Atoms and Molecules. II. *The Journal of Chemical Physics* **1962**, *36*, 3198–3208.

(47) Nesbet, R. K. In *Advances in Chemical Physics*; John Wiley & Sons, Ltd: 1965, pp 321–363.

(48) Langhoff, S. R.; Davidson, E. R. Configuration interaction calculations on the nitrogen molecule. *International Journal of Quantum Chemistry* **1974**, *8*, 61–72.

(49) Pople, J. A.; Binkley, J. S.; Seeger, R. Theoretical models incorporating electron correlation. *International Journal of Quantum Chemistry* **1976**, *10*, 1–19.

(50) Meyer, W.; Rosmus, P. PNO–CI and CEPA studies of electron correlation effects. III. Spectroscopic constants and dipole moment functions for the ground states of the first-row and second-row diatomic hydrides. *The Journal of Chemical Physics* **1975**, *63*, 2356–2375.

(51) Galbraith, J. M.; Shaik, S.; Danovich, D.; Braïda, B.; Wu, W.; Hiberty, P.; Cooper, D. L.; Karadakov, P. B.; Dunning, T. H. Valence Bond and Molecular Orbital: Two Powerful Theories that Nicely Complement One Another. *Journal of Chemical Education* **2021**, *98*, 3617–3620.

(52) Cooper, D. L.; Gerratt, J.; Raimondi, M. In 2007, pp 319–397.

(53) Gerratt, J.; Cooper, D. L.; Karadakov, P. B.; Raimondi, M. Modern valence bond theory. *Chemical Society Reviews* **1997**, *26*, 87.

(54) Heitler, W.; London, F. Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik. *Zeitschrift für Physik* **1927**, *44*, 455–472.

(55) Coulson, C.; Fischer, I. XXXIV. Notes on the molecular orbital treatment of the hydrogen molecule. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1949**, *40*, 386–393.

(56) Hurley, A. C.; Lennard-Jones, J. E.; Pople, J. A. The molecular orbital theory of chemical valency XVI. A theory of paired-electrons in polyatomic molecules. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953**, *220*, 446–455.

(57) Goddard, W. A.; Dunning, T. H.; Hunt, W. J.; Hay, P. J. Generalized Valence Bond Description of Bonding in Low-Lying States of Molecules. *Accounts of Chemical Research* **1973**, *6*, 368–376.

(58) Gerratt, J.; Lipscomb, W. N. Spin-Coupled Wave Functions for Atoms and Molecules. *Proceedings of the National Academy of Sciences* **1968**, *59*, 332–335.

(59) Gerratt, J. In *Advances in Atomic and Molecular Physics*; C; Academic Press: 1971; Vol. 7, pp 141–221.

(60) Bobrowicz, F. W.; Goddard, W. A. In *Methods of Electronic Structure Theory*; Springer US: Boston, MA, 1977, pp 79–127.

(61) Cullen, J. Generalized valence bond solutions from a constrained coupled cluster method. *Chemical Physics* **1996**, *202*, 217–229.

(62) Van Voorhis, T.; Head-Gordon, M. Connections between coupled cluster and generalized valence bond theories. *Journal of Chemical Physics* **2001**, *115*, 7814–7821.

(63) Van Voorhis, T.; Head-Gordon, M. Implementation of generalized valence bond-inspired coupled cluster theories. *The Journal of Chemical Physics* **2002**, *117*, 9190–9201.

(64) Small, D. W.; Head-Gordon, M. Tractable spin-pure methods for bond breaking: Local many-electron spin-vector sets and an approximate valence bond model. *Journal of Chemical Physics* **2009**, *130*, 84103.

(65) Small, D. W.; Head-Gordon, M. Post-modern valence bond theory for strongly correlated electron spins. *Physical Chemistry Chemical Physics* **2011**, *13*, 19285.

(66) Dunning, T. H.; Xu, L. T.; Takeshita, T. Y.; Lindquist, B. A. Insights into the Electronic Structure of Molecules from Generalized Valence Bond Theory. *The Journal of Physical Chemistry A* **2016**, *120*, 1763–1778.

(67) Dunning, T. H.; Xu, L. T.; Cooper, D. L.; Karadakov, P. B. Spin-Coupled Generalized Valence Bond Theory: New Perspect i ves on the Electronic Structure of Molecules and Chemical Bonds. *The Journal of Physical Chemistry A* **2021**, *125*, 2021–2050.

(68) Sinanoğlu, O. In *Advances in Chemical Physics*; John Wiley & Sons, Ltd: 1964, pp 315–412.

(69) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. A simple DFT-based diagnostic for nondynamical correlation. *Theoretical Chemistry Accounts* **2013**, *132*, 1291.

(70) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871.

(71) Parr, R. G.; Weitao, Y., *Density-Functional Theory of Atoms and Molecules*; 1; Oxford University Press: 1995.

(72) Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *Journal of Physical Chemistry* **1996**, *100*, 12974–12980.

(73) Nagy, Á. Density functional. Theory and application to atoms and molecules. *Physics Report* **1998**, *298*, 1–79.

(74) Capelle, K. A bird's-eye view of density-functional theory. *Brazilian Journal of Physics* **2006**, *36*, 1318–1343.

(75) Burke, K.; Wagner, L. O. DFT in a nutshell. *International Journal of Quantum Chemistry* **2013**, *113*, 96–101.

(76) Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics* **2014**, *140*, 18A301.

(77) Thomas, L. H. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, *23*, 542–548.

(78) Fermi, E. Un Metodo Statistic per la Determinazione di Alcune Paioriete dell' Atome. *Rendicontia dell Accademia del Lincei* **1927**, *6*, 602–607.

(79) Dirac, P. A. Note on Exchange Phenomena in the Thomas Atom. *Mathematical Proceedings of the Cambridge Philosophical Society* **1930**, *26*, 376–385.

(80) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, *140*, A1133.

(81) Perdew, J. P. In *AIP Conference Proceedings*, AIP: 2001; Vol. 577, pp 1–20.

(82) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

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

(84) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, *38*, 3098–3100.

(85) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation. *Physical Review Letters* **1999**, *82*, 2544–2547.

(86) Kurth, S.; Perdew, J. P.; Blaha, P. Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs. *International Journal of Quantum Chemistry* **1999**, *75*, 889–909.

(87) Van Voorhis, T.; Scuseria, G. E. A novel form for the exchange-correlation energy functional. *The Journal of Chemical Physics* **1998**, *109*, 400–410.

(88) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

(89) Kim, M.-C.; Sim, E.; Burke, K. Understanding and Reducing Errors in Density Functional Calculations. *Physical Review Letters* **2013**, *111*, 073003.

(90) Kim, M.-C.; Sim, E.; Burke, K. Ions in solution: Density corrected density functional theory (DC-DFT). *The Journal of Chemical Physics* **2014**, *140*, 18A528.

(91) Vuckovic, S.; Song, S.; Kozlowski, J.; Sim, E.; Burke, K. Density Functional Analysis: The Theory of Density-Corrected DFT. *Journal of Chemical Theory and Computation* **2019**, *15*, 6636–6646.

(92) Song, S.; Vuckovic, S.; Sim, E.; Burke, K. Density-Corrected DFT Explained: Questions and Answers. *Journal of Chemical Theory and Computation* **2022**, *18*, 817–827.

(93) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics* **2001**, *115*, 3540–3544.

(94) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004**, *393*, 51–57.

(95) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics* **2017**, *115*, 2315–2372.

(96) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615.

(97) 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. *Physical Chemistry Chemical Physics* **2014**, *16*, 9904.

(98) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics* **2006**, *124*, 034108.

(99) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory. *Journal of Chemical Theory and Computation* **2014**, *10*, 3669–3680.

(100) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. Multiconfiguration Pair-Density Functional Theory: A New Way To Treat Strongly Correlated Systems. *Accounts of Chemical Research* **2017**, *50*, 66–73.

(101) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787–1799.

(102) 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. *Journal of Chemical Physics* **2010**, *132*, 154104.

