RANK REDUCTION APPROACHES IN ELECTRONIC STRUCTURE THEORY

A Thesis
Presented to
The Academic Faculty

by

Robert M. Parrish

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

Georgia Institute of Technology
August 2015

RANK REDUCTION APPROACHES IN ELECTRONIC STRUCTURE THEORY

Approved by:

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

Professor Edmond Chow
School of Computational Science and Engineering
Georgia Institute of Technology

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

Professor Jean-Luc Brédas
Solar and Photovoltaics Engineering Research Center
King Abdullah University of Science and Technology

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

Date Approved: 16 April 2015

To my grandfathers, Urban Parrish and Roger Whitmore, who taught me to always be a good man before trying to be anything else.

PREFACE

Electronic structure theory is simultaneously one of the most promising and one of the most frustrating areas of science. Through electronic structure theory and extensions thereof, we have the formal tools to describe any chemical phenomenon to essentially exact precision, if we could only solve the necessary equations. This last bit is, however, the sticky wicket - a truly remarkable realization is that one can write down the exact electronic Schrödinger equation and the exact molecular Hamiltonian in four lines, and yet many lifetimes have been spent looking for tractable and accurate solutions to this problem. The advent of the computer and extensive mathematical and numerical developments over the past eighty years have helped markedly, to the point that the predictions of theoretical chemistry can meet experimental accuracy for many problems. However, many important problems remain unsolved, particularly those involving large systems with multiple length- and time-scales and those with multiple physical processes contributing to the observable variables of the system. A classical example is the proposed use of theoretical chemistry in the problem of *in silico* drug design. Here, one must describe the complicated and delicate noncovalent interaction between the ligand and target with a high degree of accuracy, and then also account for the important and difficult of effects of conformational averaging and solvent interactions with a similar level of accuracy.

With some notable exceptions, the overall methodology of electronic structure theory is well developed, and most of what remains is to improve the computational efficiency and utility of the hierarchy of these methods. With regard to efficiency, much progress has been made in recent years in the use of rank reduction and locality approaches to reduce the computational burden of the simulation. These approaches

are particularly elegant in that they work by taking advantage of the physics underlying the chemical process, often yielding robust approximations with enormous tractability gains. With regard to utility, there has recently been renewed interest in extracting more chemically useful information from the results of extensive electronic structure computations. These approaches often provide robust partitions or explanations of chemical phenomena, and provide a compelling complement to chemical intuition. Together, these advances in efficiency and utility are vastly improving our ability to apply electronic structure to important real-world problems.

ACKNOWLEDGEMENTS

One of the most critical lessons I have learned in graduate school is the importance of having good people in my life, both professionally and personally. In this regard, I have been extremely fortunate, and am very glad to have this small opportunity to thank my friends and colleagues for all of their support.

Professor David Sherrill has been a major force in my life since I first met him. David found me as a rather dejected young engineering student, and suggested that maybe quantum chemistry would be a more rewarding calling for me. He was certainly correct in that recommendation - to me this is manifest in the great joy I take in going to work every morning. I earnestly love all of the far-reaching goals, elegant developments, and dirty numerical tricks of electronic structure theory, far beyond any professional success I have had in this area. For introducing me to my career, I will be forever grateful to David. Besides introducing me to the field, David has been a superb advisor. He has supported the difficult transition between engineering and chemistry far more generously than I deserved, given me unlimited support and freedom in my projects and endeavors, and has dispatched me to many neat places throughout the world to further my personal and professional goals. Moreover, David is just about the nicest guy there is, and I have thoroughly enjoyed the thousands of conversations we have had over the past few years. Thanks Doc, it has been a great six years.

I have also been very fortunate to have myriad excellent colleagues and co-authors during my time in graduate school. Among the very best of these is my old friend and mentor, Prof. Edward Hohenstein. Ed is blindingly brilliant, has a deviously clever mind, and absolutely does not stand BS. I attribute much of my success to just trying

to keep up with him. More than that, we have had some great times, from our days of back-to-back cubicles in David's lab, to weeks at Ed's cabin on Skyline Drive above Stanford, to seeing Ed marry his Joanna and welcome their daughter Joy. One thing I am not at all ashamed of is the number of pubs we've been asked to leave along the way (usually for fighting with each other over some aspect of the electron repulsion integral tensor)! Another good friend and mentor (and now my postdoc advisor) is Prof. Todd Martínez, who has taught me a uniquely fearless and colorful approach to research. I am very excited for some great years to come in Palo Alto, where I will do my best to distill my own research approach from a mix of David's serene and rigorous style and Todd's aggressive and intuitive style.

Further thanks goes to Dr. Justin Turney and Dr. Andy Simmonett, who have been my excellent C++ instructors. I am also grateful to Dr. Nicolas Schunck, my practicum advisor at Lawrence Livermore National Lab, who showed me what it was to be a successful and well-rounded young scientist. I have also been very lucky to have had the support and encouragement of many wonderful people affiliated with CCQC and Psi4 - Prof. Fritz Schaefer, Prof. Daniel Crawford, Prof. Justin Fermann, Prof. Rollin King, Prof. Johnny Galbraith, and Prof. Francesco Evangelista. To Dr. Michael Marshall and Dr. Matthew Kennedy, thanks for reminding me to have a life outside the integrals - you guys really made an impact on me. To my fellow Sherrill Group members, Dr. Lori Burns, Dr. Jerome Gonthier, Dr. Ryan Richard, Dr. Sahan Thanthiriwatte, Prof. Ugur Bozkaya, Prof. Eugene Deprince, Mr. Trent Parker, Mr. Brandon Bakr, Mr. D.J. Crowell, and Mr. Chris "VV10" Cook, my earnest thanks for putting up with this irascible engineer - I really enjoyed all of our discussions and efforts. To my REU friends Kevin Hannon and Andrew James - I sincerely enjoyed your company in the summer of 2013 and beyond, and wish you the best in your respective endeavors at Emory and Virginia Tech. Further thanks go to my many expert and enthusiastic coauthors, especially including Dr. John Sears,

Prof. Thomas Körzdörfer, Ms. Sara Kokkila, Prof. Lee-Ping Wang, and Mr. Daniel Smith.

I also want to acknowledge the efforts and guidance of my thesis committee, consisting of Prof. David Sherrill, Prof. Ken Brown, Prof. Jean-Luc Brédas, Prof. Angelo Bongiorno, and Prof. Edmond Chow. In particular, Ken has always been up for a Cypress run to discuss the dizzying array of topics that run through the mind of a former Hertz Fellow, and has always held me in far higher regard than I deserve. Jean-Luc has supported my efforts academically and financially, leading to a number of interesting publications from our work in non-empirically tuned long-range corrected functionals with Dr. John Sears and Prof. Thomas Körzdörfer. Moreover, he once had the excellent foresight to hire absolutely the most dazzling and marvelous summer undergraduate that has ever been, leading to my headlong pursuit of the Lady Monika. Finally, I owe sincere thanks to Dr. Cameron Tyson and Prof. David Collard of Georgia Tech, who encouraged and guided me through my initial matriculation to chemistry.

Financial support for my Ph.D. has been provided by the Department of Energy's Computational Science Fellowship (Grant DE-FG02-97ER25308), by the Georgia Tech Presidential Fellowship, and by the Penny and E. Roe Stamps IV Leadership Scholarship through the Georgia Tech President's Scholarship program. The combined financial support of these three programs sums to a staggering amount. I can only hope that my work over the next few decades is adequate recompense for the generosity shown to me by these foundations.

I have also been fortunate to have the support and encouragement of many people outside of the academic realm. To my dear friends Steven DeBoer, Jenny Jacobs, Jaydeep Srimani, and Miller Templeton, thanks for keeping me sane and happy these past few years. To my parents, Jeanine and Jack Parrish, my sincere thanks for bringing me up to always be smart and pursue my dreams. Dad, I always wanted

to be a Hurricane Hunter like you, but I am afraid I got a little distracted with the Schrödinger Equation and never properly made it to the Navier-Stokes Equations. Mom, thank you above all else for teaching me to be kind - it has made all the difference in how I see the world. Mark and Maggie, thanks for being good foils and never letting me get too uppity with the academics. I know you both are destined for great things. And finally, to my lady Monika Williams, who has been my constant companion and friend across eight hundred miles of separation. I have never felt so alive as I have in any of our dozens of adventures and misadventures, and am ecstatic to be starting a new life with you in California and beyond.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Quantum chemistry is plagued by the presence of high-rank quantities, stemming from the N-body nature of the electronic Schrödinger equation. These high-rank quantities present a significant mathematical and computational barrier to the computation of chemical observables, and also drastically complicate the pedagogical understanding of important interactions between particles in a molecular system. The application of physically-motivated rank reduction approaches can help address these to problems. This thesis details recent efforts to apply rank reduction techniques in both of these arenas.

With regards to computational tractability, the representation of the $1/r_{12}$ Coulomb repulsion between electrons is a critical stage in the solution of the electronic Schrödinger equation. Typically, this interaction is encapsulated via the order-4 electron repulsion integral (ERI) tensor, which is a major bottleneck in terms of generation, manipulation, and storage. Many rank reduction techniques for the ERI tensor have been proposed to ameliorate this bottleneck, most notably including the order-3 density fitting (DF) and pseudospectral (PS) representations. Here we detail a new and uniquely powerful factorization - tensor hypercontraction (THC). THC decomposes the ERI tensor as a product of five order-2 matrices (the first wholly order-2 compression proposed for the ERI) and offers great flexibility for low-scaling algorithms for the manipulations of the ERI tensor underlying electronic structure theory. THC is shown to be physically-motivated, quite accurate, and uniquely efficient for some of the most difficult operations encountered in modern quantum chemistry.

On the front of chemical understanding of electronic structure theory, we present our recent work in developing robust two-body partitions for *ab initio* computations

of intermolecular interactions. Noncovalent interactions are the critical and delicate forces which govern such important processes as drug-protein docking, enzyme function, crystal packing, and zeolite adsorption. These forces arise as weak residual interactions leftover after the binding of electrons and nuclei into molecule, and, as such, are extremely difficult to accurately quantify or systematically understand. Symmetry-adapted perturbation theory (SAPT) provides an excellent approach to rigorously compute the interaction energy in terms of the physically-motivated components of electrostatics, exchange, induction, and dispersion. For small intermolecular dimers, this breakdown provides great insight into the nature of noncovalent interactions. However, SAPT abstracts away considerable details about the N-body interactions between particles on the two monomers which give rise to the interaction energy components. In the work presented herein, we step back slightly and extract an effective 2-body interaction for each of the N-body SAPT terms, rather than immediately tracing all the way down to the order-0 interaction energy. This effective order-2 representation of the order-N SAPT interaction allows for the robust assignment of interaction energy contributions to pairs of atoms or functional groups (the A-SAPT or F-SAPT partitions), allowing one to discuss the interaction in terms of atom- or functional-group-pairwise interactions. These A-SAPT and F-SAPT partitions can provide deep insight into the origins of complicated noncovalent interactions, e.g., by clearly shedding light on the long-contested question of the nature of the substituent effect in substituted sandwich benzene dimers.

CHAPTER I

INTRODUCTION TO TENSOR HYPERCONTRACTION

1.1 The Electron Repulsion Integral

1.1.1 Origins

This section serves to explain where and how the electron repulsion integral tensor arises in the standard approach to electronic structure theory.

The major technical goal in electronic structure theory is the solution of the electronic Schrödinger equation, e.g., in its time-independent form,[107,264,265,33,290,109]

$$\hat{H}|\Psi_n\rangle = E_n|\Psi_n\rangle, \tag{1}$$

together with the constraints of wavefunction orthonormality,

$$\langle\Psi_m|\Psi_n\rangle = \delta_{mn}, \tag{2}$$

and fermionic antisymmetry,

$$|\Psi_n(\ldots,\vec{x}_i,\ldots,\vec{x}_j,\ldots)\rangle = -|\Psi_n(\ldots,\vec{x}_j,\ldots,\vec{x}_i,\ldots)\rangle. \tag{3}$$

In the above, $|\Psi_n\rangle$ is the electronic wavefunction for state n and E_n is the corresponding electronic energy. Each electron carries coordinates of position and spin $\vec{x} \equiv < \vec{r}, s >$. $\hat{H}$ is the electronic Hamiltonian,

$$\hat{H} \equiv -\sum_i \frac{1}{2}\nabla^2 - \sum_{i,A}\frac{Z_A}{r_{iA}} + \sum_{i>j}\frac{1}{r_{ij}}. \tag{4}$$

Here i and j label electrons, while A labels nuclei. The first term corresponds to the kinetic energy of the electrons, the second to the electron-nuclear attraction, and the third to the electron-electron repulsion. The scalar Z_A is the charge of the A-th nucleus. Note the use of atomic units here and throughout.[290]

A major source of complexity in electronic structure theory is the electron-electron repulsion term, which provides a coupling element between pairs of electrons. This, in turn, implies that the wavefunction $|\Psi_n\rangle$ is an true many-body quantity, with nonzero correlations between any 2-, 3-, 4-, etc-tuples of electrons. Working directly with N-body wavefunctions is prohibitively difficult (both computationally and mathematically). Therefore, an immediate formal simplification is usually invoked to treat the N-body wavefunction as a linear combination of antisymmetrized direct products of a set of orthonormal one-electron wavefunctions $\{\psi_p(\vec{x})\}$ (often referred to as "spin-orbitals," assumed to be real unless otherwise noted),

$$|\Psi_n\rangle \equiv \sum_{\vec{T}=\{\ldots p\ldots q\ldots\}} C_{\vec{T}}|\Phi_{\vec{T}}\rangle. \tag{5}$$

Here $\vec{T} = \{\ldots p\ldots q\ldots\}$ is an occupation number vector, determining the combination of N spin orbitals to occupy, and the sum runs over all possible combinations. $C_{\vec{T}}$ is the amplitude of the $\vec{T}$-th antisymemtrized direct-product wavefunction. The antisymmetrized direct-product wavefunctions or "Slater determinants" are,[279]

$$|\Phi_{\vec{T}}\rangle \equiv \frac{1}{\sqrt{N!}}\mathcal{A}^{(N)}\left[\ldots\psi_p(\vec{x}_p)\ldots\psi_q(\vec{x}_q)\ldots\right]. \tag{6}$$

Here $\mathcal{A}^{(N)}$ is an N-body antisymmetrizer, which generates all possible permutations of N electrons in N spin orbitals, additionally with a parity factor of -1 applied for every pairwise swap.

The space of all possible N-electron Slater determinants for K spin orbitals is called a "Fock space," (technically, an N-electron Fock subspace) and thus this general parametrization of the wavefunction is often termed a "Fock-space approach."[109] Note that each Slater determinant is fully antisymmetric, and thus the full wavefunction will automatically obey the antisymmetry property. However, aside from fermionic correlation, the particles in a single Slater determinant are uncorrelated. Thus, the Coulomb correlation induced by the electron-electron repulsion term in the

Hamiltonian is built in by taking a weighted sum over all Slater determinants. In principle, this Fock-space approach is formally exact, if a complete set of one-particle spin orbitals $\{\psi_p(\vec{x})\}$ is used to build the Fock space, and if all Slater determinants in the Fock space are retained in the parametrization of the wavefunction. In practice, a finite number of spin orbitals must be used for computational tractability (typically accomplished via a finite-basis expansion). Moreover, in the overwhelming majority of cases, the number of Slater determinants included in the Fock-space expansion must be restricted [restricted configuration interaction (CI) approaches],[274,271] or else the coefficients of the Slater determinants must be approximated in some advantageous way [restricted coupled cluster (CC) approaches].[45,270] The truncation of the Fock-space expansion is absolutely critical for computational tractability, as the dimension of the Fock space grows as $K!/(N!(K-N)!)$, i.e., factorially in the system size.

Note that there do exist alternative methods to parametrize the wavefunction in terms of 2-, 3-, etc-particle objects.[131,165,300,296,305] These "geminal expansion" approaches offer compellingly fast convergence to the exact wavefunction, but are often enormously expensive.

In Fock-space approaches, a critical step is to develop expressions for the representation of the Hamiltonian operator in the basis of Slater determinants,

$$H_{\vec{T},\vec{U}} \equiv \langle \Phi_{\vec{T}} | \hat{H} | \Phi_{\vec{U}} \rangle. \tag{7}$$

The quantity $H_{\vec{T},\vec{U}}$ is a Hermitian matrix, which may be directly diagonalized to yield the electronic energies and coefficients of the Slater determinants which build up the full wavefunctions. This general type of approach is referred to as "configuration interaction" or CI. A remarkable realization is that the matrix elements $H_{\vec{T},\vec{U}}$, while seemingly N-body expectation values over the operator $\hat{H}$, can always be written in terms of one- and two-electron quantities. The reason for this is that the Hamiltonian is (at most) a two-particle operator, so the operation of $\hat{H}$ on the Slater determinant

$|\Phi_{\vec{U}}\rangle$ can always be written as a linear combination of N-electron Slater determinants differing from $|\Phi_{\vec{U}}\rangle$ by at most two spin orbitals. For a nonzero matrix element $H_{\vec{T},\vec{U}}$, the bra $\langle\Phi_{\vec{T}}|$ must overlap with some contribution from the ket $\hat{H}|\Phi_{\vec{U}}\rangle$. Therefore, since the Slater determinants are orthonormal basis functions of the Fock space, nonzero matrix elements $H_{\vec{T},\vec{U}}$ exist only for determinants with occupations $\vec{T}$ and $\vec{U}$ which differ by at most two spin orbitals.

The Slater-Condon rules[279,49] provide explicit formulae for the nonzero matrix elements. For identical Slater determinants:

$$H_{\vec{T},\vec{U}} = \sum_{p \in \vec{T}} \langle p|\hat{h}|p\rangle + \frac{1}{2} \sum_{p,q \in \vec{T}} \langle pq||pq\rangle. \tag{8}$$

For Slater determinants differing by one spin orbital, e.g., $\vec{T} = \{\ldots m \ldots\}$ and $\vec{U} = \{\ldots n \ldots\}$,

$$H_{\vec{T},\vec{U}} = \langle m|\hat{h}|n\rangle + \sum_{p \in \vec{T}} \langle mp||np\rangle. \tag{9}$$

For Slater determinants differing by two spin orbitals, e.g., $\vec{T} = \{\ldots mp \ldots\}$ and $\vec{U} = \{\ldots nq \ldots\}$,

$$H_{\vec{T},\vec{U}} = \langle mp||nq\rangle. \tag{10}$$

Note that for these rules to work, the Slater determinants must be first permuted to be in maximum coincidence, keeping track of parity factors as this is done.

The one-electron matrix elements are the one-electron integrals,

$$\langle p|\hat{h}|q\rangle \equiv \int_{\mathbb{R}^4} \mathrm{d}^4x_1 \; \psi_p(\vec{x}_1) \left[-\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{r_{1A}}\right] \psi_q(\vec{x}_1), \tag{11}$$

and the two-electron matrix elements are the antisymmetrized electron repulsion integrals,

$$\langle pq||rs\rangle = \langle pq|rs\rangle - \langle pq|sr\rangle. \tag{12}$$

The electron repulsion integrals are written in terms of the physicist's electron repulsion integrals over spin orbitals,

$$\langle pq|rs\rangle \equiv \iint_{\mathbb{R}^8} \mathrm{d}^4x_1 \; \mathrm{d}^4x_2 \; \psi_p(\vec{x}_1)\psi_q(\vec{x}_2)\frac{1}{r_{12}}\psi_r(\vec{x}_1)\psi_s(\vec{x}_2), \tag{13}$$

which are alternatively expressed in terms of the chemist's electron repulsion integrals over spin orbitals,

$$[pq|rs] \equiv \iint_{\mathbb{R}^8} \mathrm{d}^4x_1 \; \mathrm{d}^4x_2 \; \psi_p(\vec{x}_1)\psi_q(\vec{x}_1)\frac{1}{r_{12}}\psi_r(\vec{x}_2)\psi_s(\vec{x}_2) = \langle pr|qs\rangle. \tag{14}$$

The one-electron integrals form an order-2 tensor, which is relatively easy to generate and manipulate. The electron repulsion integrals form an order-4 tensor, which is much harder to deal with. Therefore, much of the lore of Fock-space approaches is devoted to mitigating or amortizing the complexity and storage of the electron repulsion integral tensor.

1.1.2 Definition

The electron repulsion integral (ERI), written in the chemist's notation in spatial form, is,

$$(pq|rs) \equiv \iint_{\mathbb{R}^6} \mathrm{d}^3r_1 \; \mathrm{d}^3r_2 \; \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)\frac{1}{r_{12}}\phi_r(\vec{r}_2)\phi_s(\vec{r}_2). \tag{15}$$

For a one-particle spatial basis set $\{\phi_p(\vec{r})\}$, this encapsulates the electrostatic interaction of a particle in coordinate $\vec{r}_1$ represented by the product of basis functions $\rho_{pq}(\vec{r}_1) = \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)$ with another particle in coordinate $\vec{r}_2$ represented by the product of basis functions $\rho_{rs}(\vec{r}_2) = \phi_r(\vec{r}_2)\phi_s(\vec{r}_2)$. In the limit that the basis functions approach Dirac delta functions, the charge distributions of the two electrons approach point charges, e.g., $\rho_{pq}(\vec{r}_1) \to Q_P\delta(\vec{r}_1 - \vec{r}_P)\delta_{pq}$, and the classical limit results,

$$(pq|rs) \to \delta_{pq}\frac{Q_PQ_Q}{r_{PQ}}\delta_{rs}. \tag{16}$$

Here we have generalized the one-particle spatial basis $\{\phi_p(\vec{r})\}$ slightly from the orthonormal one-particle spin orbitals $\{\psi_p(\vec{x})\}$ in the Fock-space discussion above. In the present discussions, a one-particle basis set $\{\phi_p(\vec{r})\}$ could refer to a non-orthogonal atomic-orbital basis, the spatial part of an orthonormal one-particle spin-orbital basis (e.g., Hartree-Fock orbitals), any subset thereof, or really any other one-particle

quantity. Unless explicitly stated, we will consider only the case of an exponentially decaying atom-centered basis set (e.g., Gaussian or Slater functions in finite systems).

1.1.2.1 Properties

The ERI is an order-4 tensor providing the spectral representation or image of the symmetric-positive definite $1/r_{12}$ Coulomb operator in the basis of direct products of basis functions $\{\rho_{pq}(\vec{r}) \equiv \phi_p(\vec{r})\phi_q(\vec{r})\}$ (the "pair space"). The properties of both the Coulomb operator and the pair space lead to a number of interesting and important properties in the ERI tensor.

The position-space form of the Coulomb operator $G \equiv 1/r_{12}$ is the spatial representation of the Green's function of the Poisson equation on $\mathbb{R}^3$,

$$-\frac{1}{4\pi}\nabla^2 V(\vec{r}_1) = \rho(\vec{r}_1). \tag{17}$$

The Poisson equation is an elliptic, linear, second-order partial differential equation which relates the charge density $\rho(\vec{r})$ to the corresponding electrostatic potential (ESP) $V(\vec{r})$. As the Green's function, the Coulomb operator is the solution of the Poisson equation for a Dirac delta source,

$$-\frac{1}{4\pi}\nabla^2 G(\vec{r}_1, \vec{r}_2) = \delta(\vec{r}_1 - \vec{r}_2). \tag{18}$$

This is easily solved by Fourier transformation to reciprocal space, in which the problem is algebraic, and then back-transformation to position space, yielding the expected solution $G(\vec{r}_1, \vec{r}_2) = 1/r_{12}$.

The Coulomb operator is a local, finite-ranged, symmetric positive definite, two-particle operator. Locality implies that the operator depends only on the span of the pair space, and cannot see the details of the individual basis functions in each basis function product. An equivalent way to define this property is to say that the Coulomb integrals reduce to the 6-space integral above, rather than the more-general

12-space integral from a generic nonlocal operator,

$$\langle pr|\hat{o}|qs\rangle \equiv \iint_{\mathbb{R}^{12}} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \mathrm{d}^3 r_1' \; \mathrm{d}^3 r_2' \; \phi_p(\vec{r}_1)\phi_r(\vec{r}_2)\hat{o}(\vec{r}_1,\vec{r}_2,\vec{r}_1',\vec{r}_2')\phi_q(\vec{r}_1')\phi_s(\vec{r}_2'). \quad (19)$$

The property of locality will prove invaluable in the development of reduced rank representations for the ERI for two reasons. One is that locality allows one to exploit rank redundancy in the pair space, paving the way for density fitting and related approaches. Another is that locality allows for multiplicative separability of the basis functions in position space, eventually yielding the pseudospectral and tensor hypercontraction representations.

The finite-ranged property of the Coulomb operator, which distinguishes it from zero-range operators, such as the Skyrme force[278] or Slater exchange model[280], is actually a prize understatement. The Coulomb operator is essentially the worst-case scenario in terms of its spectral behavior. The operator exhibits a nasty singularity as $r_{12} \to 0^+$, and is markedly long-ranged as $r_{12} \to \infty$. The harmonic decay at long range implies that the operator may never be accurately truncated at some suitably large value of r_{12}. The singularity at the origin correspondingly implies that the operator will have slow convergence in reciprocal space, due to the high frequency transients as $r_{12} \to 0^+$. The singularity is a further source of complexity if grid-based resolutions of the ERI are attempted in position space.

The symmetric positive definite (SPD) property of the Coulomb operator is related to the indistinguishability of particles (symmetry) and to the principle that like charges repel (positivity). This property will be carried through in several forms in the ERI tensor, and is an important property to maintain in approximate resolutions of the ERI.

The pair space is composed of the span of direct products of the orbital basis functions, i.e., $\{\rho_{pq}(\vec{r}) \equiv \phi_p(\vec{r})\phi_q(\vec{r})\}$. In principle, a complete orbital basis $\phi_p(\vec{r})$ would span the complete square-integrable Hilbert space in three-space, e.g., $L_2(\mathbb{R}^3)$, and thus the pair space would span the space of integrable functions in three-space,

e.g., $L_1(\mathbb{R}^3)$. In practice, a finite set of smooth functions is usually chosen for the orbital basis, and thus the pair space also lives in the space of smooth functions. Naïvely, the pair space basis scales quadratically with the size of the orbital basis (which itself scales linearly with the size of the system). However, for atom-centered basis sets, the pair space is often markedly numerically redundant, implying that many pairs or combinations of pairs are not needed for accurate representation of the electronic structure problem. This is particularly the case for large molecules and/or large basis sets. There are two reasons for this. The first the spatial sparsity of the rapidly-decaying atom-centered basis functions implies that asymptotically only a linear number of basis function pairs are nonvanishing, particularly in large molecules. The second is that many of the nonvanishing products generate the same class of shapes in the pair space, particularly in large basis sets. For instance, the same-center product of an S and a D orbital gives a pair-space function which looks like a D function, just as a product of two P functions also looks like a D function. In practice, these two considerations jointly imply that the naïvely quadratic scaling pair-space basis can be accurately replaced by a linear-scaling auxiliary representation.

For real orbitals, the ERI tensor obeys the following 8-fold permutational symmetry,

$$\begin{aligned}(pq|rs) = (pq|sr) = (qp|rs) = (qp|sr) \\ = (rs|pq) = (rs|qp) = (sr|pq) = (sr|qp)\end{aligned} \tag{20}$$

This symmetry may be used to avoid computing or storing redundant integrals, though the logic of where each integral contributes yields some additional code complexity.

The ERI tensor is SPD with respect to the matrix of chemist's integrals,

$$D_{pq}(pq|rs)D_{rs} > 0 \;\forall\; \mathbf{D} \tag{21}$$

and also SPD with respect to the matrix of physicist's integrals,

$$D_{pq}\langle pq|rs\rangle D_{rs} > 0 \; \forall \; \mathbf{D} \tag{22}$$

Note the use of the generalized Einstein summation convention for tensor contractions here and throughout. The SPD property of the chemist's integrals corresponds to the principle that like charges repel. It is critical to maintain the SPD property in the chemist's integrals or approximations thereof, particularly in variational procedures such as self-consistent field (SCF) theory. Failure to do so can lead to situations where electrons *attract* each other within certain parts of the pair space, and may collapse toward the smallest wavefunction possible within the span of the finite basis employed. This "variational collapse" problem is a major barrier to certain classes of numerical approximation, such as local flavors of density fitting.[194] The SPD property of the physicist's integrals appears to be less practically important, and many of the most popular ERI approximations do not obey this property.

An important practical consideration is the rapid growth of the ERI tensor with respect to the system size. In general, the ERI tensor scales as $\mathcal{O}(N^4)$ where N is some measure of the system size (typically the primary basis size N_p). Even with 8-fold permutational symmetry built in, this quickly becomes a major storage bottleneck. Using double precision (64-bit) integrals, at 100 basis functions, the ERI tensor consumes roughly 100 MB, and can safely be assumed to reside in core memory. By 400 orbitals, the ERI tensor consumes 26 GB, and is approaching the limits of core memory on all but the heaviest of nodes. By 1000 basis functions, the ERI tensor consumes 1 TB, and is approaching the limits of disk storage on most heavy nodes. Note that the largest production-level jobs typically encountered in modern quantum chemistry research may have 5000-10000 basis functions. For the largest of these, the ERI tensor would consume 10 PB of storage. Clearly, algorithms or schemes to attenuate or eliminate the storage of the ERI tensor are necessary to perform these computations.

1.1.3 Sparsity Considerations

In finite basis atom centered basis sets, such as those composed of Gaussian or Slater functions, the exponential decay of the basis functions can often be practically exploited to accelerate computations, e.g., by means of the Cauchy-Schwarz inequality,[105]

$$|(pq|rs)| \leq \sqrt{(pq|pq)(rs|rs)} \tag{23}$$

As the centers of the basis functions p and q (or r and s) are pulled apart, their pair distribution $\rho_{pq}(\vec{r})$ will decay exponentially, extinguishing the integral $(pq|rs)$. The Cauchy-Schwarz inequality can be used to provide an upper bound for the integral's magnitude under this decay, and can be used to determine if an integral need be computed at all. For example, many codes utilize a user-specified integral cutoff parameter δ, which is often in the range of 10^{-8} to 10^{-12} Hartree. If the Cauchy-Schwarz estimate for an integral is below δ, the integral is not computed, and is effectively clamped to zero. In practice, this sieving must be done by basis shell (all basis functions of differing angular momentum generated from the same radial Gaussian or Slater function) rather than basis function, to preserve rotational invariance.

In tight, small basis sets such as 6-31G*[102] or cc-pVDZ,[70] Cauchy-Schwarz sieving works reasonably well, and yields an asymptotically quadratic number of integrals that must be computed. However, the use of augmented or large basis sets quickly overwhelms the pair space with pairs containing one or two diffuse functions. These pairs decay much more slowly with distance than in tight, small basis sets, markedly decreasing the utility of Cauchy-Schwarz sieving. Moreover, Cauchy-Schwarz sieving depends strongly on the geometry of the molecular system. Linear systems will exhibit rapid onset of the quadratic regime, while globular systems will persist in the quartic region for much longer.

Other problems with Cauchy-Schwarz sieving exist. One is that the method estimates many integrals to be near δ, but they are then clamped all the way to *zero*,

potentially ignoring the buildup of many small contributions to a nontrivial total. An additional concern (which is inherent in any cutoff-based method) is that discontinuities in potential surfaces are certain to exist.

Additional considerations may augment or accelerate the application of Cauchy-Schwarz sieving. In density-based screening, the Cauchy-Schwarz bound is applied in concert with the magnitudes of elements of the density matrix or other quantity that the ERI is multiplied by to generate an upper bound of the ERI-vector product contribution.[105] These contributions may be neglected if they fall below a product threshold γ, which is usually of roughly the same order as δ. Density based sieving is particularly useful when the state vector is sparse and couples at least one basis function from each particle in the ERI (an "exchange-type" contraction), as the sparsity of the state vector will overcome the slow $1/r_{12}$ decay of the ERI, often leading to asymptotically linear scaling.

Note that the Cauchy-Schwarz or density-based sieves do not factor any information about the coordinate r_{12} into their estimates, and thus do not produce tight upper bounds. As discussed previously, $1/r_{12}$ is extraordinarily long-ranged, and it is not usually considered numerically acceptable to screen based on distance in r_{12} with anything but the smallest of thresholds. However, in some cases, the observed decay of the relevant ERIs is faster than $1/r_{12}$. For example, if a theory needs only ERIs which excite from one subspace to another, orthogonal subspace (e.g., the hole-particle excitations in some correlated theories), the relevant part of the product space carries only dipolar and higher products. This implies that the ERIs will decay as at least $1/r_{12}^3$, and might be profitably screened based on distance in r_{12}. The family of multipole-bound integral estimate (MBIE) methods[167] provide tighter bounds than Cauchy-Schwarz sieving to take advantage of this decay.

In practice, sieving methods are reasonably easy to implement, and can produce marked gains in efficiency for large systems with small, tight basis sets. The price

one must pay for these gains is that much effort must be placed on producing efficient codes for the evaluation of the electron repulsion integrals (or common products thereof). This is a major mathematical and computer science problem, and has led to myriad schemes for the evaluation of the needed integrals. E.g., for Gaussian basis sets, the McMurchie-Davidson,[193] Obara-Saika,[210] and Rys[258] approaches are but three of the most popular of many possible schemes. In all of these schemes, it typically costs tens to thousands of floating point operations (FLOPs) to evaluate a given integral, and it has proven quite difficult to make these routines run at the peak performance available on a given machine.

1.1.4 Existing Factorizations of the ERI

A critical realization that has slowly and continuously been occurring over the past fifty years is that it is never the ERIs that are needed, but rather, their effective contractions with state vectors such as density matrices and wavefunction amplitudes. For instance, in Hartree-Fock (HF) self consistent field (SCF) theory, the rate-limiting steps are the construction of the Coulomb matrix,[290]

$$J_{pq} \equiv (pq|rs)D_{rs}, \tag{24}$$

and the exchange matrix,

$$K_{pq} \equiv (pr|qs)D_{rs}. \tag{25}$$

Here D_{rs} is the SCF one-particle density matrix (OPDM). To form these contractions, the most straightforward approach is to build the necessary ERIs analytically and then contract with the relevant parts of the OPDM. This is the basis for integral-direct (ID) SCF methods,[5,105] which rely heavily on Cauchy-Schwarz and density-based sieving for efficiency, among other tricks.

An alternative approach is to directly build the J_{pq} and K_{pq} contractions in some effective way. One technique which has enjoyed considerable success in recent years is the idea of "early contraction" in McMurchie-Davidson and other ERI generation

schemes, in which the contraction of the ERIs with the density matrix is built into the integral evaluation routines themselves, often saving considerably in terms of FLOPs.[321] However, this only reduces the prefactor of ID-SCF approaches, and does not affect the scaling or general reliance on ERI evaluation codes.

Another extremely interesting general approach is to approximate rewrite the ERI tensor as a product of several low-rank factor tensors. This provides several possible advantages:

1. The generation and storage of the low-rank factor tensors may be markedly easier than the generation and storage of the full ERI tensor.

2. The composition of the ERI tensor in terms of factor tensors may allow for the reduced-scaling construction of an ERI-state vector product in terms of several low rank intermediates.

3. The computational effort is shifted to tensor contraction operations, which are markedly efficient and scalable on modern computer hardware (e.g., through the use of `GEMM` calls using well-developed `BLAS` implementations[170]).

Factorization approaches do carry a number of possible disadvantages:

1. In molecular problems with atom-centered basis sets, the factorization of the ERI tensor is necessarily a numerical approximation, which must be carefully controlled to ensure that the desired accuracy is retained in the chemical observables being computed.

2. Factorization approaches often introduce additional complexity into quantum chemistry codes, both in terms of the formation of the factor tensors and in their use to form effective ERI-state vector products.

For canonical example, consider the (not so) hypothetical case of an accurate

factorization of the ERI with the topology,

$$(pq|rs) \approx b^C_{pq} b^C_{rs}. \tag{26}$$

That is, the order-4 chemist's integrals $(pq|rs)$ are factored as an outer product of order-3 factor tensors b^C_{pq}. With this representation, we can reduce the complexity of the Coulomb contraction from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^3)$ through the judicious use of a small intermediate,

$$J_{pq} \approx \left[b^C_{pq} \underbrace{\left[b^C_{rs} D_{rs} \right]}_{d^C} \right]. \tag{27}$$

The exchange contraction cannot be factored, due to the "pinned" appearance of the p and q indices in the factor tensor b^C_{pq}. However, we can take advantage of the low-rank formulation of the OPDM in SCF theory in terms of the occupied orbitals,

$$D_{pq} = C_{pi} C_{qi}, \tag{28}$$

where i is the occupied orbital index (much smaller than the AO-basis index p). Together with the low-rank ERI tensor, this gives,

$$K_{pq} \approx \underbrace{\left[b^C_{pr} C_{ri} \right]}_{E^C_{pi}} \underbrace{\left[b^C_{qs} C_{si} \right]}_{E^C_{qi}}. \tag{29}$$

This scales as a low-prefactor $\mathcal{O}(N^4)$, and is formulated entirely in terms of efficient `GEMM` operations. This example, together with the "density fitting" (DF) approach to form the factor tensors b^C_{pq}, forms the basis for density-fitted SCF (DF-SCF).[312] For small- to medium-sized molecules in large basis sets, DF-SCF is among the most efficient approaches to SCF theory.

What remains is to develop an algorithm for the accurate and efficient factorization of the ERI tensor. Many approaches to this problem exist, with varying topologies of factorization possible. The best approaches are deemed "physically-motivated rank reduction approaches" and operate by developing a rapidly converging approximate

factorization of the ERI tensor based on observations of the physics of the ERI itself. That is, these approaches do not attempt to brute-force a factorization, but rather work with the existing structure of the ERI to let a natural rank sparsity drive the factorization.

1.1.4.1 Density Fitting

One of the most successful rank reduction approaches for the ERI tensor is density fitting (DF).[324,67,68,156,76,299,245,144,312] DF asserts that the quadratic number of generalized densities in the pair space $\rho_{pq}(\vec{r}) \equiv \phi_p(\vec{r})\phi_q(\vec{r})$ can be accurately represented in a linear number of auxiliary functions $\chi_A(\vec{r})$,

$$\rho_{pq}(\vec{r}) \approx d^A_{pq}\chi_A(\vec{r}). \tag{30}$$

Here, d^A_{pq} is the problem-specific tensor of density fitting coefficients. In this representation, the ERI tensor may be decomposed as,

$$(pq|rs) \approx d^A_{pq}(A|B)d^B_{rs}, \tag{31}$$

where $(A|B)$ is a two-index ERI between auxiliary functions,

$$(A|B) \equiv \iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \ \mathrm{d}^3 r_2 \ \chi_A(\vec{r}_1)\frac{1}{r_{12}}\chi_B(\vec{r}_1). \tag{32}$$

The density fitting coefficients are typically solved by minimizing the square of the residual in the generalized densities with respect to an error metric $o(\vec{r}_{12})$,

$$O[d^A_{pq}] \equiv \iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \ \mathrm{d}^3 r_2 \ \left[\rho_{pq}(\vec{r}_1) - \underbrace{\tilde{\rho}_{pq}(\vec{r}_1)}_{d^A_{pq}\chi_A(\vec{r}_1)}\right] o(r_{12}) \left[\rho_{pq}(\vec{r}_2) - \underbrace{\tilde{\rho}_{pq}(\vec{r}_2)}_{d^A_{pq}\chi_A(\vec{r}_2)}\right]. \tag{33}$$

Many choices for the error metric have been proposed: $\delta(\vec{r}_{12})$ (the overlap metric),[324] $1/r_{12}$ (the Coulomb metric),[67] $\mathrm{erfc}(\omega r_{12})/r_{12}$ (the complementary Ewald metric),[142] etc. By far the most popular is the Coulomb metric, which eventually leads to the density fitting factorization,

$$(pq|rs) \approx (pq|A)(A|B)^{-1}(B|rs), \tag{34}$$

where $^{-1}$ indicates a matrix inverse, and $(pq|A)$ are the three-index ERIs,

$$(pq|A) \equiv \iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)\frac{1}{r_{12}}\chi_A(\vec{r}_2). \tag{35}$$

Defining the symmetrically-fitted three-index tensor,

$$b^C_{pq} \equiv (pq|A)(A|C)^{-1/2}, \tag{36}$$

this leads to the symmetric DF factorization,

$$(pq|rs) \approx b^C_{pq} b^C_{rs}. \tag{37}$$

This is precisely the factorization used in the discussion of Coulomb and exchange matrices above, and inherits all of the corresponding merits and limitations.

Note that for general density fitting metrics, it has been demonstrated by Dunlap that the alternative robust DF formulation,[69]

$$(pq|rs) \approx d^B_{pq}(B|rs) + (pq|A)d^A_{rs} - d^A_{pq}(A|B)d^B_{rs}, \tag{38}$$

will be free of first-order errors due to the residuals in the generalized densities $\delta\rho_{pq}(\vec{r}) \equiv \rho_{pq}(\vec{r}) - \tilde{\rho}_{pq}(\vec{r})$. That is, the principle component of the error in the robust DF formulation is quadratic in the density residuals. This may significantly increase the fidelity of the DF approximation, at the cost of replacing the one-term "SVS" representation[299] of Equation 31 with the three-term expansion of Equation 38. However, for the special case of the Coulomb metric, the SVS and robust representations are identical, and both reduce to the simple, one-term formulae of Equations 34 or 37. This largely explains the widespread adoption of Coulomb-metric density fitting - the one-term symmetric formula of Equation 37 may be used, and yet the factorization remains free of first-order error.

One complication of density fitting is the need to introduce an auxiliary basis set $\{\chi_A(\vec{r})\}$. In general, this is accomplished by optimizing an auxiliary basis set for a given primary basis set and class of electronic structure problems. For example, the

cc-pVXZ-JKFIT sets[312] are designed to provide an accurate representation of the J and K matrices of SCF theory for the Dunning cc-pVXZ basis sets,[70] whereas the cc-pVXZ-MP2FIT sets[313] are designed to accurately represent the hole-particle excitation space for MP2 and other correlated methods. In practice, optimized auxiliary basis sets have between $2-5\times$ the number of functions as the primary basis sets, e.g., in the range of several dozen to a few hundred auxiliary functions per atom. Lower ratios of auxiliary functions to primary functions occur for large and/or diffuse bases sets (e.g., aug-cc-pV5Z), for which the pair-space is more redundant. Thus, efficiency of the DF approximation *increases* for larger and more-diffuse basis sets, providing an important complement to sparsity-based accelerations which break down in these environments.

DF preserves the eightfold symmetry of the ERI tensor, along with the positive definiteness of the chemist's ERIs. The formal positive definiteness of the physicist's ERIs is lost, but this is rarely if ever found to be an issue in practical computations.

1.1.4.2 Cholesky Decomposition

A closely related alternative to density fitting is the pivoted Cholesky decomposition (CD) of the ERI tensor.[24,254,151,8,6] The resultant factorization is,

$$(pq|rs) \approx L^C_{pq} L^C_{rs}. \tag{39}$$

The factors L^C_{pq} are found by applying the recursive pivoted Cholesky-Crout algorithm to the SPD chemist's integrals,

1. Set the residual matrix $R_{pq,rs}$ to the original ERI tensor $(pq|rs)$.

2. Choose the next expansion vector $C \equiv pq_{\max}$ where $pq_{\max}$ is the pair index with maximal element on the diagonal of the residual tensor $R_{pq,pq}$. This selection is referred to as "pivoting." If $R_{C,C} < \delta$, terminate the algorithm.