(103) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

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

(105) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Physical Review Letters* **2004**, *92*, 246401.

(106) Vydrov, O. A.; Van Voorhis, T. Improving the accuracy of the nonlocal van der Waals density functional with minimal empiricism. *Journal of Chemical Physics* **2009**, *130*, 104105.

(107) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional Made Simple. *Physical Review Letters* **2009**, *103*, 063004.

(108) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *Journal of Chemical Physics* **2010**, *133*, 244103.

(109) 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 function. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

(110) Zhao, Y.; Truhlar, D. G. Applications and validations of the Minnesota density functionals. *Chemical Physics Letters* **2011**, *502*, 1–13.

(111) Kirkpatrick, J. et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* **2021**, *374*, 1385–1389.

(112) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density functional theory is straying from the path toward the exact functional. *Science* **2017**, *355*, 49–52.

(113) 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. *Journal of Chemical Theory and Computation* **2018**, *14*, 1969–1981.

(114) Coulson, C. A. Present State of Molecular Structure Calculations. *Reviews of Modern Physics* **1960**, *32*, 170–177.

(115) Neese, F.; Atanasov, M.; Bistoni, G.; Maganas, D.; Ye, S. Chemistry and Quantum Mechanics in 2019: Give Us Insight and Numbers. *Journal of the American Chemical Society* **2019**, *141*, 2814–2824.

(116) Mulliken, R. S. Electronic population analysis on LCAO-MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *The Journal of Chemical Physics* **1955**, *23*, 1841–1846.

(117) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics* **1950**, *18*, 365–375.

(118) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *The Journal of Chemical Physics* **1985**, *83*, 735–746.

(119) Mayer, I. Atomic Orbitals from Molecular Wave Functions: The Effective Minimal Basis. *The Journal of Physical Chemistry* **1996**, *100*, 6249–6257.

(120) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals. *Journal of Chemical Physics* **2004**, *120*, 2629–2637.

(121) Laikov, D. N. Intrinsic minimal atomic basis representation of molecular electronic wavefunctions. *International Journal of Quantum Chemistry* **2011**, *111*, 2851–2867.

(122) Knizia, G. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *Journal of Chemical Theory and Computation* **2013**, *9*, 4834–4843.

(123) Bader, R. F. W. Atoms in molecules. *Accounts of Chemical Research* **1985**, *18*, 9–15.

(124) Bader, R. F. W. A quantum theory of molecular structure and its applications. *Chemical Reviews* **1991**, *91*, 893–928.

(125) Becke, A. D., *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Matta, C. F., Boyd, R. J., Eds.; John Wiley & Sons.: 2007, pp 1–527.

(126) Cassam-Chenaï, P.; Jayatilaka, D. Some fundamental problems with zero flux partitioning of electron densities. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2001**, *105*, 213–218.

(127) Shahbazian, S. Why Bond Critical Points Are Not "Bond" Critical Points. *Chemistry - A European Journal* **2018**, *24*, 5401–5405.

(128) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry* **1990**, *11*, 361–373.

(129) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta* **1977**, *44*, 129–138.

(130) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *Journal of Chemical Physics* **2007**, *126*, 144111.

(131) Stone, A.; Alderton, M. Distributed multipole analysis. *Molecular Physics* **1985**, *56*, 1047–1064.

(132) Karen, P.; McArdle, P.; Takats, J. Toward a comprehensive definition of oxidation state (IUPAC Technical Report). *Pure and Applied Chemistry* **2014**, *86*, 1017–1081.

(133) Karen, P.; McArdle, P.; Takats, J. Comprehensive definition of oxidation state (IUPAC Recommendations 2016). *Pure and Applied Chemistry* **2016**, *88*, 831–839.

(134) Karen, P. Oxidation State, A Long-Standing Issue! *Angewandte Chemie International Edition* **2015**, *54*, 4716–4726.

(135) Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113.

(136) Krylov, A. I. From orbitals to observables and back. *The Journal of Chemical Physics* **2020**, *153*, 080901.

(137) Bartlett, R. J. Towards an exact correlated orbital theory for electrons. *Chemical Physics Letters* **2009**, *484*, 1–9.

(138) Huzinaga, S.; Arnau, C. Virtual orbitals in hartree-fock theory. *Physical Review A* **1970**, *1*, 1285–1288.

(139) Huzinaga, S.; Arnau, C. Virtual orbitals in hartree-fock theory. II. *The Journal of Chemical Physics* **1971**, *54*, 1948–1951.

(140) Perdew, J. P.; Levy, M. Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities. *Physical Review Letters* **1983**, *51*, 1884–1887.

(141) Janak, J. F. Proof that dE/dn_i=ϵ in density-functional theory. *Physical Review B* **1978**, *18*, 7165–7168.

(142) Boys, S. F. Construction of Molecular orbitals to be minimally variant for changes from one molecule to another. *Reviews of Modern Physics* **1960**, *32*, 296–299.

(143) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics* **1989**, *90*, 4916–4926.

(144) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Reviews of Modern Physics* **1963**, *35*, 457–464.

(145) Frenking, G.; Krapp, A. Unicorns in the world of chemical bonding models. *Journal of Computational Chemistry* **2007**, *28*, 15–24.

(146) Sano, T. Elementary Jacobi rotation method for generalized valence bond perfect-pairing calculations combined with simple procedure for generating reliable initial orbitals. *Journal of Molecular Structure: THEOCHEM* **2000**, *528*, 177–191.

(147) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L. L.; Knizia, G. Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals. *Journal of Chemical Theory and Computation* **2017**, *13*, 4063–4078.

(148) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *The Journal of Chemical Physics* **2005**, *123*, 114108.

(149) Schütz, M.; Yang, J.; Chan, G. K.-L.; Manby, F. R.; Werner, H.-J. The orbital-specific virtual local triples correction: OSV-L(T). *The Journal of Chemical Physics* **2013**, *138*, 054109.

(150) Meyer, W. Ionization energies of water from PNO-CI calculations. *International Journal of Quantum Chemistry* **1971**, *5*, 341–348.

(151) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *Journal of Chemical Physics* **2013**, *138*, 34106.

(152) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Reviews of Modern Physics* **1960**, *32*, 300–302.

(153) Heßelmann, A. Local Molecular Orbitals from a Projection onto Localized Centers. *Journal of Chemical Theory and Computation* **2016**, *12*, 2720–2741.

(154) Høyvik, I. M.; Olsen, J.; Jørgensen, P. Generalising localisation schemes of orthogonal orbitals to the localisation of non-orthogonal orbitals. *Molecular Physics* **2017**, *115*, 16–25.

(155) Weinstein, H.; Pauncz, R. Molecular orbital set determined by a localization procedure. *Symposia of the Faraday Society* **1968**, *2*, 23–31.

(156) Bauschlicher, C. W. The construction of modified virtual Orbitals (MVO's) which are suited for configuration interaction calculations. *The Journal of Chemical Physics* **1980**, *72*, 880–885.

(157) Boughton, J. W.; Pulay, P. Comparison of the boys and Pipek-Mezey localizations in the local correlation approach and automatic virtual basis selection. *Journal of Computational Chemistry* **1993**, *14*, 736–740.

(158) Aquilante, F.; Bondo Pedersen, T.; de Merás, A.; Koch, H.; Sánchez de Merás, A.; Koch, H. Fast noniterative orbital localization for large molecules. *The Journal of Chemical Physics* **2006**, *125*, 174101.

(159) Michalak, A.; Mitoraj, M.; Ziegler, T. Bond orbitals from chemical valence theory. *Journal of Physical Chemistry A* **2008**, *112*, 1933–1939.

(160) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital localization using fourth central moment minimization. *The Journal of Chemical Physics* **2012**, *137*, 224114.

(161) Lehtola, S.; Jónsson, H. Pipek–Mezey Orbital Localization Using Various Partial Charge Estimates. *Journal of Chemical Theory and Computation* **2014**, *10*, 642–649.

(162) Høyvik, I.-M. M.; Jørgensen, P. Characterization and generation of local occupied and virtual Hartree–Fock orbitals. *Chemical Reviews* **2016**, *116*, 3306–3327.