3. Form the diagonal Cholesky element $L^C_C = \sqrt{R_{C,C}}$.

4. Form the rest of the C-th Cholesky vector as $L_{pq}^{C} = R_{pq}^{C}/L_{C}^{C}$ for $pq \neq C$.

5. Repeat from 2.

In practice, the residual matrix $R_{pq,rs}$ is not explicitly stored. Instead, the diagonal elements $R_{pq,pq}$ are stored and the other elements are computed on the fly via $R_{pq,rs} = (pq|rs) - L_{pq}^{C} L_{rs}^{C}$.

Note that due to Sylvester's criterion [the property that every principal minor of $(pq|rs)$ must be SPD for $(pq|rs)$ to be SPD] implies that this algorithm reduces Chebyshev norm (the maximum residual) of the chemist's ERI in each step. If the algorithm is carried out to N_p^2 factors C, the residual will be exactly zero. In practice, the factorization is terminated after only a linear number of expansion vectors N_C, which is typically in the range of $3-5\times$ the size of the primary basis N_p, corresponding to δ in the range of 10^{-4} to 10^{-7}.

There is a deep connection between DF and CD approaches. In the former, the auxiliary basis is constructed of *a priori* empirically specified auxiliary functions. In the latter, a system-specific auxiliary basis is constructed on the fly by examining the span of the pair space to successively include the next-most-optimal auxiliary function C. In general, this function C is not centered on any one atom. In the often-invoked one-center CD approach (1C-CD),[7] the pivots C are restricted to be taken from the set of pq with both p and q on the same atom. This approximation is no longer formally guaranteed to converge to zero error as the auxiliary basis is expanded, but is often quite accurate in practice. Moreover, 1C-CD places CD and DF on a one-to-one footing, as both now use an auxiliary basis constructed of atom-centered Gaussians.[6]

In practice the CD factorization of the ERI tensor is not used as often as DF, even though the CD factorization is somewhat more efficient in terms of number of expansion vectors N_C for a given accuracy. One reason for this is that the CD algorithm requires that every existing vector L_{pq}^{C} be used in the construction of the

next Cholesky vector. Thus, the entire L_{pq}^{C} tensor must be stored in memory to avoid an excessive number of disk operations in forming the Cholesky vectors, limiting the maximum tractable system size to roughly 1000 primary basis functions. Another reason is that analytical gradients of the Cholesky ERIs are notoriously difficult to derive - tractable progress along these lines has thus far only been achieved in the context of 1C-CD and the concomitant duality with DF.[10] CD is used widely in certain contexts - especially in high accuracy computations on small molecules (such as CD coupled cluster approaches[60]), or in theories such as coupled-cluster of symmetry-adapted perturbation theory for which custom auxiliary bases are not available for the DF approach. Generally, one finds that the -JKFIT and -MP2FIT auxiliary basis sets are sufficiently accurate for general Fock-type and correlation-type methods, respectively, but the application of CD allows one to be more certain that the appropriate accuracy is retained in the ERI tensor. From this point forward, when we refer to DF, it should be understood that we view CD as a particular flavor of DF and are also including it within the discussion.

1.1.4.3 Pseudospectral Representation

The DF approach provides excellent compression of the ERI tensor to order-3 objects, recasts the problem in terms of tensor contractions, and succeeds in breaking the scalings of Coulomb-type contractions. However, it is unable to break the scalings of exchange-type contractions, due to the pinned occurrence of the p and q indices in the same three-index tensor. An alternative approach which manages to "unpin" the p and q indices is the pseudospectral (PS) representation.[80,81,82,83,251,250,93,188,190,187,187,246,85,150,208,162]

The PS approach starts by contracting over $\vec{r}_2$ in the ERI, yielding a representation of the ERI tensor written as an integral over only $\vec{r}_1$,

$$(pq|rs) = \int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)V_{rs}(\vec{r}_1), \tag{40}$$

where $V_{rs}(\vec{r}_1)$ is the ESP of the generalized density $\rho_{pq}(\vec{r}_2)$ in coordinate $\vec{r}_1$,

$$V_{rs}(\vec{r}_1) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \frac{1}{r_{12}} \phi_r(\vec{r}_2)\phi_s(\vec{r}_2). \tag{41}$$

At this point, it is noted that the ESPs of the generalized densities are all finite - the $1/r_{12}$ singularity has been extinguished by contraction against the finite span of the $\rho_{rs}(\vec{r}_2)$ pair-space functions. The schwerpunkt of pseudospectral theory is the introduction of a quadrature grid $\{< \vec{r}_P, w_P >\}$ to resolve the integral in $\vec{r}_1$ via $\int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; f(\vec{r}_1) \approx w_P f(\vec{r}_P)$. This exposes the multiplicative separability of the pair space in position-space, unpinning the indices p and q,

$$(pq|rs) \approx w_P R_p^P R_q^P V_{rs}^P. \tag{42}$$

Here $R_p^P \equiv \phi_p(\vec{r}_P)$ is the collocation matrix (the values of the basis function on the grid), and $V_{rs}^P \equiv V_{rs}(\vec{r}_P)$ is the collocation of the ESP to the grid. V_{rs}^P may easily be computed from standard nuclear attraction integral routines. For conceptual convenience, we may write this factorization as,

$$(pq|rs) \approx X_p^P X_q^P V_{rs}^P, \tag{43}$$

where $X_p^P \equiv \sqrt{w_P} R_p^P$ is the weighted collocation matrix.

It is apparent that the PS approach decomposes the ERI as a product of three tensors - two order-2 collocation matrices and an order-3 ESP matrix. Thus the order of PS is the same as DF. Whereas DF expands the ERI over an auxiliary basis of smooth functions, PS expands the ERI over an effective auxiliary basis of Dirac delta functions. In practice, the DF auxiliary basis is usually much smaller (e.g., a factor of tenfold smaller) than the PS quadrature grid. Therefore, PS does not usually provide the same level of compression that DF provides. The principal benefit of PS is the enhanced flexibility of the factorization obtained by unpinning the p and q factors. For canonical instance, the exchange matrix may now be formed in $\mathcal{O}(N^3)$ operations,

$$K_{pq} \approx \left[X_r^P \left[V_{qs}^P \left[X_s^P D_{rs} \right] \right] \right]. \tag{44}$$

This particular factorization underlies the extremely successful Jaguar[31] and ORCA[207] packages for large-scale SCF computations.

Note that in the full PS approach, a slightly different form of the factorization is actually used,[81]

$$(pq|rs) \approx Q_p^P R_q^P V_{rs}^P. \tag{45}$$

This arises from the understanding of the ERI as an expectation value of the ESP of each rs pair in the primary basis space,

$$(pq|rs) = \langle p|\hat{V}_{rs}|q\rangle = \int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \phi_p(\vec{r}_1) V_{rs}(\vec{r}_1) \phi_q(\vec{r}_1). \tag{46}$$

In this understanding, the PS approach may be written as,

$$(pq|rs) \approx Q_p^P V_{rs}^P R_q^P. \tag{47}$$

Here R_q^P is an operator used to project from spectral space q to a discrete representation of physical space P. The ESP operator $\hat{V}_{rs}$ is then applied in physical space, where it is diagonal. The operator Q_p^P then projects the result back onto spectral space. If R_p^P is used for the second projector, components of the physical-space representation of $\hat{V}_{rs}|q\rangle$ outside of the primary basis will be accidentally projected onto $\langle p|$, if a finite quadrature grid is used (alias). This effect can be attenuated by the use of a dealiasing projector,

$$Q_p^P = P_{pt} \underbrace{[R_t^Q w_Q R_u^Q]^{-1}}_{\tilde{S}_{tu}^{-1}} R_u^P w_P. \tag{48}$$

Here the labels t and u refer to a united basis set built as the union of the primary basis set $\{\phi_p(\vec{r})\}$ and a "dealias basis set" $\{\xi_z(\vec{r})\}$ which is chosen to be orthogonal to the primary basis set. The projector $P_{pt} = \delta_{pt}$ annihilates the dealias basis space in the last step in the product. Broadly speaking, this approach decomposes the spectrum of any physical-space signal into contributions from the primary basis set and the dealias basis set, and discards the latter. This means that any alias in the signal $V_{rs}^P R_q^P$ that

lies in the span of the dealias basis set will be removed from the PS approximation, lowering the error. Practically speaking, this approach can lower the error of the PS approximation by roughly an order of magnitude for a given grid size, with no increase in the computation time. Formally, the error in the PS approximation will be zero if either the quadrature grid or the dealias basis are complete, but in practice, both must be expanded together to prevent the matrix $\tilde{S}_{tu}$ from becoming singular.

Physical-space quadrature grids for molecular geometries[83,20,88,205,297,204,180,87,73,44] are found in several areas in electronic structure theory, including Kohn-Sham density functional theory (KS-DFT)[153] and PS theory. These quadrature grids are extremely interesting objects. Molecular geometries are notoriously difficult to accurately discretize, due to the large range of length scales involved. For instance, a standard rectilinear or spherical grid will fail to efficiently capture the rapid oscillations of molecular integrands near the nuclei and the smooth but rather slow decay of the integrals in the tails of the density field. Practical grids are instead designed to track the molecular geometry. These are typically constructed as a sum of atom-centered spherical grids, each of which is further decomposed as a direct product of angular[171] and radial grids.[83,20,205,297,204,87] Smaller angular grids may be used near the nucleus to prevent crowding, or near the periphery, where a slowly varying integrand is expected, leading to "pruned" atomic grids.[83,88] These atomic grids are then stitched together via an atomic weighting scheme.[20,297] Such a scheme assigns the physical space nearest each atom to that atom's spherical grid, allowing for the most efficient spherical integration grid to capture each part of the integral. The atomic weighting scheme is designed so that each point in space is counted exactly once over all of the atomic grids, and so that the partitioning weights vary smoothly from one atom to the next. Both PS[83] and DFT[20,88,205,297,204,180,87,73,44] grids were developed in roughly the same timeframe. In recent years, the development of DFT grid technology has progressed more than PS, so modern PS implementations typically use small DFT

grids for their quadratures.

Dealias basis sets must also be specified for each primary basis set in the full PS theory. Typically these are found by placing functions of larger and smaller radial extent around and between the functions present in the primary basis set, and also by adding functions of one or two higher angular momenta than present in the primary basis set. Dealiasing has fallen somewhat out of favor in recent years, due to concerns of $p \leftrightarrow q$ interchange symmetry and the additional complexity of specifying the dealiasing basis set. For instance the "chain-of-spheres exchange" [COS(X)] approach[208] eschews dealiasing, and instead directly uses the representation of Equation 43. Note that this avoidance of dealiasing may be a bit premature - an unpublished "Löwdin-style" dealiasing scheme for PS which preserves $p \leftrightarrow q$ interchange symmetry is,

$$X_p^P \equiv P_{pt}[R_t^Q w_Q R_u^Q]^{-1/2} R_u^P \sqrt{w_P}. \tag{49}$$

Note that PS loses many of the formal properties of the ERI tensor. The $pq \leftrightarrow rs$ particle interchange symmetry is lost, as is the $p \leftrightarrow q$ symmetry in the full PS theory. Additionally, the chemist's and physicist's ERIs are no longer Hermitian, and thus cannot be positive definite. Moreover, the practical accuracy of PS is much lower than DF. The root cause of this is the extremely ill-conditioned nature of the $V_{rs}(\vec{r}_1)$ contributions to the integrands. For example, as the $\rho_{rs}(\vec{r})$ densities approach Dirac delta functions at center A (this occurs in the core-core integrals), the ESP approaches $1/r_{1A}$, and exhibits a marked transient for small r_{1A}. This is particularly an issue for SCF-type theories, where the core integrals contribute significantly. One approach to attenuate this issue is to replace certain one-, two-, or even three-center ERIs by their analytical counterparts.[93] This complicates the implementation of PS codes (and especially their gradients), but can significantly improve accuracy. Even with analytical corrections, the PS grid space remains significantly larger than the corresponding DF space. For sub-mH accuracy, on the order of 1000 points per atom

are often required, e.g., roughly tenfold greater than the corresponding number of DF auxiliary functions.

1.2 Tensor Hypercontraction

1.2.1 Motivation

To this point, the successes and shortcomings of DF and PS theories exhibit remarkable complementarity - DF is less flexible but more accurate, and vice-versa for PS. This suggests that there might exist another approach which merges the best properties of these two methods, and might have additional advantages. Such an approximate representation of the ERI tensor might exhibit the following properties,

1. Respect the full eightfold symmetry and at least the SPD nature of the chemist's integrals, as DF does.

2. Provide enhanced flexibility by unpinning the p and q indices, as PS does. If this and the previous point are achieved together, this necessarily provides for separability of the r and s indices, which is not possible in DF or PS.

3. Retain accuracy properties which are more closely related to DF than to PS, i.e., sub-mH errors in computed energies with auxiliary expansions with at most a few hundred vectors per atom.

4. If possible, compress the ERI to order-2 factor tensors. This is not possible in either DF or PS.

These goals were attained by the introduction of the tensor hypercontraction factorization (THC) for the ERI tensor.

1.2.2 Development

THC has the factorization topology,

$$(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q, \tag{50}$$

and arises for physically-motived reasons. The THC representation was developed from 2011 to the present by Edward Hohenstein, Robert Parrish, Todd Martínez, and David Sherrill, with additional technical contributions from Sara Kokkila, Chenchen Song and Lee-Ping Wang. A brief historical sketch of the development is included below, to illuminate some of the unusual terminology and interpretations that have occurred throughout the literature reports on this topic,

1. We considered the idea of "double pseudospectral" factorization of ERI tensor, wherein a PS-type quadrature was applied in *both* coordinates of the ERI tensor. This idea immediately is plagued by the singularity of the $1/r_{12}$ operator, but it was suggested that this method might have some utility in ERIs involving the Ewald operator $\mathrm{erf}(\omega r_{12})/r_{12}$. This was the first time the order-2 THC factorization $(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q$ was encountered in our work, and it was shown that this could reduce second-order Møller-Plesset Perturbation Theory (MP2) to $\mathcal{O}(N^4)$ for the first time.

2. We first solved the problem of the singularity of $1/r_{12}$ by applying PARAFAC decomposition to the three-index overlap integrals of SVS-type overlap-metric density fitting, leading to "PF-THC." This was shown to accurately reduce MP2 *and* MP3 to $\mathcal{O}(N^4)$ with minimal loss in accuracy.[121]

3. We attempted to apply PS quadrature and DF renormalization of the $1/r_{12}$ operator to avoid the costly PARAFAC decomposition in PF-THC, leading to an approach known as PS-THC. This approach required too many quadrature points to be practically useful, due to an interesting mismatch in the spans of the DF auxiliary basis and the PS quadrature.

4. We circumvented the failure of PS-THC by replacing DF renormalization of $1/r_{12}$ with direct least-squares renormalization, leading to the very-accurate and simple grid-based LS-THC method.[216]

5. We applied LS-THC to the opposite-spin amplitudes of coupled-cluster singles and doubles (CCSD), proving that this led to an accurate and wholly $\mathcal{O}(N^4)$ factorization of the previously $\mathcal{O}(N^6)$ method.[122]

6. We investigated a number of technologies for the construction of the LS-THC quadrature grid, in the context of THC-MP2. Discrete Variable Representation (DVR) technology was found to produce the most efficient grids, but THC-MP2 was found to be largely robust to the specifics of the grid generation.[217]

7. We demonstrated that PS-THC and LS-THC are identical and exact in the polynomial basis sets encountered in nuclear structure computations, leading to the exact THC (X-THC) representation in this limit. Besides its lossless utility in nuclear physics, X-THC demonstrates why LS-THC succeeds but PS-THC fails in the molecular problem.[218]

8. We applied LS-THC ERIs to produce efficient implementations of approximated coupled cluster (CC2) and excited-state CC2 (EOM-CC2) methods, leading to accurate $\mathcal{O}(N^4)$ implementations of these methods which are competitive with the best DF-based implementations.[119,120]

9. We applied a new flavor of "weighted" LS-THC ERIs to the notoriously approximation-resistant particle-particle ladder term in CCSD, providing a 10× acceleration of that term and a 2.4× speedup of the overall CCSD method over the best DF implementation.[223]

Below, we present a review of the salient details of THC ERIs. We begin by presenting X-THC to demonstrate how the factorization arises naturally in a perfect basis set. Then we describe the various machinery (PF-THC, PS-THC, LS-THC, etc) required to develop an accurate approximation in finite basis sets composed of atom-centered functions, for use in electronic structure theory.

1.2.2.1 Exact Tensor Hypercontraction

For a 1D problem, consider a basis set of N_o functions, each built of a polynomial $P_p(x)$ of up to degree N_p, multiplied by a weight function $v(x)$ which is the same for all basis functions,

$$\phi_p(x) \equiv P_p(x)v(x). \tag{51}$$

The pair space for this problem $\{\rho_{rs}(x)\} = \{\phi_p(x)\phi_q(x)\}$ naïvely contains N_p^2 functions, but is easily seen to be spanned by the $2N_p$-function auxiliary basis,

$$\chi_A(x) \equiv P_A(x)v^2(x), \tag{52}$$

where $P_A(x)$ is a polynomial of up to degree $2N_p$. It is also noted that an exact $2N_p$-node Gaussian quadrature $\{< x_P, w_P >\}$ exists for the overlap matrix of the auxiliary basis set,[2]

$$(AB) \equiv \int_{\mathbb{D}} \mathrm{d}^1x \ \chi_A(x)\chi_B(x) = w_P\chi_A(x_P)\chi_B(x_P). \tag{53}$$

The exact Gaussian quadrature for an arbitrary weight function $v(x)$ and domain $\mathbb{D}$ can always be found, e.g., by the Golub-Welsch algorithm.[91]

Now consider the equivalent of the ERI tensor in these coordinates, i.e., the expectation value of a local, spatial, two-particle interaction operator $\hat{o}$,

$$(pq|\hat{o}|rs) \equiv \iint_{\mathbb{D}^2} \mathrm{d}^1x_1 \ \mathrm{d}^1x_2 \ \phi_p(x_1)\phi_q(x_1)o(x_1,x_2)\phi_r(x_2)\phi_s(x_2). \tag{54}$$

The first stage is to introduce an exact resolution of the identity for the pair space, $\hat{1} \equiv |A)(A|$ (for convenience the auxiliary basis is assumed to be orthonormal), leading to the SVS-type density fitting representation,[299]

$$(pq|\hat{o}|rs) = (pqA)(A|\hat{o}|B)(Ars), \tag{55}$$

where,

$$(pqA) \equiv \int_{\mathbb{D}} \mathrm{d}^1x \ \phi_p(x)\phi_q(x)\phi_A(x), \tag{56}$$

and,

$$(A|\hat{o}|B) \equiv \iint_{\mathbb{D}^2} \mathrm{d}^1 x_1 \ \mathrm{d}^1 x_2 \ \chi_A(x_1) o(x_1, x_2) \chi_B(x_2). \tag{57}$$

Note that this is exact for this problem due to the polynomial closure of the pair space. This exact DF representation has been known and exploited in nuclear structure computations[38,294,90] for at least as long as approximate DF representations have been employed in electronic structure theory. The second, less obvious, step is to note that the three-center overlap integrals are all representable by linear combinations of two-center auxiliary basis overlap integrals, and thus may be directly resolved by the Gaussian quadrature for the auxiliary basis overlap metric,

$$(pqA) = X_p^P X_q^P Y_A^P. \tag{58}$$

Here we define the collocations, $X_p^P \equiv \phi_p(x_P)$ and $Y_A^P \equiv w_P \chi_A(x_P)$. Now, collecting a useful intermediate,

$$Z^{PQ} \equiv Y_A^P (A|\hat{o}|B) Y_B^Q, \tag{59}$$

we arrive at the X-THC representation,[217]

$$(pq|rs) = X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q. \tag{60}$$

I.e., an exact and separable compression of the N_p^4 integral tensor in terms of a product of $N_p \times 2N_p$ collocation matrices and a $2N_p \times 2N_p$ grid-to-grid interaction matrix. Note that the operator Z^{PQ} has a rich interpretation as the physical-space collocation of a renormalized or "dealiased" $\hat{o}$ operator which has been cleaned up by projection through the auxiliary basis to remove components that lie outside the span of the pair space. This renormalization allows the operator to be collocated at the finite set of quadrature points without picking up any alias, regardless of the details of $\hat{o}$.

Note that the auxiliary basis $\{\chi_A(x)\}$ and Gaussian quadrature $\{< x_P, w_P >\}$ must be chosen together to provide for an exact representation. If a larger auxiliary

basis is selected, there will be less dealiasing of the $\hat{o}$ operator, and a larger Gaussian quadrature will be required to remove the remaining parts of the renormalized operator which lie outside the pair space. Moreover, the auxiliary basis must be chosen to (at least) span the pair space, or the dealiasing step will falsely remove valid parts of the $\hat{o}$ operator which lie in the pair space. This will shortly explain the practical failure of PS-THC vs. LS-THC.

For ND problems, a basis selected as a direct product of 1D polynomials will exhibit an X-THC factorization of very similar structure as in the 1D case.

1.2.2.2 PARAFAC Tensor Hypercontraction

The first tractable THC-type representation of the ERI tensor in molecular electronic structure theory was actually proposed in a somewhat obscure paper by Boys and Rajagopal in 1966,[36] and involved a form of double pseudospectral plus a heuristic renormalization of the $1/r_{12}$ operator at short range based on intermediates often seen in Boys's work on analytical ERI formulae in Gaussian basis sets. This approach was not widely adopted, primarily because of Boys's other excellent work on analytical integral evaluation, which proved to be much faster (and more accurate) for small molecules than his THC proposal. That this elegant and extremely forward-looking paper was overlooked is a considerable shame - for instance, we were not even aware of this work's existence until our fourth paper in the THC series.

In our own work, the first tractable THC-type representation of the ERI tensor was PARAFAC tensor hypercontraction (PF-THC, earlier referred to as THC-DF).[121] This approach starts from the SVS-form of DF within the overlap metric,[299]

$$(pq|rs) \approx (pqA)(AB)^{-1}(B|C)(CD)^{-1}(Drs). \tag{61}$$

Here (AB) are two-index overlap integrals,

$$(AB) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \chi_A(\vec{r}_1)\chi_B(\vec{r}_1), \tag{62}$$

and (pqA) are the all-important three-index overlap integrals,

$$(pqA) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)\chi_A(\vec{r}_1). \tag{63}$$

It is then noted that for a primary and auxiliary basis composed of exponentially decaying functions, the three-center overlap integral tensor is sparse and low-rank, and may be accurately represented by the PARAFAC decomposition,[42]

$$(pqA) \approx X_p^P X_q^P Y_A^P. \tag{64}$$

Here P is a linear-scaling index whose size is specified by the user prior to locating the X and Y factors. If we then define,

$$Z^{PQ} \equiv Y_A^P (AB)^{-1} (B|C) (CD)^{-1} Y_B^Q, \tag{65}$$

we arrive at a THC form,

$$(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q. \tag{66}$$

In PF-THC, no physical form for the X and Y factor tensors is proposed. Rather, the PARAFAC decomposition of the three-center overlap integrals is asserted, and later found to be accurate with a linear number of auxiliary factors P. The factors X and Y are determined by Alternating Least Squares (ALS) PARAFAC decomposition.[281] A random guess for X and Y is made. Then, the optimal X is determined by least-squares fitting to the overlap integrals, under the current choice for Y. Next, the optimal Y is determined for the new X. This procedure is continued until the factors stop changing to within some threshold ϵ. It was found that mH accuracy could be obtained for MP2 and MP3 correlation energies with a PARAFAC expansion with twice the number of P factors as functions in the primary basis set.

One issue with PF-THC is the reliance on SVS-type DF in the overlap matrix, which may be much less accurate than Coulomb-metric DF. This was later found to not be a problem - one may also apply PF-THC within Coulomb-metric density

fitting by taking the PARAFAC decomposition of the three-index Coulomb integrals $(pq|A)$. These integrals are less element-wise sparse than (pqA), but exhibit the same PARAFAC-type rank-sparsity. However, by the time this realization occurred, LS-THC had already superseded PF-THC as our preferred approach.

1.2.2.3 Pseudospectral Tensor Hypercontraction

A major issue with PF-THC was the large computational effort and reliance on a random guess inherent in the PARAFAC decomposition of (pqA). It was quickly realized that a physically-motivated and direct PARAFAC of (pqA) existed: quadrature resolution on a molecular integration grid of the form $\{< \vec{r}_P, w_P >\}$,

$$(pqA) \approx X_p^P X_q^P Y_A^P, \tag{67}$$

where $X_p^P \equiv \phi_p(\vec{r}_P)$ and $Y_A^P \equiv w_P \chi_A(\vec{r}_P)$. This pseudospectral THC (PS-THC) was implemented, but was surprisingly found to be practically inefficient - on the order of 1000 points per atom were required for mH accuracy. An understanding of the origins of the failure of PS-THC was provided by X-THC - in PS-THC, the auxiliary basis and quadrature grids are chosen independently, precluding the existence of a proper dealiasing step. For instance, one could imagine increasing the fidelity of the auxiliary basis within a fixed molecular quadrature grid - this should presumably make PS-THC more accurate, by lowering the underlying DF error. However, in the limit of a complete auxiliary basis set, the PS-THC Z^{PQ} operator trends to the raw $1/r_{PQ}$ operator that plagues double pseudospectral theory, meaning that an *infinite* quadrature grid is required. Thus, increasing the fidelity of the auxiliary basis actually causes the PS-THC error to go *up*. In principle, the quadrature grid can also be made larger to compensate, but in practice it was found that by the time mH accuracy is achieved, the PS-THC grid is too large for practical deployment of the THC ERIs.

Note that one need not start from overlap-metric SVS DF for PS-THC. For instance, one may start from Coulomb-metric DF i.e., $(pq|rs) \approx (pq|A)(A|B)^{-1}(B|rs)$

and apply a PS-THC representation with X_p^P as above and a Z operator chosen as,

$$Z^{PQ} \equiv V_A^P (A|B)^{-1} V_B^Q, \tag{68}$$

where,

$$V_A^P \equiv w_P \int_{\mathbb{R}^3} \mathrm{d}^3 r_2 \, \frac{1}{r_{2P}} \chi_A(\vec{r}_2). \tag{69}$$

This corresponds to applying the PARAFAC $(pq|A) \approx X_p^P X_q^P V_A^P$ via quadrature in $\vec{r}_1$. This inherits the same problems as overlap-metric PS-THC, but was later used accurately and effectively in the context of atomic symmetry-adapted perturbation theory (A-SAPT)[222] with much larger grids.

1.2.2.4 Least-Squares Tensor Hypercontraction

A much more accurate representation is least-squares THC (LS-THC),[216] which arises from the realization that in an X-THC-like scheme, one does not need the explicit form of the renormalized interaction operator, but rather its collocation to the grid points, i.e., the matrix elements Z^{PQ}. In LS-THC, we determine Z^{PQ} to provide for the minimal vector 2-norm of the residual of the THC ERIs, e.g., the Z operator that minimizes the sum of the squares of the error of the approximate integrals. This is accomplished by minimizing the quadratic positive-definite metric,

$$O(Z^{PQ}) = \frac{1}{2} \left\| (pq|rs) - X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q \right\|_{\mathcal{F}}^2. \tag{70}$$

Owing to the linearity of the THC ERIs in Z, there exists an analytical solution,

$$Z^{PQ} = \underbrace{\left[X_{pq}^P X_{pq}^{P'} \right]^{-1}}_{\left[S^{PP'}\right]^{-1}} \underbrace{X_{p'q'}^{P'} (p'q'|r's') X_{r's'}^{Q'}}_{E^{P'Q'}} \underbrace{\left[X_{rs}^{Q'} X_{rs}^{Q} \right]^{-1}}_{\left[S^{Q'Q}\right]^{-1}}. \tag{71}$$

In the above $S^{PP'}$ is the grid metric matrix, and accounts for the differing spans of the pair space and grid space. The matrix $E^{P'Q'}$ is the projection of $1/r_{12}$ through the pair space and from there onto the raw quadrature grid (the dealiasing step). The formation of $E^{P'Q'}$ is the rate-limiting step in the factorization - this scales as $\mathcal{O}(N^5)$

if exact ERIs are used, or $\mathcal{O}(N^4)$ if DF ERIs are decomposed, using any choice of metric.

One ambiguity in LS-THC is the use of the quadrature grid weights - if the computations were carried out at infinite precision, it would not matter what weights were applied in the collocation matrix, as the least-squares fitting procedure effectively determines its own weights. However, in practice, the grid metric matrix $S^{PP'}$ is extremely ill-conditioned, due to redundant linear combinations of grid points. In these cases, it is considered best practice to construct the metric matrix so that its eigenvalues convey the magnitude of physical significance of each eigenvector, and then to discard contributions from eigenvalues below a certain threshold when forming matrix powers. Typically, small eigenvalues are discarded to provide for a given relative condition number in the metric matrix, e.g., to use the part of the metric that provides for a relative condition number of 10^{10}. In an atom-centered Becke-type grid (discussed above in the review of PS theory), this can be accomplished by exploiting the information contained in the quadrature grid weights $\{w_P\}$. We have found that the choice $X_p^P \equiv \sqrt[4]{w_P}\phi_p(\vec{r}_P)$ makes sense topologically (as it allows the quadrature grid to integrate the pair-space overlap matrix exactly once if no least-squares fitting is applied), and significantly improves the condition number of $S^{PP'}$. Moreover, this application of weights imparts knowledge of the volume element into the grid metric matrix, providing for a more-physically-sensible elimination of redundant combinations of grid points in the pseudoinversion.

Note that besides a grid-based renormalization process, LS-THC can also be viewed as the first stage of a more-general ALS PARAFAC decomposition wherein *both* the X and Z tensors are optimized,

$$O(Z^{PQ}, X_p^P) = \frac{1}{2}\left\|(pq|rs) - X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q\right\|_{\mathcal{F}}^2. \tag{72}$$

The solution of this latter "optimal THC" or OT-THC procedure would necessarily require an iterative procedure due to the nonlinear couplings between and within the

X and Z tensors, but might provide better accuracy than LS-THC. As it is, LS-THC is very simple because it is only the first stage in OT-THC, but proves to be quite accurate because a physically-motivated choice was made for the "guess" of the factor tensor X_p^P.

In practice, LS-THC has proven to be markedly accurate and robust in small grids with at most a few hundred points per atom. For example, in our first study on the method, LS-THC-MP2/cc-pVDZ correlation energies were found to be accurate to roughly 0.05 kcal mol^{-1} in molecular systems with roughly 60 atoms, using grids with roughly 150 points per atom. Note that no serious attempt was made to optimize the grids for this paper - optimized grids with half as many points have been found to provide comparable accuracy.[217,154] Moreover, the errors in energy differences computed with LS-THC ERIs generally benefit from favorable error cancellation of at least one order of magnitude,[154] as is seen with DF methodology.

1.2.2.5 Weighted Least-Squares Tensor Hypercontraction

There exists an interesting dichotomy between DF and LS-THC - both employ a least-squares fitting procedure, but the nature of the objective function is quite different between the two cases. One particularly extreme exposition of the difference is that, for a set DF auxiliary basis or LS-THC quadrature grid, the DF approach can "cheat" and continue to provide accurate chemical results if the primary basis is expanded, while the LS-THC approach will necessarily diminish in accuracy as the primary basis is expanded. The origin of this effect stems from the nature of the least-squares fitting equations. For DF, the fitting is separately optimized for each pair-space function. E.g., if we apply DF to a primary basis including the pair-space function $\rho_{pq}(\vec{r})$, and then expand the basis to also include $\rho_{p'q'}(\vec{r})$ pair-space functions, the fit for the $\rho_{pq}(\vec{r})$ function is unchanged. This is not the case in LS-THC - all of the pair-space residuals contribute to a joint objective function. This implies that when the pair space span

is increased, LS-THC compromises by making a larger error for all integrals. Thus, LS-THC does not possess one of DF's best strengths - that the auxiliary basis may be constructed to provide for the highest accuracy in integrals that are likely to be most relevant to chemical observables while tolerating larger relative error in less-important integrals.

Weighted LS-THC (W-LS-THC)[223] is designed to partially overcome this problem. In this scheme, we modify the LS-THC objective function by including weights to inform the least-squares fitting procedure of the expected relative importance of various integrals. For instance, if a set of weights $\{\xi_p\}$ exists which encapsulates the importance of each orbital p (in any desired rotation of the orbital basis), the LS-THC objective function may be modified as,

$$O(Z^{PQ}) = \frac{1}{2} \left\| \xi_p \xi_q \xi_r \xi_s \left[(pq|rs) - X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q \right] \right\|_{\mathcal{F}}^2 . \tag{73}$$

This implies only a minor modification of the formula for the W-LS-THC Z^{PQ},

$$Z^{PQ} = \underbrace{\left[X_{pq}^P \xi_{pq}^2 X_{pq}^{P'} \right]^{-1}}_{\left[S^{PP'}\right]^{-1}} \underbrace{\xi_{p'q'}^2 X_{p'q'}^{P'} (p'q'|r's') X_{r's'}^{Q'} \xi_{r's'}^2}_{E^{P'Q'}} \underbrace{\left[X_{rs}^{Q'} \xi_{rs}^2 X_{rs}^{Q} \right]^{-1}}_{\left[S^{Q'Q}\right]^{-1}}, \tag{74}$$

where $\xi_{pq}^2 = \xi_p \xi_p \xi_q \xi_q$.

The initial application of W-LS-THC was THC-CCSD. In this approach, an MP2 natural orbital (NO) computation[295] is first performed as a model for the forthcoming (and much more expensive) CCSD computation.[239,270] The MP2 NOs provide a rotation of the occupied and virtual spaces to a basis in which the contributions to the correlation energy diminish at a near-maximal rate (the NO eigenbasis), along with a metric for how much each NO function contributes to the correlation process (the NO eigenvalue or occupation, η_p). For reasons of dimensionality, the square roots of the occupations are the appropriate weights for W-LS-THC, e.g., $\xi_p \equiv \sqrt{|\eta_p|}$, with the W-LS-THC fitting performed in the MP2 NO basis. This approach was found to reduce the errors of THC-CCSD by a factor of tenfold relative to standard LS-THC without

any increase in computational complexity, and is particularly critical in larger basis sets such as cc-pVTZ. One can view this overall idea as a machine learning approach - the MP2 computation provides a cheap model to discover the relevant portions of the orbital space, following which the more-expensive CCSD model is applied with an acceleration scheme that takes advantage of the knowledge obtained in the MP2 stage of the computation.

1.2.3 Tensor Hypercontraction Publications

The remainder of this half of the thesis document consists of four papers detailing the development and deployment of THC ERIs in electronic structure theory. These works include,

1. The development and interpretation of X-THC integrals within Hermite function basis sets, adapted from Ref. 218.

2. The development of LS-THC ERIs and deployment in THC-MP2, adapted from Ref. 216.

3. The introduction and exploration of new grid technologies for LS-THC ERIs in the context of THC-MP2, adapted from Ref. 217.

4. The development of W-LS-THC and practical application to THC-CCSD, adapted from Ref. 223.

The interested reader is referred to the other papers in this series,

1. Our original THC paper, describing PF-THC and applications to $\mathcal{O}(N^4)$ MP2 and MP3.[121]

2. The use of the LS-THC factorization for the compression of wavefunction amplitudes to produce $\mathcal{O}(N^4)$ CCSD and related methods.[122]

3. The application of LS-THC ERIs to provide $\mathcal{O}(N^4)$ THC-CC2.[119]

4. The application of LS-THC ERIs to provide $\mathcal{O}(N^4)$ THC-EOM-CC2 for excited states.[120]

CHAPTER II

EXACT TENSOR HYPERCONTRACTION

The following is adapted from Ref. 218.

2.1 Abstract

Configuration-space matrix elements of N-body potentials arise naturally and ubiquitously in the Ritz-Galerkin solution of many-body quantum problems. For the common specialization of local, finite-range potentials, we develop the eXact Tensor HyperContraction (X-THC) method, which provides a quantized renormalization of the coordinate-space form of the N-body potential, allowing for a highly separable tensor factorization of the configuration-space matrix elements. This representation allows for substantial computational savings in chemical, atomic, and nuclear physics simulations, particularly with respect to difficult "exchange-like" contractions.

2.2 Introduction

The physics of many-body quantum systems is often captured by local, finite-range N-body potentials $\hat{V}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, where $\boldsymbol{x}$ is any convenient parametrization of the physical space, e.g., position space ($\boldsymbol{x} \equiv \boldsymbol{r}$) or momentum space ($\boldsymbol{x} \equiv \boldsymbol{k}$). Given some real, finite, one-particle Ritz-Galerkin basis set $\{\psi_i(\boldsymbol{x})\}$, the configuration-space representation of $\hat{V}$ is the integral tensor,

$$\langle i \ldots n|\hat{V}|i' \ldots n'\rangle = \int \mathrm{d}\boldsymbol{x}_1 \ldots \int \mathrm{d}\boldsymbol{x}_N \tag{75}$$

$$\psi_i(\boldsymbol{x}_1) \ldots \psi_n(\boldsymbol{x}_N)\hat{V}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\psi_{i'}(\boldsymbol{x}_1) \ldots \psi_{n'}(\boldsymbol{x}_N).$$

The generation, manipulation, and storage of this tensor is a major hurdle in many-body quantum simulations. In order to overcome the computational difficulties inherent to such high order tensors, it is common to introduce simplifying approximations. For example, the Slater approximation[280] has been applied to reduce the numerical expense of treating exchange terms involving the local, two-body Coulomb potential. Unfortunately, such approximations can fail, as exemplified by the often spectacular self-interaction errors induced by local approximations to exchange interactions[227]. Another canonical example is nuclear density functional theory (DFT), in which even more severe approximations are often invoked, e.g., the substitution of actual three-body potentials by density-dependent two-body terms[306], and the use of zero-range two-body Skyrme-like potentials[278] in place of the more realistic finite-range two-body Gogny potential[59]. These approximations often lead to qualitative errors such as divergences in the two-body pairing channel[209], instabilities in random-phase approximation computations with zero-range three-body forces[30], and the breakdown of standard post-mean-field methods[65,66,166,26] for potentials with fractional density dependences. It is clear that an improved technique for faithful and direct treatment of arbitrary local potentials would be preferable to computationally-motivated range limitation and truncations.

In this Letter, we show that an *exact* and *separable* decomposition exists for any local potential in a finite basis set built from polynomial functions in any desired parameterization of the physical space. This decomposition is motivated by our recently introduced Tensor HyperContraction (THC) method for electronic structure[121,216,122], which provided a phenomenological approximation for the electron repulsion integrals involving the Coulomb potential in non-polynomial basis sets. The new eXact Tensor HyperContraction (X-THC) representation reveals two points of great importance for both electronic and nuclear structure problems. First, THC approximation of the Coulomb interaction is exact for basis sets which can be expressed in polynomial form

(and thus the approximation in electronic structure arises only because the basis functions used were of non-polynomial form). Secondly, THC approximation is applicable not only to the two-body Coulomb interaction but also to arbitrary local potentials commonly encountered in nuclear structure (such as the Coulomb, Gogny, local forms of realistic three-body potentials, etc.). Since the nuclear problem is already commonly formulated in terms of polynomial basis sets, this implies that many problems in nuclear structure can now be treated exactly with the *lossless scaling reduction* afforded by the X-THC representation. The first of these points may aid markedly in the search for more efficient THC approximations in electronic structure, while the second may yield unprecedented physical fidelity in nuclear structure computations (especially within the context of nuclear DFT).

Below, we first demonstrate the key features of the X-THC representation through the representative example of a one-dimensional, two-body problem in Cartesian coordinates using Hermite functions. The D-dimensional, N-body generalization of X-THC is then presented. Finally, we present an example implementation of X-THC for the finite-range Gaussian potential in a basis of Hermite functions, demonstrating that X-THC is both lossless and markedly efficient in practice.

2.3 Theory

2.3.1 X-THC Example

Consider a one-dimensional ($D = 1$) problem in Cartesian coordinates, involving a finite basis of $M + 1$ Hermite functions $\{\psi_i(x)\}$ (labeled from 0 to M) with a local two-body ($N = 2$) potential $\hat{V} \equiv \hat{V}(x_1, x_2)$. The potential matrix elements are,

$$\langle ij|\hat{V}|i'j'\rangle \equiv \iint \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{76}$$
$$\psi_i(x_1)\psi_j(x_2)\hat{V}(x_1, x_2)\psi_{i'}(x_1)\psi_{j'}(x_2).$$

The first stage in X-THC is to note that all $(M + 1)^2$ products $\psi_i(x_1)\psi_{i'}(x_1)$ are exactly spanned by an orthonormal "auxiliary" basis $\{\chi_A(x_1)\}$ consisting of $2M + 1$

Hermite functions with a slightly modified spatial range, $\chi_A(x_1) \equiv \psi_A(\sqrt{2}x_1)$,

$$\psi_i(x_1)\psi_{i'}(x_1) = \sum_A [ii'A]\chi_A(x_1), \tag{77}$$

where,

$$[ii'A] \equiv \int_{\mathbb{R}} \mathrm{d}x_1 \; \psi_i(x_1)\psi_{i'}(x_1)\chi_A(x_1). \tag{78}$$

This resolution is well known in the context of nuclear physics[38,294,90], and is analogous to the popular Density Fitting (DF) procedure of electronic structure theory[324,67,299]. In this context, the decomposition is exact thanks to the closure properties of the polynomial-based Hermite functions. The integrals are now given as,

$$\langle ij|\hat{V}|i'j'\rangle = \sum_{AB} [ii'A][jj'B]G^{AB}, \tag{79}$$

where,

$$G^{AB} \equiv \iint_{\mathbb{R}^2} \mathrm{d}x_1 \; \mathrm{d}x_2 \; \chi_A(x_1)\chi_B(x_2)\hat{V}(x_1, x_2). \tag{80}$$

Thus, the fourth-order integral tensor is expressed as a product of second- and third-order tensors. Even though we have compressed the fourth-order tensor, this representation still precludes scaling reduction in "exchange-like" terms. A canonical example of such a term is the pairing field in Hartree-Fock-Bogoliubov theory,

$$\Delta_{ij} \equiv \sum_{i'j'} \langle ij|\hat{V}|i'j'\rangle \kappa_{i'j'} = \sum_{ABi'j'} [ii'A][jj'B]G^{AB}\kappa_{i'j'}, \tag{81}$$

where κ is the pairing tensor. Despite the factorization, computing this term still scales as $\mathcal{O}(M^4) = \mathcal{O}(M^{2ND})$.