(163) Albaugh, A. et al. Advanced Potential Energy Surfaces for Molecular Simulation. *Journal of Physical Chemistry B* **2016**, *120*, 9811–9832.

(164) Sosso, G. C.; Chen, J.; Cox, S. J.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chemical Reviews* **2016**, *116*, 7078–7116.

(165) Gleiter, R.; Haberhauer, G.; Werz, D. B.; Rominger, F.; Bleiholder, C. From Noncovalent Chalcogen–Chalcogen Interactions to Supramolecular Aggregates: Experiments and Calculations. *Chemical Reviews* **2018**, *118*, 2010–2041.

(166) Ponder, J. W.; Case, D. A. In *Protein Simulations*; Advances in Protein Chemistry, Vol. 66; Academic Press: 2003, pp 27–85.

(167) Stone, A., *The Theory of Intermolecular Forces*; Oxford University Press: 2013.

(168) Scheiner, S., *Molecular interactions: from van der Waals to strongly bound complexes*; Wiley: 1997, p 359.

(169) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chemical Reviews* **1994**, *94*, 1887–1930.

(170) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 254–272.

(171) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 304–326.

(172) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *The Journal of Chemical Physics* **2014**, *140*, 094106.

(173) Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Computational Molecular Science* **2020**, *10*, DOI: 10.1002/wcms.1452.

(174) Jansen, G. Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 127–144.

(175) Hohenstein, E. G.; Sherrill, C. D. Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes. *The Journal of Chemical Physics* **2010**, *132*, 184111.

(176) Xie, W.; Gao, J. Design of a Next Generation Force Field: The X-POL Potential. *Journal of Chemical Theory and Computation* **2007**, *3*, 1890–1900.

(177) Casimir, H. B.; Polder, D. The influence of retardation on the London-van der Waals forces. *Physical Review* **1948**, *73*, DOI: 10.1103/PhysRev.73.360.

(178) Gould, T.; Toulouse, J.; Ángyán, J. G.; Dobson, J. F. Casimir–Polder Size Consistency: A Constraint Violated by Some Dispersion Theories. *Journal of Chemical Theory and Computation* **2017**, *13*, 5829–5833.

(179) Lamoreaux, S. K. Demonstration of the casimir force in the 0.6 to 6 μm range. *Physical Review Letters* **1997**, *78*, 5–8.

(180) Brezgunova, M. E.; Lieffrig, J.; Aubert, E.; Dahaoui, S.; Fertey, P.; Lebègue, S.; Ángyán, J. G.; Fourmigué, M.; Espinosa, E. Chalcogen bonding: Experimental and theoretical determinations from electron density analysis. geometrical preferences driven by electrophilic-nucleophilic interactions. *Crystal Growth and Design* **2013**, *13*, 3283–3289.

(181) Fanfrlík, J.; Přáda, A.; Padělková, Z.; Pecina, A.; Macháček, J.; Lepšík, M.; Holub, J.; Růžička, A.; Hnyk, D.; Hobza, P. The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics. *Angewandte Chemie International Edition* **2014**, *53*, 10139–10142.

(182) Deng, S.; Wang, Q.; Ren, P. Estimating and modeling charge transfer from the SAPT induction energy. *Journal of Computational Chemistry* **2017**, *38*, 2222–2231.

(183) 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. *The Journal of Physical Chemistry A* **2007**, *111*, 8753–8765.

(184) Horn, P. R.; Mao, Y.; Head-Gordon, M. Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals. *Physical Chemistry Chemical Physics* **2016**, *18*, 23067–23079.

(185) Mao, Y.; Loipersberger, M.; Horn, P. R.; Das, A.; Demerdash, O.; Levine, D. S.; Prasad Veccham, S.; Head-Gordon, T.; Head-Gordon, M. From Intermolecular Interaction Energies and Observable Shifts to Component Contributions and Back Again: A Tale of Variational Energy Decomposition Analysis. *Annual Review of Physical Chemistry* **2021**, *72*, 641–666.

(186) Aldossary, A.; Head-Gordon, M. Non-iterative method for constructing valence antibonding molecular orbitals and a molecule-adapted minimum basis. *The Journal of Chemical Physics* **2022**, *157*, 094102.

(187) Gimferrer*, M.; Aldossary*, A.; Salvador, P.; Head-Gordon, M. Oxidation State Localized Orbitals: A Method for Assigning Oxidation States Using Optimally Fragment-Localized Orbitals and a Fragment Orbital Localization Index. *Journal of Chemical Theory and Computation* **2022**, *18*, 309–322.

(188) Maslen, P. E.; Ochsenfeld, C.; White, C. A.; Lee, M. S.; Head-Gordon, M. Locality and sparsity of Ab initio one-particle density matrices and localized orbitals. *Journal of Physical Chemistry A* **1998**, *102*, 2215–2222.

(189) Schmidt, M. W.; Hull, E. A.; Windus, T. L. Valence Virtual Orbitals: An Unambiguous ab Initio Quantification of the LUMO Concept. *Journal of Physical Chemistry A* **2015**, *119*, 10408–10427.

(190) Reed, A. E.; Weinhold, F. Natural bond orbital analysis of near-Hartree-Fock water dimer. *The Journal of Chemical Physics* **1983**, *78*, 4066–4073.

(191) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor—Acceptor Viewpoint. *Chemical Reviews* **1988**, *88*, 899–926.

(192) Fukui, K. Role of frontier orbitals in chemical reactions. *Science* **1982**, *218*, 747–754.

(193) Fukui, K. The Role of Frontier Orbitals in Chemical Reactions (Nobel Lecture), 1982.

(194) Hoffmann, R. Building Bridges Between Inorganic and Organic Chemistry (Nobel Lecture), 1982.

(195) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Electron donation in the water-water hydrogen bond. *Chemistry - A European Journal* **2009**, *15*, 851–855.

(196) Hurley, A. C.; Lennard-Jones, J. E. The molecular orbital theory of chemical valency XIV. Paired electrons in the presence of two unlike attracting centres. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953**, *218*, 327–333.

(197) Lee, M. S.; Head-Gordon, M. Polarized atomic orbitals for self-consistent field electronic structure calculations. *Journal of Chemical Physics* **1997**, *107*, 9085–9095.

(198) Iwata, S. Valence type vacant orbitals for configuration interaction calculations. *Chemical Physics Letters* **1981**, *83*, 134–138.

(199) Derricotte, W. D.; Evangelista, F. A. Localized Intrinsic Valence Virtual Orbitals as a Tool for the Automatic Classification of Core Excited States. *Journal of Chemical Theory and Computation* **2017**, *13*, 5984–5999.

(200) Wang, Q.; Zou, J.; Xu, E.; Pulay, P.; Li, S. Automatic Construction of the Initial Orbitals for Efficient Generalized Valence Bond Calculations of Large Systems. *Journal of Chemical Theory and Computation* **2018**, *15*, 141–153.

(201) Lee, M. S.; Head-Gordon, M. Extracting Polarized Atomic Orbitals from Molecular Orbital Calculations. *International Journal of Quantum Chemistry* **2000**, *76*, 169–184.

(202) Hampel, C.; Werner, H. J. Local treatment of electron correlation in coupled cluster theory. *Journal of Chemical Physics* **1996**, *104*, 6286–6297.

(203) Schütz, M.; Hetzer, G.; Werner, H. J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *Journal of Chemical Physics* **1999**, *111*, 5691–5705.

(204) Schütz, M.; Werner, H. J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *Journal of Chemical Physics* **2001**, *114*, 661–681.

(205) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K.-L. L. Tensor factorizations of local second-order Møller–Plesset theory. *The Journal of Chemical Physics* **2011**, *134*, 044123.

(206) Kurashige, Y.; Yang, J.; Chan, G. K.-L.; Manby, F. R. Optimization of orbital-specific virtuals in local Mller-Plesset perturbation theory. *Journal of Chemical Physics* **2012**, *136*, 124106.

(207) Yang, J.; Chan, G. K.-L.; Manby, F. R.; Schütz, M.; Werner, H. J. The orbital-specific-virtual local coupled cluster singles and doubles method. *Journal of Chemical Physics* **2012**, *136*, 144105.

(208) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *The Journal of Chemical Physics* **2009**, *131*, 064103.

(209) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *The Journal of Chemical Physics* **2009**, *130*, 114108.

(210) Roos, B. O. In Roos2007, 2007, pp 399–445.

(211) Cooper, D. L.; Gerratt, J.; Raimondi, M. Spin-coupled valence bond theory. *International Reviews in Physical Chemistry* **1988**, *7*, 59–80.

(212) Cooper, D. L.; Gerratt, J.; Raimondi, M. Applications of Spin-Coupled Valence Bond Theory. *Chemical Reviews* **1991**, *91*, 929–964.

(213) Small, D. W.; Lawler, K. V.; Head-Gordon, M. Coupled cluster valence bond method: Efficient computer implementation and application to multiple bond dissociations and strong correlations in the acenes. *Journal of Chemical Theory and Computation* **2014**, *10*, 2027–2040.

(214) Lehtola, S.; Parkhill, J.; Head-Gordon, M. Orbital optimisation in the perfect pairing hierarchy: applications to full-valence calculations on linear polyacenes. *Molecular Physics* **2018**, *116*, 547–560.

(215) Stück, D.; Baker, T. A.; Zimmerman, P.; Kurlancheek, W.; Head-Gordon, M. On the nature of electron correlation in C60. *Journal of Chemical Physics* **2011**, *135*, 194306.

(216) Ibeji, C. U.; Ghosh, D. Singlet-triplet gaps in polyacenes: A delicate balance between dynamic and static correlations investigated by spin-flip methods. *Physical Chemistry Chemical Physics* **2015**, *17*, 9849–9856.

(217) Bender, C. F.; Davidson, E. R. Theoretical calculation of the potential curves of the Be2 molecule. *The Journal of Chemical Physics* **1967**, *47*, 4972–4978.

(218) Bonaccorsi, R.; Petrongolo, C.; Scrocco, E.; Tomasi, J. Configuration-interaction calculations for the ground state of OF2, NO2-, CN-: Canonical orbitals and exclusive orbitals. *Theoretica Chimica Acta* **1969**, *15*, 332–343.

(219) Wasilewski, J. Modified virtual orbitals (MVO) in limited CI calculations. *International Journal of Quantum Chemistry* **1991**, *39*, 649–656.

(220) Beran, G. J. O.; Austin, B.; Sodt, A.; Head-Gordon, M. Unrestricted Perfect Pairing: The Simplest Wave-Function-Based Model Chemistry beyond Mean Field. *The Journal of Physical Chemistry A* **2005**, *109*, 9183–9192.

(221) Lawler, K. V.; Small, D. W.; Head-Gordon, M. Orbitals that are unrestricted in active pairs for generalized valence bond coupled cluster methods. *Journal of Physical Chemistry A* **2010**, *114*, 2930–2938.

(222) Löwdin, P.-O. In *Advances in Chemical Physics*, Prigogine, I., Ed.; Advances in Chemical Physics, Hoboken, NJ, USA, 1958, pp 207–322.

(223) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chemical Physics Letters* **1988**, *153*, 503–506.

(224) Kapuy, E.; Csépes, Z.; Kozmutza, C. Application of the many-body perturbation theory by using localized orbitals. *International Journal of Quantum Chemistry* **1983**, *23*, 981–990.

(225) Kapuy, E.; Bartha, F.; Bogár, F.; Csépes, Z.; Kozmutza, C. Applications of the MBPT in the localized representation. *International Journal of Quantum Chemistry* **1990**, *38*, 139–147.

(226) Taube, A. G.; Bartlett, R. J. Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory. *Collection of Czechoslovak Chemical Communications* **2005**, *70*, 837–850.

(227) Taube, A. G.; Bartlett, R. J. Frozen natural orbital coupled-cluster theory: Forces and application to decomposition of nitroethane. *The Journal of Chemical Physics* **2008**, *128*, 164101.

(228) Salter, E. A.; Trucks, G. W.; Fitzgerald, G.; Bartlett, R. J. Theory and application of MBPT(3) gradients: The density approach. *Chemical Physics Letters* **1987**, *141*, 61–70.

(229) Trucks, G. W.; Salter, E. A.; Sosa, C.; Bartlett, R. J. Theory and implementation of the MBPT density matrix. An application to one-electron properties. *Chemical Physics Letters* **1988**, *147*, 359–366.

(230) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A direct MP2 gradient method. *Chemical Physics Letters* **1990**, *166*, 275–280.

(231) Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics* **1955**, *23*, 1833–1840.

(232) Davidson, E. R. Electronic population analysis of molecular wavefunctions. *The Journal of Chemical Physics* **1967**, *46*, 3320–3324.

(233) Roby, K. R. Quantum theory of chemical valence concepts I. Definition of the charge on an atom in a molecule and of occupation numbers for electron density shared between atoms. *Molecular Physics* **1974**, *27*, 81–104.

(234) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M. Are atoms intrinsic to molecular electronic wavefunctions? II. Analysis of fors orbitals. *Chemical Physics* **1982**, *71*, 51–64.

(235) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are atoms intrinsic to molecular electronic wavefunctions? III. Analysis of FORS configurations. *Chemical Physics* **1982**, *71*, 65–78.

(236) Mayer, I. Non-orthogonal localized orbitals and orthogonal atomic hybrids derived from Mulliken's population analysis. *Chemical Physics Letters* **1995**, *242*, 499–506.

(237) Epifanovsky, E. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *The Journal of Chemical Physics* **2021**, *155*, 084801.

(238) Liang, W.; Shao, Y.; Ochsenfeld, C.; Bell, A. T.; Head-Gordon, M. Fast evaluation of a linear number of local exchange matrices. *Chemical Physics Letters* **2002**, *358*, 43–50.

(239) Subotnik, J. E.; Shao, Y.; Liang, W.; Head-Gordon, M. An efficient method for calculating maxima of homogeneous functions of orthogonal matrices: Applications to localized occupied orbitals. *The Journal of Chemical Physics* **2004**, *121*, 9220–9229.

(240) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. Localized orbital theory and ammonia triborane. *Physical Chemistry Chemical Physics* **2007**, *9*, 5522.

(241) Pipek, J. Localization measure and maximum delocalization in molecular systems. *International Journal of Quantum Chemistry* **1989**, *36*, 487–501.

(242) Shavitt, I. In *Methods of Electronic Structure Theory*; Springer, Boston, MA: 1977, pp 189–275.

(243) Siegbahn, P. E. A new direct CI method for large CI expansions in a small orbital space. *Chemical Physics Letters* **1984**, *109*, 417–423.

(244) Knowles, P. J.; Handy, N. C. A new determinant-based full configuration interaction method. *Chemical Physics Letters* **1984**, *111*, 315–321.

(245) Levine, B. G.; Durden, A. S.; Esch, M. P.; Liang, F.; Shu, Y. CAS without SCF - Why to use CASCI and where to get the orbitals. *Journal of Chemical Physics* **2021**, *154*, 090902.

(246) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(247) 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. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

(248) Van Voorhis, T.; Head-Gordon, M. A geometric approach to direct minimization. *Molecular Physics* **2002**, *100*, 1713–1721.

(249) Hait, D.; Head-Gordon, M. Orbital Optimized Density Functional Theory for Electronic Excited States. *The Journal of Physical Chemistry Letters* **2021**, *12*, 4517–4529.

(250) Hait, D.; Head-Gordon, M. Excited State Orbital Optimization via Minimizing the Square of the Gradient: General Approach and Application to Singly and Doubly Excited States via Density Functional Theory. *Journal of Chemical Theory and Computation* **2020**, *16*, 1699–1710.

(251) Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. Molecular dynamics in low-spin excited states. *Journal of Chemical Physics* **1998**, *108*, 4060–4069.