The critical step in THC is to resolve the three-index overlap integral $[ii'A]$ to "unpin" the indices i and i' across some additional linear-scaling index P. That is, we seek a decomposition of the form $[ii'A] = \sum_P X_i^P X_{i'}^P Y_A^P$, where the range of P is $\mathcal{O}(M)$. Thanks to the choice of a polynomial basis, the overlap integral is exactly integrated by a $2M + 1$-node Gaussian quadrature (in this case, Gauss-Hermite)

defined by the nodes and weights $\{< x_P, w_P >\}$[2]. Therefore, the quadrature grid index provides a natural decomposition of the overlap integral,

$$[ii'A] = \sum_P w_P \psi_i(x_P)\psi_{i'}(x_P)\chi_A(x_P) = \sum_P X_i^P X_{i'}^P Y_A^P, \tag{82}$$

where $X_i^P \equiv \psi_i(x_P)$ and $Y_A^P \equiv w_P \chi_A(x_p)$. This is reminiscent of the discrete variable representation[103,61,177,19] or pseudospectral[80] techniques of chemical physics. Defining the intermediate $Z^{PQ} = \sum_{AB} Y_A^P G^{AB} Y_B^Q$, the full integral (76) is thus expressed as,

$$\langle ij|\hat{V}|i'j'\rangle = \sum_{PQ} X_i^P X_j^Q Z^{PQ} X_{i'}^P X_{j'}^Q. \tag{83}$$

This X-THC representation of the integral tensor is the key for the exact $\mathcal{O}(M^3) = \mathcal{O}(M^{ND+1})$ treatment of the pairing term, via several intermediate summations, indicated here by brackets for clarity,

$$\Delta_{ij} = \sum_{PQi'j'} X_i^P X_j^Q Z^{PQ} X_{i'}^P X_{j'}^Q \kappa_{i'j'} \tag{84}$$

$$= \sum_P X_i^P \left[\sum_Q X_j^Q \left[Z^{PQ} \left[\sum_{i'} X_{i'}^P \left[\sum_{j'} X_{j'}^Q \kappa_{i'j'} \right] \right] \right] \right].$$

2.3.2 Interpretation

At first glance, the Z operator is a mere mathematical intermediate, but there exists a much richer interpretation: it is a quantized renormalization of the coordinate-space representation of the potential operator $\hat{V}$. To see this, we first consider the continuous, renormalized potential operator $\bar{V}$, defined as,

$$\bar{V}(x_1, x_2) \equiv \sum_{AB} \chi_A(x_1)\chi_B(x_2) G^{AB}. \tag{85}$$

This operator is not equivalent to the original in physical space, i.e., $\hat{V}(x_1, x_2) \neq \bar{V}(x_1, x_2)$, yet the matrix elements of both operators are identical, i.e., $\langle ij|\hat{V}|i'j'\rangle = \langle ij|\bar{V}|i'j'\rangle$. The renormalized operator is simply the raw operator $\hat{V}$ with all components outside of the finite product space $\{\psi_i(x_1)\psi_{i'}(x_1)\} \Leftrightarrow \{\chi_A(x_1)\}$ projected out

in each coordinate. This projection is serendipitous: the coordinate-space integrand involving $\bar{V}$ and the products of basis functions are exactly resolved by the Gaussian quadrature for the auxiliary basis, while the corresponding integrand for $\hat{V}$ is not exact under any finite quadrature due to the presence of "alias" components outside of $\{\psi_i(x_1)\psi_{i'}(x_1)\}$. Applying the Gaussian quadrature, we can quantize the renormalized operator $\bar{V}$ to produce the discrete operator $\tilde{V}$, adding quadrature weights to account for the spatial contribution of each point,

$$\tilde{V}(x_1, x_2) \equiv w_P w_Q \delta(x_1 - x_P)\delta(x_2 - x_Q)\bar{V}(x_1, x_2). \tag{86}$$

As with $\bar{V}$, the matrix elements of $\tilde{V}$ are identical to those of $\hat{V}$. Integrating $\tilde{V}$ instead of $\hat{V}$ naturally exposes the X-THC factorization,

$$\begin{aligned} \langle ij|\hat{V}|i'j'\rangle &= \langle ij|\tilde{V}|i'j'\rangle \\ &= \iint \mathrm{d}x_1 \; \mathrm{d}x_2 \; \psi_i(x_1)\psi_j(x_2)\tilde{V}(x_1, x_2)\psi_{i'}(x_1)\psi_{j'}(x_2) \\ &= \sum_{PQ} X_i^P X_j^Q Z^{PQ} X_{i'}^P X_{j'}^Q. \end{aligned} \tag{87}$$

Here, the elements Z^{PQ} are simply the quantized values of the renormalized potential, with the weights rolled in, i.e., $Z^{PQ} = w_P w_Q \bar{V}(x_P, x_Q)$. An example involving a Gaussian potential in Hermite functions is shown in Figure 1. The renormalized potential (right) clearly shows the effects of projection from the raw potential (left). The locations of the quantization to Z^{PQ} (the positions at which $\bar{V}$ can be discretized in a lossless manner) are indicated with small white x's on the right.

This understanding of the Z operator reveals that while X-THC is built from DF and DVR techniques, the resultant supersedes both of the originals. In the context of local potentials and polynomial basis sets, DF is always exact, but does not provide separability of the i and i' indices, precluding scaling reductions. DVR techniques do provide separability, but are only exact when an infinite quadrature is used, for an arbitrary choice of local potential. By contrast, X-THC's particular merger of

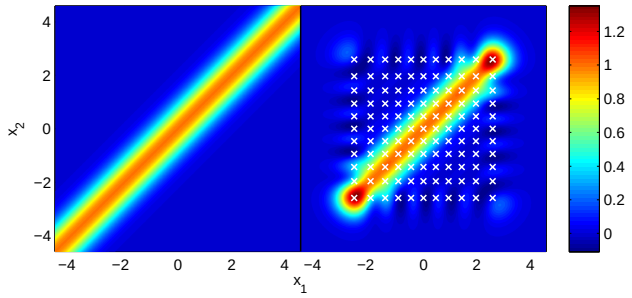

Figure 1: Example of the X-THC process for a one-dimensional, two-body Gaussian potential $\hat{V}(x_1, x_2) = \exp(-x_{12}^2)$ in Hermite functions $\{\psi_i(x)\}$ up to $M = 5$. Left: raw $\hat{V}(x_1, x_2)$. Right: renormalized, quantizable $\bar{V}(x_1, x_2)$. White x's indicate the collocation locations of the Gauss-Hermite quadrature to the quantized operator $\tilde{V}(x_1, x_2)$.

DF and DVR yields a perfect dealiasing renormalization within a finite quadrature, providing a decomposition that is both exact and separable for an arbitrary choice of local potential.

2.3.3 Generalized X-THC

The generalization of the one-dimensional, two-body, Hermite function example above to N-body potentials in D-dimensions and other choices of polynomial direct-product bases is straightforward. For X-THC to hold, the one-particle basis must be of the D-dimensional direct-product polynomial type, i.e., $\psi_i(\boldsymbol{r}) \equiv \prod_{\mu=1}^{D} P_{i_\mu}(r_\mu) v_\mu(r_\mu)$. In each dimension μ, P_{i_μ} is a polynomial of up to degree i_μ, and v_μ is an arbitrary weight function (analogous to the Gaussian term in the Hermite functions above). Such basis sets are widely used in atomic and nuclear many-body physics in various coordinate systems. Use of a direct-product polynomial basis automatically guarantees closure: for the $M_\mu + 1$ functions in the μ^{th} dimension, the span $< \psi_{i_\mu}(r_\mu)\psi_{i'_\mu}(r_\mu) >$ lies wholly inside a $2M_\mu + 1$-function auxiliary basis, defined by a set of polynomials orthogonal with respect to the weight $|v_\mu(r_\mu)|^4$. Additionally, all quadratic products of auxiliary functions are exactly integrated by a $2M_\mu + 1$-node Gaussian quadrature $\{< r_{P_\mu}, w_{P_\mu} >\}$ which can always be found, e.g., by the Golub-Welsch algorithm[91].

These properties allow for the X-THC factorization,

$$\langle i \dots n | \hat{V} | i' \dots n' \rangle = \tag{88}$$

$$\sum_{P \dots W} X_i^P \dots X_n^W Z^{P \dots W} X_{i'}^P \dots X_{n'}^W,$$

with each X_i^P being the direct product of the D underlying $X_{i_\mu}^{P_\mu}$. $Z^{P \dots W}$ is the generalization of Z^{PQ} to the case with N-body auxiliary integrals $G^{A \dots N}$.

Within the X-THC representation, the representative generalization of the pairing term, $\Delta_{i \dots n} \equiv \langle i \dots n | \hat{V} | i' \dots n' \rangle \kappa_{i' \dots n'}$, now scales as $\mathcal{O}(M_\mu^{ND+1})$, rather than $\mathcal{O}(M_\mu^{2ND})$, with no approximation or restriction on the form of the local, finite-range potential $\hat{V}$.

It is worth noting that common techniques to reduce the cost of treating exchange-like terms involve approximating the potential to be direct-product separable over N_w terms, e.g., by approximating the Coulomb operator as a sum of separable Gaussians[64,253]. This reduces the conventional or DF cost of forming the generalized pairing tensor to $\mathcal{O}(M_\mu^{ND+N})$. X-THC can be applied to this approximate separable potential, producing an $\mathcal{O}(M_\mu^{ND+1})$ implementation. However, the separable form gives no particular scaling advantage in the X-THC formalism, and can only reduce the prefactor and memory requirements. A more severe approximation is the invocation of a zero-range potential. This is typically formulated as a DVR-type quadrature in coordinate space, which can be exact depending on the form of the zero-range operator[63]. The asymptotic scaling of a pairing term involving a zero-range potential is $\mathcal{O}(M_\mu^{ND+1})$, due to the first or last transformation into or out of the grid index. Remarkably, this is the same asymptotic scaling as X-THC. The zero-range potential will generally have lower prefactor than X-THC (as there is only one grid coordinate in the zero-range potential), but the asymptotic scalings are identical, and thus the tractability limits should be comparable. A summary of the scaling reductions afforded with various factorization approaches and local potentials is shown in Table

1.

Table 1: Computational scalings for the pairing term of an arbitrary local potential in several approaches. M_μ is the order of the polynomial basis in the μ^{th} degree of freedom, and the potential is N-body in D dimensions. For simplicity, we consider the isotropic case where M_μ is the same in all dimensions in this comparison. N_w is the number of terms retained in a separable approximation to the potential.

Approach	General Local	Separable Local	Zero-Range
Conventional	$\mathcal{O}(M_\mu^{2ND})$	$\mathcal{O}(N_w M_\mu^{ND+N})$	$\mathcal{O}(M_\mu^{ND+1})$
X-THC	$\mathcal{O}(M_\mu^{ND+1})$	$\mathcal{O}(N_w M_\mu^{ND+1})$	$\mathcal{O}(M_\mu^{ND+1})$

2.3.4 Practical Demonstration

To illustrate the numerical equivalence and practical utility of the X-THC approach, a hybrid MATLAB/C++ code was developed to produce generalized pairing fields for D-dimensional, N-body forces in Hermite functions. A complete description of the code is presented in the supplemental material.

We have verified that the X-THC generalized pairing fields are exact within machine precision (as expected mathematically). Figure 2 shows the computational gains which can be achieved from the X-THC factorization using a representative example of $N = 2$ and $D = 1, 2, 3$. For a general local potential, X-THC is several orders of magnitude faster than conventional approaches for the largest M_μ studied here. When the potential is written in separable form, the X-THC scaling advantage is less dramatic, but X-THC becomes less costly for the largest M_μ used in Figure 2. The X-THC approach allows one to retain the general local potential and calculate the *exact* pairing tensor in similar (or even less) computational effort as with an *approximate* separable potential.

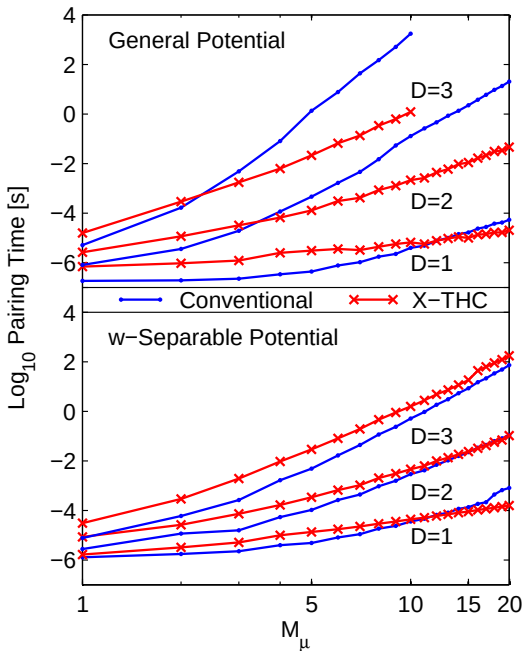

Figure 2: Wall times for pairing tensor formation as a function of M_μ for $N = 2$ (log-log scale). $N_w = 8$ for the separable potential.

2.4 Summary and Outlook

In this Letter, we have demonstrated that an exactly quantized renormalization of any local, finite-range N-body potential exists in any situation where the primary basis set may be composed of polynomial-based functions. This X-THC representation provides for substantial computational scaling reduction of contractions involving the local potential integral tensor, for instance by reducing the formation of a representative generalization of the pairing tensor from $\mathcal{O}(M_\mu^{2ND})$ to $\mathcal{O}(M_\mu^{ND+1})$.

In electronic structure, the concept of X-THC helps to codify and rationalize our existing Least-Squares Tensor HyperContraction (LS-THC) approximation for non-polynomial basis sets[216]. The least squares procedure introduced in that work actually performs an implicit renormalization of the potential. Since the basis sets used in our previous applications of LS-THC were not direct products of polynomials (but rather atom-centered Gaussian functions), the decomposition was necessarily an

approximation. As the X-THC limit is approached, the fidelity of the approximation will depend on both the basis set resemblance to a set of polynomial-based functions and the efficiency of the quadrature. The physical picture provided by X-THC's explicit renormalization process will almost certainly aid in the search for enhanced approximate THC recipes for non-polynomial basis sets.

In nuclear structure, the potential applications for X-THC are immediate and substantial. A crucial finding of this work is that the formal scaling of operations involving arbitrary local potential operators is identical to that of zero-range operators, without any loss in accuracy. This implies that the finite-range two-body Gogny potentials of nuclear DFT can immediately be applied with the same computational complexity as the more approximate zero-range Skyrme potentials. Tractability gains should be even more marked for three-body potentials, paving the way for Hamiltonian-based nuclear energy densities derived from effective, local, finite-range two- and three-nucleon pseudopotentials, which would be free of the current artifacts of zero-range nuclear DFT.

CHAPTER III

LEAST-SQUARES TENSOR HYPERCONTRACTION

The following is adapted from Ref. 216.

3.1 Abstract

The Least-Squares Tensor HyperContraction (LS-THC) representation for the Electron Repulsion Integral (ERI) tensor is presented. Recently, we developed the generic Tensor HyperContraction (THC) ansatz, which represents the fourth-order ERI tensor as a product of five second-order tensors [E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, *J. Chem. Phys.*, **137**, 044103 (2012)]. Our initial algorithm for the generation of the THC factors involved a two-sided invocation of overlap-metric density fitting, followed by a PARAFAC decomposition, and is denoted PARAFAC Tensor HyperContraction (PF-THC). LS-THC supersedes PF-THC by producing the THC factors through a least-squares renormalization of a spatial quadrature over the otherwise singular $1/r_{12}$ operator. Remarkably, an analytical and simple formula for the LS-THC factors exists. Using this formula, the factors may be generated with $\mathcal{O}(N^5)$ effort if exact integrals are decomposed, or $\mathcal{O}(N^4)$ effort if the decomposition is applied to density-fitted integrals, using any choice of density fitting metric. The accuracy of LS-THC is explored for a range of systems using both conventional and density-fitted integrals in the context of MP2. The grid fitting error is found to be negligible even for extremely sparse spatial quadrature grids. For the case of density-fitted integrals, the additional error incurred by the grid fitting step is generally markedly smaller than the underlying Coulomb-metric density fitting error. The present results, coupled with our previously published factorizations of MP2 and MP3, provide an efficient, robust $\mathcal{O}(N^4)$ approach to both methods. Moreover,

LS-THC is generally applicable to many other methods in quantum chemistry.

3.2 Introduction

The Electron Repulsion Integral (ERI) tensor is,[290]

$$(\mu\nu|\lambda\sigma) = \iint_{\mathbb{R}^6} \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{r_{12}}\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)\ \mathrm{d}^3r_1\ \mathrm{d}^3r_2, \tag{89}$$

where the indices μ, ν, λ, and σ refer to atomic orbital (AO) basis functions. Written in chemists' notation as above, the ERI tensor is easily seen to represent the Coulomb interaction between the generalized charge densities $\rho_{\mu\nu}(\vec{r}_1) = \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)$ and $\rho_{\lambda\sigma}(\vec{r}_2) = \phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)$.

The ERI tensor is a formidable source of complexity in electronic structure theory, due to the immense computational cost of integral generation, the heavy storage requirements, and the ungainly presence of the ERIs in tensor contraction operations. To reduce this complexity, many authors have developed numerical approximation techniques ranging from sparsity considerations[105,167,191] to multipole moment approaches[322,323,287] to density fitting (DF, also known as resolution of the identity or RI)[324,67,68,299,245,144,312] or Cholesky decomposition (CD)[24,254,151,8,6] compression to pseudospectral (PS) representation.[80,81,82,83,251,250,93,188,190,187,187,85,150,208,162]

Of these techniques, density fitting/Cholesky decomposition and pseudospectral representation are particularly relevant to this study. Density fitting,[324,67,68,299,245,144,312] which has many conceptual similarities to Cholesky decomposition,[24,254,151,8,6] exploits the fact that the generalized densities $\rho_{\mu\nu}(\vec{r}_1) = \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)$ are often largely redundant and can be approximately represented with an auxiliary basis $\{\chi_A(\vec{r}_1)\}$, comprised of $\mathcal{O}(N)$ functions,

$$\phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1) \approx d^A_{\mu\nu}\chi_A(\vec{r}_1). \tag{90}$$

Here $d^A_{\mu\nu}$ are the density fitting coefficients, which are typically determined by least-squares fitting with respect to the $1/r_{12}$ fitting metric (the Coulomb metric), leading

to a factorization of the form,

$$(\mu\nu|\lambda\sigma) \approx (\mu\nu|A)(A|B)^{-1}(B|\lambda\sigma) = (\mu\nu|A)(A|C)^{-1/2}(C|B)^{-1/2}(B|\lambda\sigma) = \beta^C_{\mu\nu}\beta^C_{\lambda\sigma}. \tag{91}$$

The Einstein summation convention is used here and throughout this paper, i.e. there is an implicit sum over the repeated indices A, B, and C, which represent auxiliary basis functions. The integrals are defined as,

$$(\mu\nu|A) = \iint_{\mathbb{R}^6} \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{r_{12}}\chi_A(\vec{r}_2)\ \mathrm{d}^3r_1\ \mathrm{d}^3r_2, \tag{92}$$

and,

$$(A|B) = \iint_{\mathbb{R}^6} \chi_A(\vec{r}_1)\frac{1}{r_{12}}\chi_B(\vec{r}_2)\ \mathrm{d}^3r_1\ \mathrm{d}^3r_2. \tag{93}$$

The factorization of Equation 91 exploits the low-rank interactions between the generalized densities, and is able to reduce the fourth-order ERI tensor to the product of two third-order tensors by separating $\mu\nu$ from $\lambda\sigma$. However, the indicies μ and ν (and similarly λ and σ) remain "pinned" to each other, preventing reduced scaling factorizations of "exchange-like" terms, such as the exchange potential of Hartree-Fock theory,

$$K^s_{\mu\nu} = (\mu\lambda|\sigma\nu)D^s_{\lambda\sigma}, \tag{94}$$

where $D^s_{\lambda\sigma}$ is the one-particle density matrix corresponding to electrons of spin s. As with all exchange-like terms, the contraction indices for this term (λ and σ) appear on both sides of the chemists' ERI, precluding an $\mathcal{O}(N^3)$ factorization with density fitting, because the μ and λ (and likewise σ and ν) indices remain pinned in the DF approximation.

A conceptual improvement over density fitting is the pseudospectral method,[80,81,82,83,251,250,93,188,1] which integrates over one of the electronic coordinates using a numerical quadrature in physical space (with gridpoints $\vec{r}_P$ and quadrature weights w_P), leading to,

$$(\mu\nu|\lambda\sigma) \approx w_P R^P_\mu R^P_\nu A^P_{\lambda\sigma}. \tag{95}$$

The collocation matrix in this expression is defined as,

$$R_\mu^P = \phi_\mu^P(\vec{r}_P), \tag{96}$$

and the grid representation of the Coulomb potential is,

$$A_{\lambda\sigma}^P = \int_{\mathbb{R}^3} \frac{1}{r_{2P}} \phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)\ \mathrm{d}^3 r_2. \tag{97}$$

One aspect of the pseudospectral method which distinguishes it from previous quadrature methods used in electronic structure theory is replacement of one of the collocation matrices by a least-squares fitting matrix Q that guarantees the numerical procedure will reproduce the matrix of overlap integrals,

$$(\mu\nu|\lambda\sigma) \approx Q_\mu^P R_\nu^P A_{\lambda\sigma}^P, \tag{98}$$

where Q is determined as,

$$Q_\mu^P = S_{\mu\xi} \left[R_\xi^Q w_Q R_\eta^Q \right]^{-1} R_\eta^P w_P. \tag{99}$$

Here $S_{\mu\xi}$ is the overlap metric between basis functions ϕ_μ and ϕ_ξ. This factorization is able to separate $\mu\nu$ from $\lambda\sigma$, and *also* "unpins" the indices μ and ν, thus expressing the ERI tensor as a product of one third-order and two second-order tensors. This finally allows for the exchange potential to be factorized so that it can be evaluated in $\mathcal{O}(N^3)$ operations,

$$K_{\mu\nu}^s = (\mu\lambda|\sigma\nu) D_{\lambda\sigma}^s \approx Q_\mu^P \left[A_{\sigma\nu}^P \left[R_\lambda^P D_{\lambda\sigma}^s \right] \right]. \tag{100}$$

This unpinning of indices is an extremely desirable property, as it allows factorization of further equations that involve contraction of the integrals (for example, with density matrices or wavefunction coefficients).[187] Unfortunately, the pseudospectral method as described so far does not provide sufficient accuracy for electronic structure applications. Successful applications have required extensive dealiasing and analytic integral corrections.[81,93,208] The results can also be quite sensitive to the grids which

are used. In practice, this means that successful applications of the pseudospectral method require the development and optimization of both atomic grids and dealiasing sets. Furthermore, many of the two-electron integrals need to be calculated analytically (e.g., integrals where two or more of the basis functions are located on the same atomic center). Exploiting the factorization when some of the integrals need to be treated as full four-index quantities requires very careful programming.