(252) Filatov, M.; Shaik, S. A spin-restricted ensemble-referenced Kohn-Sham method and its application to diradicaloid situations. *Chemical Physics Letters* **1999**, *304*, 429–437.

(253) Rayez, M.-T.; Rayez, J.-C.; Sawerysyn, J.-P. Ab Initio Studies of the Reactions of Chlorine Atoms with Fluoro- and Chloro-Substituted Methanes. *The Journal of Physical Chemistry* **1994**, *98*, 11342–11352.

(254) Davis, S. R. Ab initio study of the insertion reaction of magnesium into the carbon-halogen bond of fluoro- and chloromethane. *Journal of the American Chemical Society* **1991**, *113*, 4145–4150.

(255) Fonseca Guerra, C.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD Methods for Charge Analysis. *Journal of Computational Chemistry* **2004**, *25*, 189–210.

(256) Sundin, S.; Saethre, L. J.; Sorensen, S. L.; Ausmees, A.; Svensson, S. Vibrational structure of the chloromethane series, CH4-nCln, studied by core photoelectron spectroscopy and ab initio calculations. *The Journal of Chemical Physics* **1999**, *110*, 5806–5813.

(257) Kokkonen, E.; Jänkälä, K.; Patanen, M.; Cao, W.; Hrast, M.; Bučar, K.; Žitnik, M.; Huttula, M. Role of ultrafast dissociation in the fragmentation of chlorinated methanes. *The Journal of Chemical Physics* **2018**, *148*, 174301.

(258) Senjean, B.; Sen, S.; Repisky, M.; Knizia, G.; Visscher, L. Generalization of Intrinsic Orbitals to Kramers-Paired Quaternion Spinors, Molecular Fragments, and Valence Virtual Spinors. *Journal of Chemical Theory and Computation* **2021**, *17*, 1337–1354.

(259) Hashimoto, K.; Osamura, Y.; Iwata, S. Ab initio study of structure and stability of beryllium compounds. *Journal of Molecular Structure: THEOCHEM* **1987**, *152*, 101–117.

(260) Yu, W.; Andrews, L.; Wang, X. Infrared Spectroscopic and Electronic Structure Investigations of Beryllium Halide Molecules, Cations, and Anions in Noble Gas Matrices. *The Journal of Physical Chemistry A* **2017**, *121*, 8843–8855.

(261) Coulson, C. d Electrons and Molecular Bonding. *Nature* **1969**, *221*, 1106–1110.

(262) Kiang, T.; Zare, R. N. Stepwise Bond Dissociation Energies in Sulfur Hexafluoride. *Journal of the American Chemical Society* **1980**, *102*, 4024–4029.

(263) Reed, A. E.; Weinhold, F. On the role of d orbitals in sulfur hexafluoride. *Journal of the American Chemical Society* **1986**, *108*, 3586–3593.

(264) Cooper, D. L.; Cunningham, T. P.; Gerratt, J.; Karadakov, P. B.; Raimondi, M. Chemical Bonding to Hypercoordinate Second-Row Atoms: d Orbital Participation versus Democracy. *Journal of the American Chemical Society* **1994**, *116*, 4414–4426.

(265) Kalemos, A. Hypervalent Bonding in the OF(a4Σ-), SF(a4Σ-), SF5/SF6, and OSF4 Species. *Journal of Physical Chemistry A* **2018**, *122*, 2178–2183.

(266) Jackson, B. A.; Harshman, J.; Miliordos, E. Addressing the Hypervalent Model: A Straightforward Explanation of Traditionally Hypervalent Molecules. *Journal of Chemical Education* **2020**, *97*, 3638–3646.

(267) Lyaskovskyy, V.; de Bruin, B. Redox Non-Innocent Ligands: Versatile New Tools to Control Catalytic Reactions. *ACS Catalysis* **2012**, *2*, 270–279.

(268) Minkin, V. I.; Glukhovtsev, M. N.; Simkin, B. Y., *Aromaticity and antiaromaticity*; John Wiley \& Sons, Incorporated: 1994.

(269) Balaban, A. T.; Oniciu, D. C.; Katritzky, A. R. Aromaticity as a Cornerstone of Heterocyclic Chemistry. *Chemical Reviews* **2004**, *104*, 2777–2812.

(270) Boldyrev, A. I.; Wang, L.-S. Beyond organic chemistry: aromaticity in atomic clusters. *Physical Chemistry Chemical Physics* **2016**, *18*, 11589–11605.

(271) Heitkemper, T.; Sarcevic, J.; Sindlinger, C. P. A Neutral Silicon(II) Half-Sandwich Compound. *Journal of the American Chemical Society* **2020**, *142*, 21304–21309.

(272) Min, X.; Popov, I. A.; Pan, F.-X.; Li, L.-J.; Matito, E.; Sun, Z.-M.; Wang, L.-S.; Boldyrev, A. I. All-Metal Antiaromaticity in Sb 4 -Type Lanthanocene Anions. *Angewandte Chemie International Edition* **2016**, *55*, 5531–5535.

(273) Panetier, J. A.; Letko, C. S.; Tilley, T. D.; Head-Gordon, M. Computational Characterization of Redox Non-Innocence in Cobalt-Bis(Diaryldithiolene)-Catalyzed Proton Reduction. *Journal of Chemical Theory and Computation* **2016**, *12*, 223–230.

(274) Raebiger, H.; Lany, S.; Zunger, A. Charge self-regulation upon changing the oxidation state of transition metals in insulators. *Nature* **2008**, *453*, 763–766.

(275) Resta, R. Charge states in transition. *Nature* **2008**, *453*, 735–735.

(276) Steen, J. S.; Knizia, G.; Klein, J. E. M. N. σ-Noninnocence: Masked Phenyl-Cation Transfer at Formal Ni IV. *Angewandte Chemie International Edition* **2019**, *58*, 13133–13139.

(277) Walsh, A.; Sokol, A. A.; Buckeridge, J.; Scanlon, D. O.; Catlow, C. R. A. Electron Counting in Solids: Oxidation States, Partial Charges, and Ionicity. *The Journal of Physical Chemistry Letters* **2017**, *8*, 2074–2075.

(278) Ampßler, T.; Monsch, G.; Popp, J.; Riggenmann, T.; Salvador, P.; Schröder, D.; Klüfers, P. Not Guilty on Every Count: The "Non-Innocent" Nitrosyl Ligand in the Framework of IUPAC's Oxidation-State Formalism. *Angewandte Chemie International Edition* **2020**, *59*, 12381–12386.

(279) Wang, G.; Zhou, M.; Goettel, J. T.; Schrobilgen, G. J.; Su, J.; Li, J.; Schlöder, T.; Riedel, S. Identification of an iridium-containing compound with a formal oxidation state of IX. *Nature* **2014**, *514*, 475–477.

(280) Thom, A. J. W.; Sundstrom, E. J.; Head-Gordon, M. LOBA: a localized orbital bonding analysis to calculate oxidation states, with application to a model water oxidation catalyst. *Physical Chemistry Chemical Physics* **2009**, *11*, 11297.

(281) Sit, P. H.-L.; Zipoli, F.; Chen, J.; Car, R.; Cohen, M. H.; Selloni, A. Oxidation State Changes and Electron Flow in Enzymatic Catalysis and Electrocatalysis through Wannier-Function Analysis. *Chemistry - A European Journal* **2011**, *17*, 12136–12143.

(282) Ramos-Cordoba, E.; Postils, V.; Salvador, P. Oxidation States from Wave Function Analysis. *Journal of Chemical Theory and Computation* **2015**, *11*, 1501–1508.

(283) Gimferrer, M.; Comas-Vilà, G.; Salvador, P. Can We Safely Obtain Formal Oxidation States from Centroids of Localized Orbitals? *Molecules* **2020**, *25*, 234.

(284) Catlow, C. R. A.; Stoneham, A. M. Ionicity in solids. *Journal of Physics C: Solid State Physics* **1983**, *16*, 4321–4338.

(285) Jiang, L.; Levchenko, S. V.; Rappe, A. M. Rigorous Definition of Oxidation States of Ions in Solids. *Physical Review Letters* **2012**, *108*, 166403.

(286) Walsh, A.; Sokol, A. A.; Buckeridge, J.; Scanlon, D. O.; Catlow, C. R. A. Oxidation states and ionicity. *Nature Materials* **2018**, *17*, 958–964.

(287) Mayer, I. Relation between the Hilbert space and "fuzzy atoms" analyses. *Chemical Physics Letters* **2013**, *585*, 198–200.

(288) Ramos-Cordoba, E.; Salvador, P.; Mayer, I. The atomic orbitals of the topological atom. *The Journal of Chemical Physics* **2013**, *138*, 214107.

(289) Postils, V.; Delgado-Alonso, C.; Luis, J. M.; Salvador, P. An Objective Alternative to IUPAC's Approach To Assign Oxidation States. *Angewandte Chemie International Edition* **2018**, *57*, 10525–10529.

(290) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Reviews of Modern Physics* **1960**, *32*, 296–299.

(291) Gimferrer, M.; Van der Mynsbrugge, J.; Bell, A. T.; Salvador, P.; Head-Gordon, M. Facing the Challenges of Borderline Oxidation State Assignments Using State-of-the-Art Computational Methods. *Inorganic Chemistry* **2020**, *59*, 15410–15420.

(292) Stoll, H.; Wagenblast, G.; Preuβ, H.; Preuß, H. On the use of local basis sets for localized molecular orbitals. *Theoretica Chimica Acta* **1980**, *57*, 169–178.

(293) Raimondi, M. Ab initio non-orthogonal approaches to the computation of weak interactions and of localised molecular orbitals for QM/MM procedures. *Journal of Molecular Structure: THEOCHEM* **2001**, *573*, 25–42.

(294) 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. *The Journal of Chemical Physics* **2001**, *115*, 3553–3560.

(295) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *The Journal of Chemical Physics* **2006**, *124*, 204105.

(296) Sax, A. F. Localization of molecular orbitals on fragments. *Journal of Computational Chemistry* **2012**, *33*, 1495–1510.

(297) De Silva, P.; Giebułtowski, M.; Korchowiec, J. Fast orbital localization scheme in molecular fragments resolution. *Phys. Chem. Chem. Phys.* **2012**, *14*, 546–552.

(298) 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. *The Journal of Chemical Physics* **2013**, *138*, 134119.

(299) Li, Z.; Li, H.; Suo, B.; Liu, W. Localization of Molecular Orbitals: From Fragments to Molecule. *Accounts of Chemical Research* **2014**, *47*, 2758–2767.

(300) Mo, Y.; Bao, P.; Gao, J. Energy decomposition analysis based on a block-localized wavefunction and multistate density functional theory. *Physical Chemistry Chemical Physics* **2011**, *13*, 6760.

(301) Giovannini, T.; Koch, H. Energy-Based Molecular Orbital Localization in a Specific Spatial Region. *Journal of Chemical Theory and Computation* **2021**, *17*, 139–150.

(302) Salvador, P.; Ramos-Cordoba, E. Communication: An approximation to Bader's topological atom. *The Journal of Chemical Physics* **2013**, *139*, 071103.

(303) Janowski, T. Near equivalence of intrinsic atomic orbitals and quasiatomic orbitals. *Journal of Chemical Theory and Computation* **2014**, *10*, 3085–3091.

(304) Salvador, P.; Ramos-Cordoba, E.; Gimferrer, M. APOST-3D, 2019.

(305) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of Chemical Physics* **1988**, *88*, 2547–2553.

(306) 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. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215.

(307) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *Journal of Chemical Theory and Computation* **2018**, *14*, 2596–2608.

(308) 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. *Journal of Chemical Theory and Computation* **2019**, *15*, 3610–3622.

(309) Hu, S.-X.; Li, W.-L.; Lu, J.-B.; Bao, J. L.; Yu, H. S.; Truhlar, D. G.; Gibson, J. K.; Marçalo, J.; Zhou, M.; Riedel, S.; Schwarz, W. H. E.; Li, J. On the Upper Limits of Oxidation States in Chemistry. *Angewandte Chemie International Edition* **2018**, *57*, 3242–3245.

(310) Davison, A.; Edelstein, N.; Holm, R. H.; Maki, A. H. The Preparation and Characterization of Four-Coordinate Complexes Related by Electron-Transfer Reactions. *Inorganic Chemistry* **1963**, *2*, 1227–1232.

(311) Lim, B. S.; Fomitchev, D. V.; Holm, R. H. Nickel Dithiolenes Revisited: Structures and Electron Distribution from Density Functional Theory for the Three-Member Electron-Transfer Series [Ni(S 2 C 2 Me 2) 2] 0,1 - ,2 -. *Inorganic Chemistry* **2001**, *40*, 4257–4262.

(312) Naumann, D.; Roy, T.; Tebbe, K.-F.; Crump, W. Synthesis and Structure of Surprisingly Stable Tetrakis(trifluoromethyl)cuprate(III) Salts. *Angewandte Chemie International Edition in English* **1993**, *32*, 1482–1483.

(313) Snyder, J. P. Elusiveness of CuIII Complexation; Preference for Trifluoromethyl Oxidation in the Formation of[CuI(CF3)4]- Salts. *Angewandte Chemie International Edition in English* **1995**, *34*, 80–81.

(314) Kaupp, M.; von Schnering, H. G. Formal Oxidation State versus Partial Charge—A Comment. *Angewandte Chemie International Edition in English* **1995**, *34*, 986–986.

(315) Aullón, G.; Alvarez, S. Oxidation states, atomic charges and orbital populations in transition metal complexes. *Theoretical Chemistry Accounts* **2009**, *123*, 67–73.

(316) Walroth, R. C.; Lukens, J. T.; MacMillan, S. N.; Finkelstein, K. D.; Lancaster, K. M. Spectroscopic Evidence for a 3d 10 Ground State Electronic Configuration and Ligand Field Inversion in [Cu(CF 3) 4] 1–. *Journal of the American Chemical Society* **2016**, *138*, 1922–1931.

(317) DiMucci, I. M.; Lukens, J. T.; Chatterjee, S.; Carsch, K. M.; Titus, C. J.; Lee, S. J.; Nordlund, D.; Betley, T. A.; MacMillan, S. N.; Lancaster, K. M. The Myth of d 8 Copper(III). *Journal of the American Chemical Society* **2019**, *141*, 18508–18520.

(318) Hoffmann, R.; Alvarez, S.; Mealli, C.; Falceto, A.; Cahill, T. J.; Zeng, T.; Manca, G. From Widely Accepted Concepts in Coordination Chemistry to Inverted Ligand Fields. *Chemical Reviews* **2016**, *116*, 8173–8192.

(319) Occhipinti, G.; Jensen, V. R. Nature of the Transition Metal–Carbene Bond in Grubbs Olefin Metathesis Catalysts. *Organometallics* **2011**, *30*, 3522–3529.

(320) Liu, S. Y.; Dykstra, C. E. A theory of vibrational transition frequency shifts due to hydrogen bonding. *Journal of Physical Chemistry* **1986**, *90*, 3097–3103.

(321) Huisken, F.; Kaloudis, M.; Vigasin, A. A. Vibrational frequency shifts caused by weak intermolecular interactions. *Chemical Physics Letters* **1997**, *269*, 235–243.

(322) Lalanne, P.; Tassaing, T.; Danten, Y.; Cansell, F.; Tucker, S. C.; Besnard, M. CO 2-Ethanol Interaction Studied by Vibrational Spectroscopy in Supercritical CO 2. *Journal of Physical Chemistry A* **2004**, *108*, 2617–2624.

(323) Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational stark effect. *Accounts of Chemical Research* **2015**, *48*, 998–1006.

(324) Wang, C.; Danovich, D.; Shaik, S.; Mo, Y. A Unified Theory for the Blue- and Red-Shifting Phenomena in Hydrogen and Halogen Bonds. *Journal of Chemical Theory and Computation* **2017**, *13*, 1626–1637.