Recently, we developed the PARAFAC Tensor HyperContraction (PF-THC) representation, which was the first method to provide full separability of the four indices in the ERI tensor in $\mathcal{O}(N^4)$ effort.[121] The PF-THC representation of the ERI tensor is,

$$(\mu\nu|\lambda\sigma) \approx X_\mu^P X_\nu^P Z^{PQ} X_\lambda^Q X_\sigma^Q. \tag{101}$$

We consider this to be the characteristic factorization of all variants of THC. Here, X and Z are second-order tensors, with P and Q each being a "decomposition index," analogous to the grid point indices in pseudospectral or the auxiliary basis function indices in density-fitting approaches. In general, the range of these decomposition indices should be proportional to the size of the basis set (with a modest proportionality factor) if the approach is to be useful. In the PF-THC method, the symmetric overlap-metric density fitting approximation is first invoked,[299]

$$(\mu\nu|\lambda\sigma) \approx (\mu\nu A)(AB)^{-1}(B|C)(CD)^{-1}(D\lambda\sigma), \tag{102}$$

where the new integrals are,

$$(\mu\nu A) = \int_{\mathbb{R}^3} \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\chi_A(\vec{r}_1)\ \mathrm{d}^3 r_1, \tag{103}$$

and,

$$(AB) = \int_{\mathbb{R}^3} \chi_A(\vec{r}_1)\chi_B(\vec{r}_1)\ \mathrm{d}^3 r_1. \tag{104}$$

The three-center overlap integrals $(\mu\nu A)$ are then separated by a PARAFAC decomposition,[42] which is obtained in practice by an iterative alternating least-squares

algorithm,[281]

$$(\mu\nu A) \approx X_{\mu}^{P} X_{\nu}^{P} Y_{A}^{P}, \tag{105}$$

where P is the new decomposition index, which is implicit in the PARAFAC procedure. After forming the X and Y factors via PARAFAC, the intermediate Z matrix is built as,

$$Z^{PQ} = Y_{A}^{P} (AB)^{-1} (B|C) (CD)^{-1} Y_{D}^{Q}, \tag{106}$$

which completes the factorization above. The factorization was used to produce an implementation of MP2 *and* MP3[200] with demonstrated scaling of $\mathcal{O}(N^4)$. This should be compared with the usual scaling behaviors of MP2 and MP3: $\mathcal{O}(N^5)$ and $\mathcal{O}(N^6)$, respectively. Furthermore, we showed that this factorized form yielded chemical accuracy in the energies when restricting the range of the decomposition index P to twice the number of functions in the auxiliary density-fitting basis set.

The existence of a THC factorization which can be formed in $\mathcal{O}(N^4)$ is sufficient to provide significant scaling gains for many methods in electronic structure theory; PF-THC already realizes this goal. However, several of the technical details in PF-THC are improvable. Most obviously, the requirement of a generic PARAFAC decomposition of the three-index overlap integral tensor mandates the use of an implicit, iterative algorithm. While the prefactor of such a PARAFAC algorithm could doubtless be improved by the implementation of a more advanced method than the alternating least-squares algorithm used in our last paper, a highly desirable feature of an alternative THC factorization would be the presence of an explicit factorization. Besides greatly attenuating the overhead of the THC factorization, this property would also facilitate the computation of analytic gradients and response properties. Additionally, the PF-THC method of our last paper started from a double-overlap density fitting approximation, which is generally found to be less accurate (for a given number of auxiliary basis functions) than the preferred Coulomb-metric density fitting. Thus, an alternative THC factorization that starts from either exact integrals

or Coulomb-metric density fitting is highly desirable.

Herein, we develop Least-Squares Tensor HyperContraction (LS-THC), which provides an accurate, efficient alternative to PF-THC. LS-THC begins by invoking a numerical quadrature in *both* electronic coordinates of the ERI, effectively attempting a two-sided application of pseudospectral representation. Such a quadrature is always numerically singular due to the presence of zeros in the denominator of $1/r_{12}$, so the quadrature is renormalized by least-squares fitting against the desired electron repulsion integrals. Remarkably, an analytical formula for the renormalized quadrature is obtained. This formula requires $\mathcal{O}(N^5)$ operations if exact ERIs are fitted, or $\mathcal{O}(N^4)$ operations if the ERIs are instead approximated by density fitting. The preferred Coulomb metric may be used in place of the overlap metric for this step. We detail an algorithm for the formation of the LS-THC factors using either exact integrals or DF integrals. The algorithm is flexible enough to allow distinct fits for subsets of the two-electron integrals, which can increase the accuracy in correlated methods. We also discuss the interpretation of LS-THC and PF-THC as two examples of a much larger class of potential THC-type methods. Practically, we demonstrate the error of LS-THC within our $\mathcal{O}(N^4)$ THC-MP2 code, and find the grid fitting error of LS-THC to be of the same order or or smaller than that of the Coulomb-metric DF approximation, even when using coarse atom-centered quadrature grids ($\approx$ 150 points/atom, i.e. $\approx$ 20 points/basis function). Finally, we close by commenting on some of the potential applications for LS-THC within electronic structure theory.

A rather large number of classes of indices appear in this work, so we summarize the majority below,

- $\mu, \nu, \lambda, \sigma, \xi, \eta$: Atomic Orbital (AO) basis indices, which in this work are atom-centered contracted Gaussian functions. The size of these indices is given by $N_\mu \propto N$, where N is the measure of system size (electrons or atoms).

- A, B, C: Auxiliary basis indices, which for DF are atom-centered Gaussian

functions, and for CD are implicit arbitrary-center combinations of contracted Gaussian functions. The size of these indices is given by $N_A \propto N_\mu$.

- P, Q: "Decomposition" indices, which for PF-THC are an implicit basis found by PARAFAC decomposition, and for LS-THC are points in a physical space quadrature grid. The size of these indices is given by $N_P \propto N_\mu$.

- i, j: Occupied Molecular Orbital (MO) basis indices, which, in this work, always exclude core orbitals. The size of these indices is given by $N_{\text{occ}} \propto N_\mu$. MO-basis terms with this index are preferred, as $N_{\text{occ}} \ll N_{\text{vir}} \sim N_\mu$, given reasonably large basis sets.

- a, b: Virtual MO basis indices. The size of the these indices is given by $N_{\text{vir}} \propto N_\mu$. MO-basis terms with this index are avoided where possible, as $N_{\text{vir}} \sim N_\mu \gg N_{\text{occ}}$, given reasonably large basis sets.

- p, q, r, s: Arbitrary AO- or MO-basis indices or subsets thereof, used to generalize LS-THC to AO or MO ERIs of a specific type. The size of these indices N_p depends on the class of orbital used, but is always $\propto N_\mu$.

Although all of the indices should have ranges which are $\propto N_\mu$, they each have different prefactors. We expect that the size of the ranges can be ordered roughly as $N_\mu^2 \gg N_P > N_A > N_\mu > N_{\text{vir}} > N_{\text{occ}}$. Finally, any mention of density fitting from this point forward refers to the the preferred Coulomb-metric type; density fitting of the two-sided overlap metric type was only used in the previous PF-THC method.

3.3 Theory

3.3.1 Least-Squares THC

Naïvely, we might expect to find a THC-like representation of the ERI tensor by two-sided quadrature as,

$$(\mu\nu|\lambda\sigma) \approx R_\mu^P R_\nu^P \underbrace{w_P \frac{1}{r_{PQ}} w_Q}_{Z^{PQ}} R_\lambda^Q R_\sigma^Q, \tag{107}$$

where the collocation matrix is,

$$R_\mu^P = \phi_\mu(\vec{r}_P). \tag{108}$$

Unfortunately, for any finite quadrature using the same grid on both sides, this approach will fail because of the Coulomb singularity for grid points where $\vec{r}_P = \vec{r}_Q$ (regardless of the choice for the quadrature weights w).

Therefore, we abandon a direct quadrature, and instead seek an implicit operator Z which is optimal in a least-squares sense. Given a primary basis $\{\phi_\mu\}$ and the factor R_μ^P, we wish to minimize,

$$O \equiv \frac{1}{2} \left\| (\mu\nu|\lambda\sigma) - R_\mu^P R_\nu^P Z^{PQ} R_\lambda^Q R_\sigma^Q \right\|^2, \tag{109}$$

with respect to the elements of Z. Below, the joint collocation is used for convenience,

$$R_{\mu\nu}^P = R_\mu^P R_\nu^P. \tag{110}$$

The stationary condition is,

$$\frac{\partial O}{\partial Z^{P'Q'}} = -R_{\mu\nu}^{P'} \Delta_{\mu\nu\lambda\sigma} R_{\lambda\sigma}^{Q'} = 0, \tag{111}$$

where the residual is,

$$\Delta_{\mu\nu\lambda\sigma} = (\mu\nu|\lambda\sigma) - R_{\mu\nu}^P Z^{PQ} R_{\lambda\sigma}^Q. \tag{112}$$

Inserting Equation 112 into Equation 111 and expanding yields,

$$R_{\mu\nu}^{P'} (\mu\nu|\lambda\sigma) R_{\lambda\sigma}^{Q'} = R_{\mu'\nu'}^{P'} R_{\mu'\nu'}^P Z^{PQ} R_{\lambda'\sigma'}^Q R_{\lambda'\sigma'}^{Q'}. \tag{113}$$

Leading finally to the analytical result,

$$Z^{PQ} = \left[R^{P}_{\mu'\nu'} R^{P'}_{\mu'\nu'}\right]^{-1} R^{P'}_{\mu\nu}(\mu\nu|\lambda\sigma) R^{Q'}_{\lambda\sigma} \left[R^{Q'}_{\lambda'\sigma'} R^{Q}_{\lambda'\sigma'}\right]^{-1}. \tag{114}$$

If N_P is at least $N_\mu(N_\mu+1)/2$ (the maximum rank of the space of quadratic products $\phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)$), and if the joint collocation matrices are full rank, the two-electron integral is exactly resolved, as,

$$R^{P}_{\mu\nu} \left[R^{P}_{\mu'\nu'} R^{P'}_{\mu'\nu'}\right]^{-1} R^{P'}_{\mu''\nu''} = \delta_{\mu\nu,\mu''\nu''}. \tag{115}$$

This property can be easily verified by substituting the full SVD of the joint collocation matrix into the expression on the left-hand side. This provides an interesting interpretation for LS-THC as resolving the identity $\delta_{\mu\nu,\mu''\nu''}$ approximately, in a manner that guarantees separability of the indices.

Note that while this procedure may be conceptually understood to start from a two-sided invocation of the pseudospectral approximation (which one might expect to be even less accurate than a one-sided invocation), the least-squares fitting procedure endows the factors with explicit knowledge of the two-body operator used. This is not present in even one-sided pseudospectral, so LS-THC may prove to be markedly more accurate than pseudospectral methods, albeit at higher cost of factor determination.

A more thorough derivation which does not rely on the generalized Einstein convention is available in the appendix.

3.3.2 Choice of Grid

Having outlined the generic LS-THC algorithm, a physical-space quadrature grid (both gridpoints and their associated quadrature weights) must be provided to complete the method. The grids selected for this study are of the atom-centered Becke type,[20,205,297,88] with modified radial quadratures that more properly reflect the distribution of radial spatial ranges in the products of basis functions. Pruning is applied to reduce the orders of the spherical Lebedev quadratures near the nuclei or tails, resulting in grids with approximately four times as many points as the number of auxiliary

functions in the corresponding MP2 auxiliary basis. A more extensive study identifying even sparser grids for use in LS-THC is forthcoming. However, the minimally optimized grids used in this study already exhibit negligible errors, indicating that the LS-THC fitting procedure is quite robust in dealing with a nearly arbitrary choice of spatial quadrature grid. These grids are included in supplemental material for all systems considered in this study.

3.3.3 Collocation Weighting

In practice, the physical space metric matrix S, defined as,

$$S^{PP'} \equiv R^{P}_{\mu'\nu'} R^{P'}_{\mu'\nu'}, \tag{116}$$

can be numerically ill-conditioned for two reasons. First, the grid may include points where all basis functions have negligible magnitude, e.g. in regions far from all the nuclei in the molecule. Second, there could be true numerical redundancy between linear combinations of points, which should be eliminated. Typically, pseudoinversion is invoked to deal with the latter problem. However, simply performing pseudoinversion on the above quantity also discards information about the "tail regions" of the basis functions. Even though the absolute magnitude of the basis function values at the outermost grid points may be small, nevertheless the relative values for different basis functions are meaningful. This information is valuable and it would be preferable not to discard it. Thus, we have found it useful to re-weight the collocation matrices to avoid penalizing points in the decaying regions due to their small absolute value. We then use pseudoinversion to remove the remaining near linear dependencies.

Up to this point, we have not used the weights associated with the grid points for quadrature grids. There are a number of ways to include these quadrature weights if desired. Perhaps the simplest is to refer to Equation 101. As the collocation matrices appear four times, associating the quadrature weights with the collocation matrices

as,

$$X_\mu^P = \sqrt[4]{w_P} R_\mu^P, \tag{117}$$

ensures that the weight associated with the diagonal term ($P = Q$) is the expected value. We find that incorporating the quadrature weights in the collocation matrices in this way significantly improves the condition number of the grid metric, and this definition is used in the results below. Here and throughout, we use X to denote the generalized re-weighted collocation matrix. With this weighting applied, small eigenvalues in the matrix S correspond to true near linear dependencies, and may be safely ignored below some cutoff which is slightly larger than the machine epsilon. In this work, we use a relative cutoff ϵ of 10^{-10} in the pseudoinversion,

$$[S^{PP'}]^{-1} \equiv U_{Pi} \lambda_i^{-1} U_{P'i} : \; S^{PP'} = U_{Pi} \lambda_i U_{P'i}, \text{ and } \lambda_i > \epsilon \lambda_1. \tag{118}$$

Here λ_i is the i-th eigenvalue (all ≥ 0), and λ_1 is the largest eigenvalue. The choice of X and S^{-1} described above is used for all computations in this work.

3.3.4 Low-Rank Formation

In the construction of the LS-THC factors, the first step is the formation of the "grid metric" inverse of S,

$$\left[S^{PP'}\right]^{-1} = \left[X_{\mu'\nu'}^P X_{\mu'\nu'}^{P'}\right]^{-1}. \tag{119}$$

The formation of S scales as $\mathcal{O}(N_\mu N_P^2) \propto \mathcal{O}(N^3)$, as we can exploit the definition of the joint collocation matrices, $X_{\mu\nu}^P = X_\mu^P X_\nu^P$,

$$S^{PP'} = \left[X_{\mu'}^P X_{\mu'}^{P'}\right] \left[X_{\nu'}^P X_{\nu'}^{P'}\right]. \tag{120}$$

The inversion scales formally as $\mathcal{O}(N_P^3) \propto \mathcal{O}(N^3)$. In practice, it might be possible to reduce this considerably by exploiting sparsity, but we have not explored this.

Having discussed the formation and inversion of the grid metric matrices, the next step needed for construction of Z^{PQ} in Equation 114 is the formation of what we will

call the E matrix,

$$E^{PQ} = X^P_{\mu\nu}(\mu\nu|\lambda\sigma)X^Q_{\lambda\sigma}. \tag{121}$$

This step scales as $\mathcal{O}(N^4_\mu N_P) \propto \mathcal{O}(N^5)$. Although such scaling might be acceptable for some methods (where it replaces the integral transformation), it is surely not acceptable for MP2 theory, as we have already developed a $\mathcal{O}(N^4)$ THC-MP2 code, presuming the THC factors are available. Conveniently, lower scaling can be easily achieved in this step by replacing the true ERIs in Equation 101 by their density-fitted approximations (in the Coulomb metric, as defined in Equation 91),

$$(\mu\nu|\lambda\sigma) \approx \beta^C_{\mu\nu}\beta^C_{\lambda\sigma}, \tag{122}$$

we may form the E matrix as,

$$E^{PQ} = \left[X^P_{\mu\nu}\beta^C_{\mu\nu}\right]\left[\beta^C_{\lambda\sigma}X^Q_{\lambda\sigma}\right]. \tag{123}$$

This now scales as $\mathcal{O}(N^2_\mu N_A N_P) \propto \mathcal{O}(N^4)$. Note that since we are fitting the DF ERIs in place of the exact ERIs, we expect the error to be bounded from below by the underlying, but typically quite reasonable, DF error. Moreover, if additional accuracy is required, but is limited by the DF error, the DF auxiliary basis may be increased significantly, with increased computational cost incurred in the formation of the LS-THC factorization, but not in the subsequent application of the LS-THC factors.

The final step is the formation of the Z operator as,

$$Z^{PQ} = \left[S^{PP'}\right]^{-1} E^{P'Q'} \left[S^{Q'Q}\right]^{-1}. \tag{124}$$

This requires $\mathcal{O}(N^3_P) \propto \mathcal{O}(N^3)$ operations.

3.3.5 MO-Basis Formation

In practice, it is often advantageous to perform the LS-THC fitting in the same basis as the target method for LS-THC factorization, rather than fitting the AO-basis

integrals and transforming the factors X. For example, in MP2, only integrals of the type $(ia|jb)$ appear. This yields two potential advantages.

1. The resultant Z operator will be tuned specifically to minimize the 2-norm of the residual for the target ERIs, yielding enhanced accuracy by excluding contributions to the residual from ERIs outside of the target space. Note that this means that Z need not always be positive definite, nor even symmetric. In fact, one could imagine using a different quadrature grid for each class of orbital products, in which case the Z operators need not even be square.

2. The smaller size of one or more of the target indices with respect to N_μ (for example, i) can be used to reduce the prefactor of the formation of the E matrix in either exact or DF LS-THC.

In the computational algorithm section below, we will use the generic MO-basis indices p, q, r, and s to denote that LS-THC fitting can be performed in an arbitrary basis which spans the subset of ERIs required. In the practical demonstration of MP2 accuracy later, we will use LS-THC to fit only the integrals required in frozen-core MP2, i.e., $(ia|jb)$, where the innermost core occupied orbitals are excluded.

3.3.6 Computational Algorithm

Here, we detail efficient computational algorithms for the formation of the LS-THC factors for a given set of generic integrals $(pq|rs)$, which are obtained from the AO-basis as,

$$(pq|rs) = C^p_\mu C^q_\nu (\mu\nu|\lambda\sigma) C^r_\lambda C^s_\sigma. \tag{125}$$

The C matrices are typically the Hartree-Fock orbital coefficients, though there is certainly no restriction to this case. For our purposes, we will assume that the indices are chosen so that $p \leq q$ and $r \leq s$ to minimize the floating pointer operation (FLOP) count in the formation of the E matrix, e.g., that the relevant integrals for MP2 are ordered $(ia|jb)$ instead of $(ai|bj)$.

Algorithm 1 details the generic formation of the LS-THC factorization of the ERIs. The choice of exact or DF ERIs is buried in the subroutines to form the projection tensor E, which also contain the rate limiting $\mathcal{O}(N^5)$ or $\mathcal{O}(N^4)$ operations, respectively. All other operations require at most $\mathcal{O}(N^2)$ memory and $\mathcal{O}(N^3)$ FLOPs, with the eigendecomposition for the pseudoinversion of S expected to dominate. The minimum required memory is essentially $3N_P^2$ double precision variables.

Algorithm 1 Generic LS-THC algorithm.

procedure LS_THC($\{\phi_\mu\}$,X_μ^P, C_μ^p, C_μ^q, C_μ^r, C_μ^s)
  $X_t^P = C_\mu^t X_\mu^P,\ t \in \{p,q,r,s\}$ ▷ $\mathcal{O}(N_\mu N_p N_P)$
  $E^{P'Q'}$ = BUILD_E_DF(...) or BUILD_E_EXACT(...) ▷ $\mathcal{O}(N^4)$ [DF] or $\mathcal{O}(N^5)$ [Exact]
  $S^{PP'} = \left[X_p^P X_p^{P'}\right]\left[X_q^P X_q^{P'}\right]$ ▷ $\mathcal{O}(N_p N_P^2)$
  $\left[S^{PP'}\right]^{-1} \equiv U_{Pi}\lambda_i^{-1}U_{P'i}\ \forall\ \lambda_i > \epsilon\lambda_1$ ▷ Pseudoinversion, $\mathcal{O}(N_P^3)$
  $S^{QQ'} = \left[X_r^Q X_r^{Q'}\right]\left[X_s^Q X_s^{Q'}\right]$
  $\left[S^{QQ'}\right]^{-1} \equiv U_{Qi}\lambda_i^{-1}U_{Q'i}\ \forall\ \lambda_i > \epsilon\lambda_1$ ▷ Shallow copy $\left[S^{PP'}\right]^{-1}$ if $pq = rs$
  $Z^{PQ} = \left[S^{PP'}\right]^{-1} E^{P'Q'} \left[S^{Q'Q}\right]^{-1}$ ▷ $\mathcal{O}(N_P^3)$
  return $< X_p^P, X_q^P, Z^{PQ}, X_r^Q, X_s^Q >$
end procedure

Algorithm 2 details one possible algorithm for the formation of the E matrix in $\mathcal{O}(N^5)$ operations using exact integrals and a conventional disk-based AO-to-MO transformation. This algorithm requires $\mathcal{O}(N^2)$ memory and $\mathcal{O}(N^5)$ FLOPS, with GEneralized Matrix Multiplication (GEMM) routines applicable for all of the FLOP-heavy steps. As the $\mathcal{O}(N^4)$ DF algorithm is preferred in practice, this algorithm is not heavily optimized. Many other variants involving direct algorithms with various levels of transformation exist, but are not pursued here.