(325) Mao, Y.; Head-Gordon, M. Probing Blue-Shifting Hydrogen Bonds with Adiabatic Energy Decomposition Analysis. *The Journal of Physical Chemistry Letters* **2019**, *10*, 3899–3905.

(326) Baiz, C. R. et al. Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction. *Chemical Reviews* **2020**, *120*, 7152–7218.

(327) Loew, L. M.; Simpson, L.; Hassner, A.; Alexanian, V. An unexpected blue shift caused by differential solvation of a chromophore oriented in a lipid bilayer. *Journal of the American Chemical Society* **1979**, *101*, 5439–5440.

(328) Fleming, G. R.; Cho, M. Chromophore-Solvent Dynamics. *Annual Review of Physical Chemistry* **1996**, *47*, 109–134.

(329) Pieper, J.; Rätsep, M.; Irrgang, K. D.; Freiberg, A. Chromophore-chromophore and chromophore-protein interactions in monomeric light-harvesting complex II of green plants studied by spectral hole burning and fluorescence line narrowing. *Journal of Physical Chemistry B* **2009**, *113*, 10870–10880.

(330) Rockwell, N. C.; Lagarias, J. C. A Brief History of Phytochromes. *ChemPhysChem* **2010**, *11*, 1172–1180.

(331) Kwon, M. S.; Lee, D.; Seo, S.; Jung, J.; Kim, J. Tailoring Intermolecular Interactions for Efficient Room-Temperature Phosphorescence from Purely Organic Materials in Amorphous Polymer Matrices. *Angewandte Chemie* **2014**, *126*, 11359–11363.

(332) González, R.; Mroginski, M. A. Fully Quantum Chemical Treatment of Chromophore-Protein Interactions in Phytochromes. *Journal of Physical Chemistry B* **2019**, *123*, 9819–9830.

(333) Chen, B.; Huang, W.; Su, H.; Miao, H.; Zhang, X.; Zhang, G. An Unexpected Chromophore–Solvent Reaction Leads to Bicomponent Aggregation-Induced Phosphorescence. *Angewandte Chemie - International Edition* **2020**, *59*, 10023–10026.

(334) Kramer, W. W.; McCrory, C. C. Polymer coordination promotes selective CO2 reduction by cobalt phthalocyanine. *Chemical Science* **2016**, *7*, 2506–2515.

(335) Nichols, E. M.; Derrick, J. S.; Nistanaki, S. K.; Smith, P. T.; Chang, C. J. Positional effects of second-sphere amide pendants on electrochemical CO 2 reduction catalyzed by iron porphyrins. *Chemical Science* **2018**, *9*, 2952–2960.

(336) Zhou, H.; Zou, X.; Wu, X.; Yang, X.; Li, J. Coordination Engineering in Cobalt-Nitrogen-Functionalized Materials for CO2 Reduction. *Journal of Physical Chemistry Letters* **2019**, *10*, 6551–6557.

(337) Chen, H.; Guo, X.; Kong, X.; Xing, Y.; Liu, Y.; Yu, B.; Li, Q. X.; Geng, Z.; Si, R.; Zeng, J. Tuning the coordination number of Fe single atoms for the efficient reduction of CO2. *Green Chemistry* **2020**, *22*, 7529–7536.

(338) Li, F. et al. Molecular tuning of CO2-to-ethylene conversion. *Nature* **2020**, *577*, 509–513.

(339) Wang, Y.; Liu, Y.; Liu, W.; Wu, J.; Li, Q.; Feng, Q.; Chen, Z.; Xiong, X.; Wang, D.; Lei, Y. Regulating the coordination structure of metal single atoms for efficient electrocatalytic CO 2 reduction. *Energy & Environmental Science* **2020**, *13*, 4609–4624.

(340) Ma, M.; Li, F.; Tang, Q. Coordination environment engineering on nickel single-atom catalysts for CO2electroreduction. *Nanoscale* **2021**, *13*, 19133–19143.

(341) Derrick, J. S.; Loipersberger, M.; Nistanaki, S. K.; Rothweiler, A. V.; Head-Gordon, M.; Nichols, E. M.; Chang, C. J. Templating Bicarbonate in the Second Coordination Sphere Enhances Electrochemical CO 2 Reduction Catalyzed by Iron Porphyrins. *Journal of the American Chemical Society* **2022**, *144*, 11656–11663.

(342) McDaniel, J. G.; Schmidt, J. R. Physically-motivated force fields from symmetry-adapted perturbation theory. *Journal of Physical Chemistry A* **2013**, *117*, 2053–2066.

(343) McDaniel, J. G.; Schmidt, J. Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory. *Annual Review of Physical Chemistry* **2016**, *67*, 467–488.

(344) Mao, Y.; Demerdash, O.; Head-Gordon, M.; Head-Gordon, T. Assessing Ion-Water Interactions in the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. *Journal of Chemical Theory and Computation* **2016**, *12*, 5422–5437.

(345) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P. P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371–394.

(346) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *Journal of Chemical Theory and Computation* **2019**, *15*, 4122–4139.

(347) 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. *Journal of Chemical Theory and Computation* **2019**, *15*, 5001–5013.

(348) Naseem-Khan, S.; Gresh, N.; Misquitta, A. J.; Piquemal, J. P. Assessment of SAPT and Supermolecular EDA Approaches for the Development of Separable and Polarizable Force Fields. *Journal of Chemical Theory and Computation* **2021**, *17*, 2759–2774.

(349) Das, A. K.; Liu, M.; Head-Gordon, T. Development of a Many-Body Force Field for Aqueous Alkali Metal and Halogen Ions: An Energy Decomposition Analysis Guided Approach. *Journal of Chemical Theory and Computation* **2022**, *18*, 953–967.

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

(351) Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy decomposition analysis. *WIREs Computational Molecular Science* **2018**, *8*, e1345.

(352) Su, P.; Tang, Z.; Wu, W. Generalized Kohn-Sham energy decomposition analysis and its applications. *WIREs Computational Molecular Science* **2020**, *10*, e1460.

(353) 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. *Physical Chemistry Chemical Physics* **2020**, *22*, 12867–12885.

(354) Mao, Y.; Loipersberger, M.; Kron, K. J.; Derrick, J. S.; Chang, C. J.; Sharada, S. M.; Head-Gordon, M. Consistent inclusion of continuum solvation in energy decomposition analysis: theory and application to molecular CO 2 reduction catalysts. *Chemical Science* **2021**, *12*, 1398–1414.

(355) Mao, Y.; Horn, P. R.; Head-Gordon, M. Energy decomposition analysis in an adiabatic picture. *Physical Chemistry Chemical Physics* **2017**, *19*, 5944–5958.

(356) Demerdash, O.; Mao, Y.; Liu, T.; Head-Gordon, M.; Head-Gordon, T. Assessing many-body contributions to intermolecular interactions of the AMOEBA force field using energy decomposition analysis of electronic structure calculations. *The Journal of Chemical Physics* **2017**, *147*, 161721.

(357) Ang, S. J.; Mak, A. M.; Wong, M. W. Nature of halogen bonding involving π-systems, nitroxide radicals and carbenes: a highlight of the importance of charge transfer. *Physical Chemistry Chemical Physics* **2018**, *20*, 26463–26478.

(358) Thirman, J.; Engelage, E.; Huber, S. M.; Head-Gordon, M. Characterizing the interplay of Pauli repulsion, electrostatics, dispersion and charge transfer in halogen bonding with energy decomposition analysis. *Physical Chemistry Chemical Physics* **2018**, *20*, 905–915.

(359) 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. *Physical Chemistry Chemical Physics* **2020**, *22*, 781–798.

(360) Loipersberger, M.; Mao, Y.; Head-Gordon, M. Variational Forward-Backward Charge Transfer Analysis Based on Absolutely Localized Molecular Orbitals: Energetics and Molecular Properties. *Journal of Chemical Theory and Computation* **2020**, *16*, 1073–1089.

(361) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.

(362) Chung, M. K. J.; Wang, Z.; Rackers, J. A.; Ponder, J. W. Classical Exchange Polarization: An Anisotropic Variable Polarizability Model. *The Journal of Physical Chemistry B* **2022**, DOI: `10.1021/acs.jpcb.2c04237`.

(363) 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. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.

(364) Wang, L.-P. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *The Journal of Physical Chemistry B* **2013**, *117*, 9956–9972.

(365) Huang, J.; Lopes, P. E.; Roux, B.; MacKerell, A. D. Recent advances in polarizable force fields for macromolecules: Microsecond simulations of proteins using the classical drude oscillator model. *Journal of Physical Chemistry Letters* **2014**, *5*, 3144–3150.

(366) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *The Journal of Physical Chemistry B* **2015**, *119*, 9423–9437.

(367) Inakollu, V. S.; Geerke, D. P.; Rowley, C. N.; Yu, H. Polarisable force fields: what do they add in biomolecular simulations? *Current Opinion in Structural Biology* **2020**, *61*, 182–190.

(368) Wilson, E. B. Some Mathematical Methods for the Study of Molecular Vibrations. *The Journal of Chemical Physics* **1941**, *9*, 76–84.

(369) Seki, T.; Chiang, K. Y.; Yu, C. C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M. The bending mode of water: A powerful probe for hydrogen bond structure of aqueous systems, 2020.

(370) Horn, P. R.; Head-Gordon, M. Alternative definitions of the frozen energy in energy decomposition analysis of density functional theory calculations. *Journal of Chemical Physics* **2016**, *144*, 84118.

(371) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. Extension of the SCF-MI Method to the Case of K Fragments one of which is an Open-Shell System. *Advances in Quantum Chemistry* **1998**, *31*, 251–266.

(372) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *Journal of Chemical Theory and Computation* **2013**, *9*, 452–460.

(373) Horn, P. R.; Head-Gordon, M. Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions. *Journal of Chemical Physics* **2015**, *143*, 114111.

(374) 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. *Journal of Chemical Physics* **2016**, *144*, 114107.

(375) Mao, Y.; Ge, Q.; Horn, P. R.; Head-Gordon, M. On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes. *Journal of Chemical Theory and Computation* **2018**, *14*, 2401–2417.

(376) Das, A. K.; Demerdash, O. N.; Head-Gordon, T. Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule. *Journal of Chemical Theory and Computation* **2018**, *14*, 6722–6733.

(377) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. Improved Formulas for the Calculation of the Electrostatic Contribution to the Intermolecular Interaction Energy from Multipolar Expansion of the Electronic Distribution. *The Journal of Physical Chemistry A* **2003**, *107*, 10353–10359.

(378) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981**, *59*, 341–350.

(379) Albaugh, A.; Demerdash, O.; Head-Gordon, T. An efficient and stable hybrid extended Lagrangian/self-consistent field scheme for solving classical mutual induction. *The Journal of Chemical Physics* **2015**, *143*, 174104.

(380) Albaugh, A.; Head-Gordon, T. A New Method for Treating Drude Polarization in Classical Molecular Simulation. *Journal of Chemical Theory and Computation* **2017**, *13*, 5207–5216.

(381) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society* **1992**, *114*, 7827–7843.

(382) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of Chemical Physics* **2010**, *133*, 134105.

(383) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theoretica Chimica Acta* **1990**, *77*, 123–141.

(384) Gill, P. M.; Johnson, B. G.; Pople, J. A. A standard grid for density functional calculations. *Chemical Physics Letters* **1993**, *209*, 506–512.

(385) Pulay, P.; Fogarasi, G. Geometry optimization in redundant internal coordinates. *The Journal of Chemical Physics* **1992**, *96*, 2856–2860.

(386) Bakken, V.; Helgaker, T. The efficient optimization of molecular geometries using redundant internal coordinates. *The Journal of Chemical Physics* **2002**, *117*, 9160–9174.

(387) Ramos-Cordoba, E.; Lambrecht, D. S.; Head-Gordon, M. Charge-transfer and the hydrogen bond: Spectroscopic and structural implications from electronic structure calculations. *Faraday Discussions* **2011**, *150*, 345.

(388) Weinhold, F. In *Advances in Protein Chemistry*; Elsevier: 2005; Vol. 72, pp 121–155.

(389) Weinhold, F.; Klein, R. A. What is a hydrogen bond? Mutually consistent theoretical and experimental criteria for characterizing H-bonding interactions. *Molecular Physics* **2012**, *110*, 565–579.

(390) Ronca, E.; Belpassi, L.; Tarantelli, F. A Quantitative View of Charge Transfer in the Hydrogen Bond: The Water Dimer Case. *ChemPhysChem* **2014**, *15*, 2682–2687.

(391) Stone, A. J. Natural Bond Orbitals and the Nature of the Hydrogen Bond. *The Journal of Physical Chemistry A* **2017**, *121*, 1531–1534.

(392) Wang, L.-P. P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *The Journal of Physical Chemistry Letters* **2014**, *5*, 1885–1891.

(393) Haghighatlari, M.; Li, J.; Heidar-Zadeh, F.; Liu, Y.; Guan, X.; Head-Gordon, T. Learning to Make Chemical Predictions: The Interplay of Feature Representation, Data, and Machine Learning Methods. *Chem* **2020**, *6*, 1527–1542.

(394) Haghighatlari, M.; Li, J.; Guan, X.; Zhang, O.; Das, A.; Stein, C. J.; Heidar-Zadeh, F.; Liu, M.; Head-Gordon, M.; Bertels, L.; Hao, H.; Leven, I.; Head-Gordon, T. NewtonNet: a Newtonian message passing network for deep learning of interatomic potentials and forces. *Digital Discovery* **2022**, *1*, 333–343.

(395) Akin-Ojo, O.; Song, Y.; Wang, F. Developing ab initio quality force fields from condensed phase quantum-mechanics/molecular-mechanics calculations through the adaptive force matching method. *Journal of Chemical Physics* **2008**, *129*, 064108.

(396) Jin, S.; Hao, Z.; Zhang, K.; Yan, Z.; Chen, J. Advances and Challenges for the Electrochemical Reduction of CO 2 to CO: From Fundamentals to Industrialization. *Angewandte Chemie International Edition* **2021**, *60*, 20627–20648.

(397) Freund, H.-J.; Meijer, G.; Scheffler, M.; Schlögl, R.; Wolf, M. CO Oxidation as a Prototypical Reaction for Heterogeneous Processes. *Angewandte Chemie International Edition* **2011**, *50*, 10064–10094.

(398) Kimble, M. L.; Castleman, A. W.; Mitrić, R.; Bürgel, C.; Bonacic-Koutecky, V. Reactivity of Atomic Gold Anions toward Oxygen and the Oxidation of CO: Experiment and Theory. *Journal of the American Chemical Society* **2004**, *126*, 2526–2535.

(399) Boese, A. D.; Schneider, H.; Glöß, A. N.; Weber, J. M. The infrared spectrum of Au -·CO 2. *Journal of Chemical Physics* **2005**, *122*, DOI: 10.1063/1.1875114.

(400) Knurr, B. J.; Weber, J. M. Solvent-Driven Reductive Activation of Carbon Dioxide by Gold Anions. *Journal of the American Chemical Society* **2012**, *134*, 18804–18808.

(401) Zhang, X.; Lim, E.; Kim, S. K.; Bowen, K. H. Photoelectron spectroscopic and computational study of (M-CO2)- anions, M = Cu, Ag, Au. *Journal of Chemical Physics* **2015**, *143*, DOI: 10.1063/1.4935061.

(402) Bilodeau, R. C.; Scheer, M.; Haugen, H. K. Infrared laser photodetachment of transition metal negative ions: studies on , , and. *Journal of Physics B: Atomic, Molecular and Optical Physics* **1998**, *31*, 3885–3891.

(403) Andersen, T.; Haugen, H. K.; Hotop, H. Binding Energies in Atomic Negative Ions: III. *Journal of Physical and Chemical Reference Data* **1999**, *28*, 1511–1533.

(404) 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. *Physical Chemistry Chemical Physics* **2021**, *23*, 928–943.