Algorithm 3 shows the steps necessary to form the E matrix in $\mathcal{O}(N^4)$ using DF integrals. Note that the DF fitting is postponed to the very end of the computation, after the orbital indices have been projected onto the grid index. This avoids storage of an $\mathcal{O}(N^3)$ tensor, which would involve both disk I/O and an expensive out-of-core transposition. As it stands, the algorithm requires $\mathcal{O}(N^2)$ memory and

Algorithm 2 Conventional algorithm for $E^{P'Q'}$.

procedure BUILD_E_EXACT($\{\phi_\mu\}$, X_p^P, X_q^P, X_r^P, X_s^P, C_μ^p, C_μ^q, C_μ^r, C_μ^s)
  Form $(pq|rs) = C_\mu^p C_\nu^q (\mu\nu|\lambda\sigma) C_\lambda^r C_\sigma^s$ ▷ Disk AO-to-MO transform, $\mathcal{O}(N_\mu^4 N_p)$
  Initialize $E^{P'Q'} = 0$
  for all blocks rs_{block}, bounded by memory **do**
    Read $(pq|rs)$ ▷ $\mathcal{O}(N_p^4)$
    for all blocks P'_{block}, bounded by memory **do**
      $\Gamma_{rs}^{P'} = X_p^{P'} X_q^{P'} (pq|rs)$ ▷ Rate limiting, $\mathcal{O}(N_p^4 N_P)$
      $E^{P'Q'} += \Gamma_{rs}^{P'} X_r^{Q'} X_s^{Q'}$ ▷ $\mathcal{O}(N_p^2 N_P^2)$
    end for
  end for
  return $E^{P'Q'}$
end procedure

$\mathcal{O}(N^4)$ FLOPs, with GEMM routines applicable for all of the FLOP-heavy steps. The minimum required memory is essentially $3N_A^2$ double precision variables.

Algorithm 3 Density-fitted algorithm for $E^{P'Q'}$.

procedure BUILD_E_DF($\{\phi_\mu\}$, X_p^P, X_q^P, X_r^P, X_s^P, C_μ^p, C_μ^q, C_μ^r, C_μ^s)
  $L_\nu^{P'} = C_\nu^q X_q^{P'}$ ▷ $\mathcal{O}(N_\mu N_p N_P)$
  $L_\sigma^{Q'} = C_\sigma^s X_s^{P'}$ ▷ $\mathcal{O}(N_p N_\mu N_P)$
  for all blocks A_{block} bounded by memory **do**
    Build $(A|\mu\nu)$ for the current block ▷ $\mathcal{O}(N_\mu^2 N_A)$
    $(A|p\nu) = (A|\mu\nu) C_\mu^p$ ▷ $\mathcal{O}(N_\mu^2 N_p N_A)$
    $(A|pP') = (A|p\nu) L_\nu^{P'}$ ▷ Rate-Limiting, $\mathcal{O}(N_\mu N_p N_A N_P)$
    $M_A^{P'} = (A|pP') X_p^{P'}$ ▷ $\mathcal{O}(N_p N_A N_P)$
    $(B|r\sigma) = (B|\lambda\sigma) C_\lambda^r$
    $(B|rQ') = (B|r\sigma) L_\sigma^{Q'}$
    $M_B^{Q'} = (B|rQ') X_r^{Q'}$ ▷ Shallow copy $M_A^{P'}$ if $pq = rs$
  end for
  Build $(A|B)$ ▷ $\mathcal{O}(N_A^2)$
  $(A|B)^{-1} \equiv U_{Ai} \lambda_i^{-1} U_{Bi} : \forall\ \lambda_i > \epsilon\lambda_1$ ▷ Pseudoinversion, $\mathcal{O}(N_A^3)$
  $E^{P'Q'} = M_A^{P'} (A|B)^{-1} M_B^{Q'}$ ▷ $\mathcal{O}(N_A N_P^2)$
  return $E^{P'Q'}$
end procedure

3.3.7 Optimal THC

It should be noted that the factor tensor X_μ^P need not be related to a physical-space grid, nor even defined explicitly at all. In fact, one might imagine that the ideal case

would be, for a given size N_P, to find the completely generic X and Z such that,

$$O \equiv \frac{1}{2} \left\| (\mu\nu|\lambda\sigma) - X_\mu^P X_\nu^P Z^{PQ} X_\lambda^Q X_\sigma^Q \right\|^2, \tag{126}$$

is minimized, thus producing the optimal factorization in the 2-norm sense. For a given N_P, this Optimal Tensor HyperContraction (Opt-THC) decomposition is expected to be more accurate than the above LS-THC, as the X factors provide additional degrees of freedom over those in Z. Unfortunately, the coupled determination of the X and Z factors lies outside the scope of even PARAFAC decomposition, and therefore would require a custom (and likely quite costly) iterative $\mathcal{O}(N^5)$ algorithm to form. Nonetheless, Opt-THC is useful as a pedagogical tool, as it represents essentially the maximum attainable accuracy and generality within the low-rank THC structure of the ERI tensor. All other flavors of THC are specific derivatives which remove one or more degrees of freedom from Opt-THC. LS-THC restricts the X factor to be a predefined collocation factor corresponding to a physical-space grid, but then allows the Z operator to be fully unconstrained and determined in a least-squares manner. PF-THC uses both symmetric overlap-metric density fitting and PARAFAC, and, as such, does not obtain the least-squares property for either X or Z. Yet another example that falls within the framework of Opt-THC is the seminal work of Hackbusch and co-workers.[27] Here the factorization is of the form,

$$(\mu\nu|\lambda\sigma) \approx X_\mu^P X_\nu^P X_\lambda^P X_\sigma^P, \tag{127}$$

where the factors X are now the free variables. This is equivalent to Opt-THC with the restriction that the factor Z^{PQ} be the identity δ_{PQ}. Therefore, this technique can be viewed as a doppleganger of LS-THC, in which the Z factor is preselected and the X factor is determined by least-squares. Unlike LS-THC, no analytic form exists for these X factors; their determination requires a four-way PARAFAC decomposition, severely limiting the size of systems to which the method may be applied.

Rank-reduction ideas like the THC family of approximations could also be useful in the context of wavefunction coefficients. For example, the amplitudes for double excitations in configuration interaction and coupled cluster theories have a similar four-index structure as that of the ERIs. Early work by Olsen, et al.[211] applied a tensor factorization inspired by natural orbital decompositions to single- and double-excitation configuration interaction. More recently, Bell, et al.[25] used high-order singular value decomposition for the T2 amplitudes in coupled cluster to obtain a factorization similar in spirit to THC, albeit at steep computational cost. In the future, we will explore the possibility of combining the THC factorization for both ERIs and wavefunction coefficients, and we can expect that the Opt-THC framework will provide important guidance in this endeavor.

Finally, we note that there may exist other THC variants which have advantages over LS-THC for certain applications. Additionally, alternative choices of X may exist within LS-THC that are superior to collocation matrices.

3.4 Results

3.4.1 Computational Details

The LS-THC algorithms detailed above were implemented in a development version of the open-source PSI4 quantum chemistry package.[298] The accuracy of the LS-THC method is established by comparisons of LS-THC-MP2, DF-MP2, and LS-THC-DF-MP2 energies for a range of chemical systems. Here, LS-THC-MP2 refers to THC-MP2 with LS-THC performed on the exact integrals (LS-THC-MP2 is thus an approximation to MP2), whereas LS-THC-DF-MP2 refers to THC-MP2 with LS-THC performed on the density-fitted integrals (LS-THC-DF-MP2 is thus an approximation to DF-MP2). In particular, we consider linear alkanes and alkenes with up to twenty carbon atoms, acenes from benzene to pentacene, alanine helices with up to five residues, and water clusters with up to twenty water molecules. These systems contain

both one- and three-dimensional geometries and a variety of chemical environments. Geometries for these systems are available in supplemental data, along with the grids used for the LS-THC procedure.

All MP2 computations were performed in the cc-pVDZ basis,[70] using the standard frozen core approximation. Spherical harmonics were used for all the basis sets in these calculations. For DF-MP2 or LS-THC-DF-MP2, the cc-pVDZ-RI auxiliary basis was used in the MP2 part.[313] All of the SCF computations used density fitting in the cc-pVDZ-JKFIT basis set;[312] this affects the MP2 reference and the three approximate methods identically. Additionally, after formation of the factors, the LS-THC-MP2 and LS-THC-DF-MP2 computations were carried out in our wholly $\mathcal{O}(N^4)$ THC-MP2 module, which requires an additional Laplace denominator approximation. The Laplace denominators are approximated by the extremely efficient minimax quadratures of Braess and Hackbusch,[37,293] and were chosen to produce errors at least an order of magnitude smaller than those discussed below ($< 1\mu$H); thus all errors discussed in the paper can be taken to arise solely from the DF or LS-THC approximations.

3.4.2 Accuracy Verification

The errors of THC-MP2, DF-MP2, and LS-THC-DF-MP2 energies compared to conventional MP2 are presented in Figures 3 to 7, for linear alkanes, linear alkenes, linear acenes, alanine helices, and water clusters, respectively. Note that for the alkanes and alkenes, we use the convention that N corresponds to the number of two-carbon units in the polymer; thus the $N = 10$ alkane corresponds to $C_{20}H_{42}$. A table containing the sizes of primary basis sets, auxiliary basis sets, LS-THC grid points, and active occupied/virtual orbitals for the largest system of each class is presented in Table 2.

The accuracy plots show qualitatively the same trends across all classes of chemical systems. For THC-MP2, the error is explicitly zero for the first few systems, as the

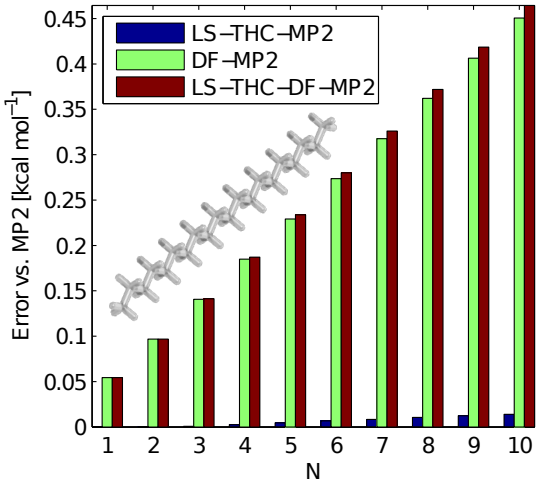

Figure 3: LS-THC-MP2 (blue bars), DF-MP2 (green bars), and LS-THC-DF-MP2 (red bars) errors compared to conventional MP2/cc-pVDZ, for linear alkanes $C_{2N}H_{4N+2}$, with up to 20 carbon atoms. Note that the blue LS-THC-MP2 bars are nonzero, but quite small on this scale.

Table 2: Number of primary basis functions, auxiliary basis functions, LS-THC grid points, and active occupied/virtual orbitals for the largest system of each class encountered in this study. The primary basis set is cc-pVDZ and the auxiliary basis set is cc-pVDZ-RI.

System	N_μ	N_A	N_P	N_{occ}	N_{vir}
$C_{20}H_{42}$	490	1708	9524	61	409
$C_{20}H_{22}$	390	1428	7084	51	319
Pentacene	378	1428	6548	51	305
Alanine Pentamer	499	1834	9316	74	399
$(H_2O)_{20}$	480	1680	9720	80	380

quadratic growth of the ov product space has not yet surpassed the linear growth of the grid size P. This numerically verifies the identity discussed in Equation 115. The critical practical finding is that the growth of the LS-THC error after this point is passed is linear (not quadratic), and possesses a remarkably small slope. This strongly implies that the LS-THC representation of the ERI tensor captures the

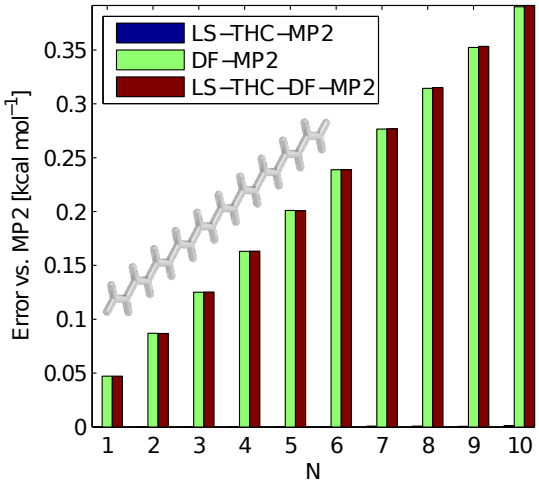

Figure 4: As in Figure 3, but for linear alkenes $C_{2N}H_{2N+2}$ with up to 20 carbon atoms. Note that the blue LS-THC-MP2 bars are nonzero, but quite small on this scale.

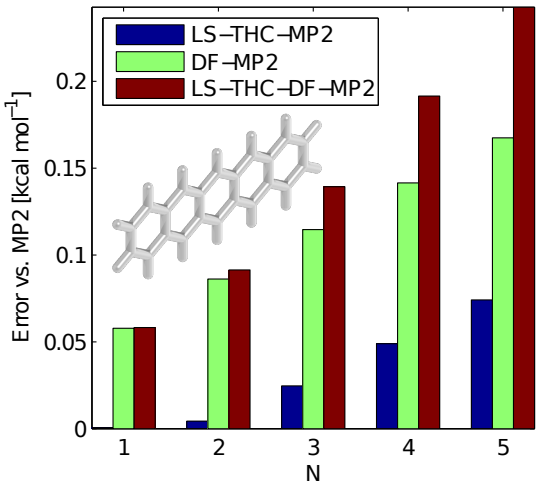

Figure 5: As in Figures 3–4, but for linear acenes from benzene (N=1) to pentacene (N=5).

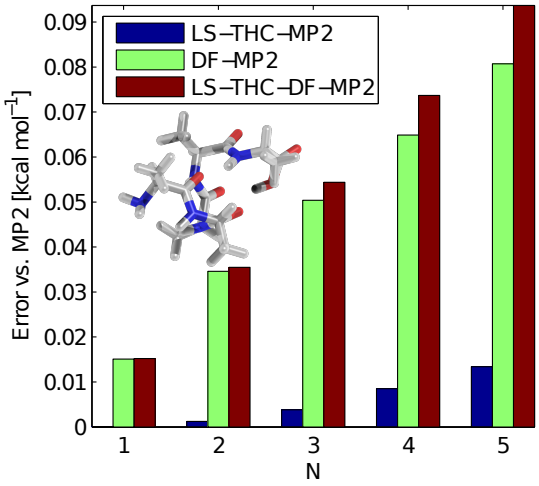

Figure 6: As in Figures 3–5, but for helical alanine polypeptides, from $(\mathrm{Ala})_1$ to $(\mathrm{Ala})_5$.

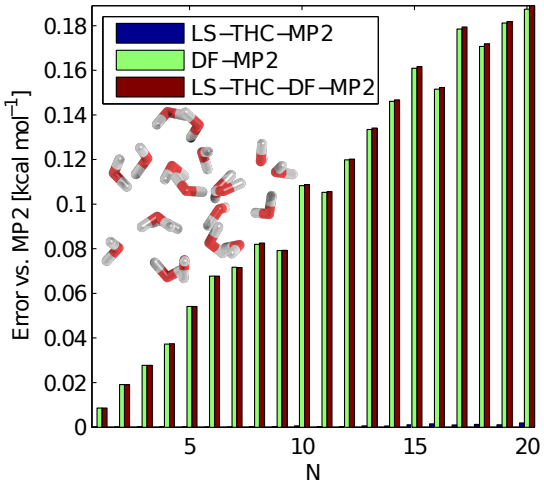

Figure 7: As in Figures 1-4, but for water clusters ranging from H_2O to $(H_2O)_{20}$.

correct physics and is more than just a forced least-squares decomposition. In fact, the pure LS-THC approximation with exact ERIs is often much more accurate than

the usual DF-MP2 approximation, for the LS-THC grids and auxiliary basis sets used here. A final piece of practical serendipity is the behavior of the LS-THC-DF-MP2 errors. Mathematically, we might expect that first approximating the ERIs by density fitting and then fitting these approximate integrals to a quadrature grid could result in a nonlinear growth of the two sources of errors. However, it appears that the LS-THC-DF-MP2 errors are essentially additive (with the LS-THC part of the error being almost negligible). This last finding allows for the existence of an accurate and explicit $\mathcal{O}(N^4)$ MP2 code using LS-THC technology, and also provides for efficient and accurate generation of the LS-THC factors in $\mathcal{O}(N^4)$ for other correlated methods.

While the accuracy plots are all qualitatively similar, the quantitative differences may give insight into the performance of the LS-THC factorization. The linear alkanes and alkenes both exhibit particularly superb accuracy, which could be attributed to the strongly one-dimensional nature of these systems, and the concomitant sparsity this induces in the ERI tensor. This trend is also borne out in the similarly excellent error properties for water clusters; though this system is truly three-dimensional, the larger average distances between functions in the *ov* product space make many of these products negligibly small, and the LS-THC capitalizes on this. However, this is not to say that LS-THC is overly dependent on sparsities in the ERI tensor; for the much less sparse cases of acenes and alanines, the LS-THC error is slightly higher, but is still entirely acceptable for any application that could have used density fitting, e.g., nearly all applications of MP2. Thus, it appears that LS-THC is robust in non-sparse chemical environments and yet can also take advantage of sparsity patterns when they are present to produce an extremely accurate approximation.

3.5 Conclusions

The LS-THC method is an analytical, accurate, and physically-motivated variant of the THC ansatz for the approximate representation of the ERI tensor. LS-THC fixes

the X factor tensors of the generic THC representation to arise from a physical-space grid collocation, and then determines the optimal Z factor tensor by least-squares minimization under the given choice of X. In a serendipitous bit of mathematics, the least-squares Z factor has an analytical expression that is relatively easy to compute. Once the Z operator is formed, LS-THC is used in the same manner as any other THC flavor, with the fourth-order ERI tensor expressed as a product of five second-order tensors, paving the way to significant scaling reductions in many methods in electronic structure theory. In this work, we have shown that LS-THC is remarkably accurate, using even a rather naïve choice of physical-space quadrature grid. In particular, LS-THC is largely robust against the sparsity patterns or bonding arrangements of the system, and thereby overcomes the severe accuracy problems associated with other grid-based methods in electronic structure. Additionally, we have shown that the LS-THC factorization can be formed in $\mathcal{O}(N^4)$ time using Coulomb-metric density-fitted integrals; the error of such a factorization is essentially the sum of the separate errors of the LS-THC and DF steps. As the LS-THC error has been shown to be of the same order or smaller as the DF error, this allows for accurate and explicit $\mathcal{O}(N^4)$ formation of the complete MP2 energy, when coupled with our previously developed THC-MP2 code.

The development of LS-THC exposes many new challenges and opportunities in electronic structure theory. The former are primarily related to the maturation of the LS-THC factorization. In particular, efficient physical-space quadrature grids (and efficient technologies to generate these grids) must be located and established. Work is already underway in our groups to this end, and preliminary results indicate that carefully selected grids may significantly decrease the number of points required for a given accuracy compared to the rather standard Becke-type grids used here. Another important consideration is efficient formation of the LS-THC factorization. In particular, the pseudoinversion of the grid metric may soon become rate-limiting

in terms of both wall time in the eigensolver and required memory, as this array is several times larger than the corresponding density fitting metric matrix. However, it is easily shown that the grid metric matrix is extremely sparse due to spatial locality of the basis functions; an efficient inversion procedure must be developed to capitalize on this fact. Yet another concern is the development of a flexible and powerful library for the generation and especially memory handling of the many Z operators that will arise for different classes of integrals in LS-THC applications to general correlated methods; the reduction from fourth-order ERIs to second-order Z operators is important, but is not such a panacea that these objects may be assumed to require negligible storage. Finally, the accuracy of LS-THC must be established for a much broader range of chemical systems, basis sets, methods, and properties. We expect that a thorough benchmark study will follow shortly, after a few intermediate developments which implement the above considerations.

In terms of opportunities, LS-THC shows great promise for the accurate scaling reduction of many difficult terms in correlated methods. As just one example, it is almost trivial to show that the most vexing $\mathcal{O}(N_{\mathrm{occ}}^2 N_{\mathrm{vir}}^4)$ terms in CCSD[239,268,267] can be evaluated in $\mathcal{O}(N_{\mathrm{occ}}^2 N_{\mathrm{vir}}^2 N_P)$ operations with LS-THC ERIs. Also, the existence of an explicit $\mathcal{O}(N^4)$ LS-THC factorization with density-fitted integrals implies that LS-THC may not be limited to methods with extremely high scalings; the crossover point may be as low as DF-MP2. Further afield, it may be possible that LS-THC can be applied to more exotic objects than ERIs. A preliminary study of the application of LS-THC to wavefunction amplitudes has already been completed in our groups.[122] Finally, the LS-THC factorization should generalize rigorously to the many-electron integrals of F12 theory,[165,300,296,305] potentially allowing for substantial savings and/or avoided approximations in these areas. In any case, LS-THC appears to combine conceptual simplicity with excellent accuracy properties, and is therefore a prime candidate for many future studies.

3.6 Alternative LS-THC Derivation

Here, we present an alternative and more thorough derivation of the explicit form of the LS-THC Z operator for the ERIs in the generic basis $(pq|rs)$, and an arbitrary set of factors X_p^P, X_q^P, X_r^Q, and X_s^Q. For clarity (and at variance with the main text of this paper), we eschew the Einstein summation convention here and write out summations explicitly.

First, we define the joint collocations, for convenience,

$$X_{pq}^P \equiv X_p^P X_q^P \text{ and } X_{rs}^Q \equiv X_r^Q X_s^Q. \tag{128}$$

The desired approximation is,

$$(pq|rs) \approx \sum_{PQ} X_{pq}^P Z^{PQ} X_{rs}^Q. \tag{129}$$

The residual tensor is,

$$\Delta_{pqrs} \equiv (pq|rs) - \sum_{PQ} X_{pq}^P Z^{PQ} X_{rs}^Q. \tag{130}$$

The elementwise square of the residual tensor is,

$$O_{pqrs} \equiv \Delta_{pqrs}\Delta_{pqrs} \tag{131}$$

$$= (pq|rs)^2 - 2(pq|rs)\sum_{P'Q'} X_{pq}^{P'} Z^{P'Q'} X_{rs}^{Q'}$$

$$+ \sum_{P'Q'} X_{pq}^{P'} Z^{P'Q'} X_{rs}^{Q'} \sum_{PQ} X_{pq}^P Z^{PQ} X_{rs}^Q.$$

The approximation above can be made to be rigorous by constraining the complete choice $\mathbf{Z} \equiv Z^{PQ}$ to minimize the square of the 2-norm (i.e., the vector Frobenius norm) of the residual tensor, which is simply the sum of the elementwise square of the residual tensor. We add a trivial factor of $1/2$ to make things easier later; this does not affect the final result,

$$O \equiv \frac{1}{2}\sum_{pqrs} O_{pqrs} = \frac{1}{2}\sum_{pqrs} \left((pq|rs) - X_{pq}^P Z^{PQ} X_{rs}^Q\right)^2 = \frac{1}{2}\left\|\Delta\right\|^2. \tag{132}$$

The weak form of the minimization condition is the stationary condition,

$$O(\mathbf{Z}^*) \leq O(\mathbf{Z}) \; \forall \; \mathbf{Z} \Rightarrow \frac{\partial O}{\partial Z^{P''Q''}} = \frac{1}{2} \sum_{pqrs} \frac{\partial O_{pqrs}}{\partial Z^{P''Q''}} = 0, \; \forall \; P'', Q''. \tag{133}$$

We must first derive the response of a single element of the square of the residual tensor to a change in the Z operator,

$$\begin{aligned} \frac{\partial O_{pqrs}}{\partial Z^{P''Q''}} &= -2 \sum_{P'Q'} (pq|rs) X_{pq}^{P'} X_{rs}^{Q'} \delta_{P'P''} \delta_{Q'Q''} \\ &+ \sum_{P'Q'} X_{pq}^{P'} X_{rs}^{Q'} \sum_{P,Q} X_{pq}^{P} Z^{PQ} X_{rs}^{Q} \delta_{P'P''} \delta_{Q'Q''} \\ &+ \sum_{PQ} X_{pq}^{P} X_{rs}^{Q} \sum_{P',Q'} X_{pq}^{P'} Z^{P'Q'} X_{rs}^{Q'} \delta_{PP''} \delta_{QQ''}. \end{aligned} \tag{134}$$

The last two terms are equivalent, and can be combined by interchange of the summation indices, leading to,

$$\begin{aligned} \frac{\partial O_{pqrs}}{\partial Z^{P''Q''}} &= \sum_{P'Q'} \delta_{P'P''} \delta_{Q'Q''} \times \\ &\left[-2(pq|rs) X_{pq}^{P'} X_{rs}^{Q'} + 2 X_{pq}^{P'} X_{rs}^{Q'} \sum_{P,Q} X_{pq}^{P} Z^{PQ} X_{rs}^{Q} \right]. \end{aligned} \tag{135}$$

Thus the total stationary condition (with the factor of 1/2) is,

$$\begin{aligned} \frac{\partial O}{\partial Z^{P''Q''}} &= \sum_{P'Q'} \delta_{P'P''} \delta_{Q'Q''} \sum_{pqrs} \\ &\left[-(pq|rs) X_{pq}^{P'} X_{rs}^{Q'} + X_{pq}^{P'} X_{rs}^{Q'} \sum_{P,Q} X_{pq}^{P} Z^{PQ} X_{rs}^{Q} \right] = 0 \; \forall \; P'', Q''. \end{aligned} \tag{136}$$

As the Kronecker deltas appear on the outside, we can discard them and simply work in terms of P' and Q'. Also factoring this expression to expose the common $X_{pq}^{P'}$ and $X_{rs}^{Q'}$ factors leads to the equation,

$$\Rightarrow - \sum_{pqrs} X_{pq}^{P'} \left[(pq|rs) - \sum_{P,Q} X_{pq}^{P} Z^{PQ} X_{rs}^{Q} \right] X_{rs}^{Q'} = 0 \; \forall \; P', Q'. \tag{137}$$

The minus sign may be trivially removed. We recognize that the quantity inside the brackets is the residual,

$$\Rightarrow \sum_{pqrs} X_{pq}^{P'} \Delta_{pqrs} X_{rs}^{Q'} = 0 \ \forall \ P', Q'. \tag{138}$$

Thus the projection of the residual onto the fitting subspace is asserted to be zero; this is the classic mathematical finding of all least-squares procedures. Rearranging Equation 137 to split the difference across the equals sign yields,

$$\Rightarrow \sum_{pqrs} X_{pq}^{P'} (pq|rs) X_{rs}^{Q'} = \sum_{pqrs} \sum_{P,Q} X_{pq}^{P'} X_{pq}^{P} Z^{PQ} X_{rs}^{Q} X_{rs}^{Q'} \ \forall \ P', Q'. \tag{139}$$

The sums over pq and rs on the right may be separated, forming the intermediate grid metric matrices $S^{P'P}$ and $S^{QQ'}$, which are symmetric positive semidefinite. Similarly, all four indices $pqrs$ on the left may be summed over, forming the intermediate $E^{P'Q'}$,

$$\Rightarrow \underbrace{\sum_{pqrs} X_{pq}^{P'} (pq|rs) X_{rs}^{Q'}}_{E^{P'Q'}} = \sum_{P,Q} \underbrace{\left[\sum_{pq} X_{pq}^{P'} X_{pq}^{P} \right]}_{S^{P'P}} Z^{PQ} \underbrace{\left[\sum_{rs} X_{rs}^{Q} X_{rs}^{Q'} \right]}_{S^{QQ'}} \ \forall \ P', Q'. \tag{140}$$

So,

$$E^{P'Q'} = \sum_{PQ} S^{P'P} Z^{PQ} S^{QQ'} \ \forall \ P', Q'. \tag{141}$$

Multiplying on the left by $\left[S^{P'P}\right]^{-1}$ and on the right by $\left[S^{QQ'}\right]^{-1}$ finally gives the analytical solution,

$$\Rightarrow Z^{PQ} = \sum_{P'Q'} \left[S^{P'P}\right]^{-1} E^{P'Q'} \left[S^{QQ'}\right]^{-1}. \tag{142}$$

Switching to matrix notation, we define $\mathbf{X_1} \equiv X_{pq,P}$, $\mathbf{X_2} \equiv X_{rs,Q}$, and $\mathbf{I} \equiv I_{pq,rs}$. In this notation, the fully expanded form of the above is,

$$\mathbf{Z} = \left[\mathbf{X_1}^\top \mathbf{X_1}\right]^{-1} \mathbf{X_1}^\top \mathbf{I} \mathbf{X_2} \left[\mathbf{X_2}^\top \mathbf{X_2}\right]^{-1} = \mathbf{S_1}^{-1} \mathbf{E} \mathbf{S_2}^{-1}. \tag{143}$$

This is the (slightly generalized) matrix-notation analog to Equation 114.

Note that only the joint collocation appears above; this means that this least-squares form holds for more complicated factors X_{pq}^{P} which are not separable in p

and q. This allows for the existence of alternative density-fitting-like factorizations. However, this finding probably has little practical utility, as the true power of LS-THC is inherently in the separability of the joint collocation.

CHAPTER IV

MOLECULAR GRIDS FOR TENSOR HYPERCONTRACTION

The following is adapted from Ref. 217.

4.1 Abstract

We investigate the application of molecular quadratures obtained from either standard Becke-type grids or discrete variable representation (DVR) techniques to the recently-developed least-squares tensor hypercontraction (LS-THC) representation of the electron repulsion integral (ERI) tensor. LS-THC uses least-squares fitting to renormalize a two-sided pseudospectral decomposition of the ERI, over a physical-space quadrature grid. While this procedure is technically applicable with any choice of grid, the best efficiency is obtained when the quadrature is tuned to accurately reproduce the overlap metric for quadratic products of the primary orbital basis. Properly selected Becke DFT grids can roughly attain this property. Additionally, we provide algorithms for adopting the DVR techniques of the dynamics community to produce two different classes of grids which approximately attain this property. The simplest algorithm is radial discrete variable representation (R-DVR), which diagonalizes the finite auxiliary-basis representation of the radial coordinate for each atom, and then combines Lebedev-Laikov spherical quadratures and Becke atomic partitioning to produce the full molecular quadrature grid. The other algorithm is full discrete variable representation (F-DVR), which uses approximate simultaneous diagonalization of the finite auxiliary-basis representation of the full position operator to produce non-direct-product quadrature grids. The qualitative features of all

three grid classes are discussed, and then the relative efficiencies of these grids are compared in the context of LS-THC-DF-MP2. Coarse Becke grids are found to give essentially the same accuracy and efficiency as R-DVR grids; however, the latter are built from explicit knowledge of the basis set and may guide future development of atom-centered grids. F-DVR is found to provide reasonable accuracy with markedly fewer points than either Becke or R-DVR schemes.

4.2 Introduction

Physical-space quadrature grids for molecular geometries are ubiquitous in chemical physics. The design of such grids is not altogether straightforward, as molecular geometries exhibit myriad 1-, 2-, and 3-dimensional conformations, with arbitrarily complex internal arrangements possible. The task is simplified somewhat in that molecular geometries can be reasonably viewed as a superposition of spherical atoms, which provides a rough natural discretization of the problem. However, desired integrands (such as the electronic density or various physical-space potentials) often vary on many length-scales, e.g., the need to capture both nuclear cusps and exponential tails in accurate integration of the electron density.[153] This inhibits the use of standard grids for common orthogonal coordinate systems. For example, if nested rectilinear grids are used in the context of the electron density, enormous refinement is required in the vicinity of the nuclei, as rectilinear quadrature rules are ill-suited for cusps and other transients at arbitrary positions. Despite these foreboding properties, many examples of efficient quadrature rules for "molecule-shaped" integrands exist in the literature. One of the most widely adopted examples is the Becke family of atom-centered grids,[20,88,205,297,204,180,87,73,44] which bears strong resemblance to the class of grids developed for pseudospectral methods by Friesner.[83] Both of these grid classes partition space into a series of atom-centered integrals, each one of which is completable to the whole of 3-dimensional space. These atom-centered integrals are

performed by a direct product of quadratures in the radial and solid-angle coordinates, which are then stitched back together to form the full molecular integral by means of a fuzzy Voronoi weighting procedure. Because these grids are built from explicit knowledge of the nuclear positions, they can be robust against length-scale changes near the nucleus, and, moreover, contain a number of points that is strictly proportional to the number of atoms in the molecule, regardless of the particulars of the molecular geometry. These grids have been used to integrate a plethora of quantities, from the electron density and Kohn-Sham exchange-correlation potential (the original objective)[20,237] to atomic volumes (e.g., the Becke exchange dipole moment method)[23] to spin and charge-transfer characteristics (e.g., constrained density functional theory)[328,329] to nonlocal correlation potentials (e.g., the VV09 and VV10 nonlocal van der Waals functionals)[62,308,309] to frequency-dependent response functions,[111,197] as just a few examples. One obvious arena in which such grids have seen less widespread use is in representation of the electronic Hamiltonian, particularly the electron repulsion integral tensor. There is good reason for this absence: evaluation of the electron repulsion integral tensor requires a double integral over the simultaneously singular and long-ranged $1/r_{12}$ Coulomb operator. A notable attempt to treat this operator with physical-space quadrature is the aforementioned pseudospectral method,[80,81,82,83,251,250,93,188,190,187,187,246,85,150,208,162] which applies a quadrature in only one coordinate of the double integral to avoid singular integrands. Recent work in our groups has identified a procedure to regularize the singularities in the physical-space $1/r_{12}$ operator to produce a finite, renormalized operator tuned to the one-particle basis used.[216] This renormalization allows for one of the first practical treatments of the electron repulsion integral by physical-space quadrature (but c.f. early work by Boys and Rajagopal along similar lines in Ref. 36). This work seeks to explore and develop tuned quadrature grids for use with this procedure. To do this, we will first identify the required integrand, and then borrow techniques from the discrete variable

representation literature of the dynamics community to adjust or wholly depart from the established Becke molecular grids.

Recently, we introduced the tensor hypercontraction (THC) representation for the electron repulsion integral (ERI) tensor.[121,216,122] This technique extends density fitting (DF)[324,67,68,299,245,144,312,24,254,151,8,6] and pseudospectral (PS)[80,81,82,83,251,250,93,188,190,187,187,246,85,15] methods to represent the four-index ERI tensor,[290]

$$(pq|rs) \equiv \int_{\mathbb{R}^6} \phi_p(\vec{r}_1)\phi_q(\vec{r}_1)\frac{1}{r_{12}}\phi_r(\vec{r}_2)\phi_s(\vec{r}_2)\ \mathrm{d}^3r_1\ \mathrm{d}^3r_2, \tag{144}$$

as,

$$(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q. \tag{145}$$

Here p, q, r, and s can refer to an arbitrary subset (including the whole set) of the primary orbital basis set $\{\phi_p(\vec{r})\}$ composed of N_p functions. The generalized Einstein summation convention is used here and throughout this paper; repeated indices present on only one side of an equation are summed over.

The least-squares variant of THC (LS-THC) preselects the X factors in some (hopefully) physically-motivated manner, and then the factor Z is tuned to produce an optimal factorization of the target integrals, in the 2-norm. Formally, LS-THC is valid with any choice of preselected factor X, though the accuracy and efficiency will depend strongly on this choice. One convenient choice selects the matrix X to be the weighted collocation matrix from a physical-space quadrature based on N_P gridpoints $\vec{r}_P$ and their corresponding weights w_P,

$$X_p^P \equiv \sqrt[4]{w_P} R_p^P = \sqrt[4]{w_P}\phi_p(\vec{r}_P). \tag{146}$$

For later convenience, we also define a joint collocation (an analogous collocation matrix to X_p^P, but in the space of charge distributions, i.e. quadratic products of basis functions):

$$X_{pq}^P \equiv X_p^P X_q^P, \tag{147}$$

and similarly for R^P_{pq}. Once the choice of factor X has been made, LS-THC constrains Z to minimize the square of the vector 2-norm in the residual tensor,

$$O(Z^{PQ}) \equiv \left\| (pq|rs) - X^P_p X^P_q Z^{PQ} X^Q_r X^Q_s \right\|^2_{\mathcal{F}}. \tag{148}$$

The stationary condition leads to an analytical formula for the LS-THC Z operator,

$$Z^{PQ} = S^{-1}_{PP'} E^{P'Q'} S^{-1}_{QQ'}. \tag{149}$$

Here the E matrix is the projection of the ERIs onto the grid, defined as:

$$E^{P'Q'} \equiv X^{P'}_{pq} (pq|rs) X^{Q'}_{rs}, \tag{150}$$

and S is the grid metric matrix,

$$S_{PP'} \equiv X^P_{pq} X^{P'}_{pq}. \tag{151}$$

As long as the X matrix is nonsingular and the number of gridpoints N_P is at least as large as the number of quadratic orbital products N_p^2, the decomposition of Equation 145 is exact because,

$$X^P_{pq} S^{-1}_{PP'} X^{P'}_{p'q'} = \delta_{pq,p'q'}. \tag{152}$$

However, the LS-THC decomposition is useful only when N_P is much less than N_p^2. Ideally, $N_P \propto N_p$, in which case decreased computational scaling can be expected when evaluating contractions with the decomposed integrals.

The results of our last three papers[216,122,119] indicate that LS-THC with a linear-scaling quadrature-based X factor is quite accurate in a wide variety of test molecules for a number of different correlation methods. To provide a physical justification for this choice of X, it is helpful to step back and consider an exact resolution of the identity:

$$(pq|rs) = \underbrace{S_{pq,p'q'} S^{-1}_{p'q',p''q''}}_{\delta_{pq,p''q''}} (p''q''|r''s'') \underbrace{S^{-1}_{r''s'',r's'} S_{r's',rs}}_{\delta_{r''s'',rs}}. \tag{153}$$

The new integrals $S_{pq,p'q'}$ are four-index overlaps,

$$S_{pq,p'q'} \equiv \int_{\mathbb{R}^3} \phi_p(\vec{r})\phi_q(\vec{r})\phi_{p'}(\vec{r})\phi_{q'}(\vec{r})\ \mathrm{d}^3r. \tag{154}$$

Next, we resolve the overlap integrals by quadrature,

$$S_{pq,p'q'} = w_P R_p^P R_q^P R_{p'}^P R_{q'}^P. \tag{155}$$

Defining $X_p^P \equiv \sqrt[4]{w_P} R_p^P$ (this is the origin of the fourth-root in Equation 146):

$$S_{pq,p'q'} = X_p^P X_q^P X_{p'}^P X_{q'}^P. \tag{156}$$

Given N_p^4 gridpoints, this quadrature can be made exact. Thus, the introduction of a quadrature does not necessarily imply any approximation. We now substitute the quadrature for the overlap integrals in Equation 153, yielding,

$$(pq|rs) = X_{pq}^P \underbrace{X_{p'q'}^P S_{p'q',p''q''}^{-1}(p''q''|r''s'')S_{r's',r''s''}^{-1}X_{r's'}^Q}_{Z^{PQ}} X_{rs}^Q, \tag{157}$$

which reveals a THC factorization,

$$(pq|rs) = X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q. \tag{158}$$

Thus, we have found an exact and explicit THC factorization of the ERI tensor which relies only on the existence of an exact quadrature for the four-index overlap integrals $S_{pq,p'q'}$. If the weighted collocation factor X from this quadrature were substituted into the LS-THC Z recipe of Equation 149, an equivalent Z operator would arise with zero residual. This provides the physical motivation for a quadrature-based X factor: in practical LS-THC, we will use a weighted collocation X from an approximate quadrature with $\mathcal{O}(N_p)$ gridpoints for the overlap integrals $S_{pq,p'q'}$. The present work explores several methods for finding efficient approximate quadrature grids for this application.

The presentation above is exact given a set of gridpoints and weights which perform the quadrature exactly. Formally, this requires a number of gridpoints $N_P \propto N_p^4$.

In practice, this may not be the case because the primary basis space $\{\phi_p(\vec{r})\}$ and quadratic product space $\{\phi_p(\vec{r})\phi_q(\vec{r})\}$ could be largely redundant. Indeed, both of these spaces lie in $L_2(\mathbb{R}^3)$, and in a complete basis they are thus asymptotically of the same complexity. In practice, this means that the N_p^2 products $\phi_p(\vec{r})\phi_q(\vec{r})$ can be accurately represented in a much smaller auxiliary basis of N_A functions $\{\chi_A(\vec{r})\}$, where ideally $N_A \propto N_p$. This is the basis for the well-known density fitting (DF) approximation,

$$\phi_p(\vec{r})\phi_q(\vec{r}) \approx d_{pq}^A \chi_A(\vec{r}). \tag{159}$$

Here d_{pq}^A are the density fitting coefficients. In practice, an accurate density fitting approximation is obtained with auxiliary bases comprised of 2-5$\times$ the number of functions in the primary basis set.

It is also apparent that the space of integrands $\{\phi_p(\vec{r})\phi_q(\vec{r})\phi_{p'}(\vec{r})\phi_{q'}(\vec{r})\}$ also lies in $L_2(\mathbb{R}^3)$, and thus shares the same asymptotic complexity class as $\phi_p(\vec{r})$. In practice, this means that we might be able to accurately integrate the four-index overlap integrals $(pqp'q')$ with a quadrature containing strictly $N_P \propto N_p$ points. A well-known piece of supporting evidence for this claim is the existence of Becke DFT-style quadrature grids;[20] these grids are arbitrarily completable for smooth functions in $\mathbb{R}^3$, with a number of points strictly proportional to the number of atoms (and thereby N_p) in the system. However, the Becke grids are typically chosen to accurately integrate the density field and exchange-correlation potential. As the functional forms of these integrands are varied and not known at initialization, the Becke DFT grids are often quite dense, with thousands of points required for each atom. Here, we devise procedures to generate grids tailored to LS-THC, leveraging the knowledge that we will be integrating functions of the form $\phi_p(\vec{r})\phi_q(\vec{r})\phi_{p'}(\vec{r})\phi_{q'}(\vec{r})$. We show that, due to the simple nature of the required integrand encountered in LS-THC, grids built from either coarse Becke grids or discrete variable representation methodology can provide remarkably compact and accurate grid representation of the ERI.

A natural simplification at this point is to work in terms of an auxiliary basis $\{\chi_A\}$ (composed of $\propto N_p$ functions) in place of the quadratic product basis $\{\phi_p\phi_q\}$, appealing to the fact that the spans are nearly equal,

$$(pqp'q') \Leftrightarrow (AA'). \tag{160}$$

Here the two-index auxiliary overlap integrals are,

$$(AA') \equiv \int_{\mathbb{R}^3} \chi_A(\vec{r})\chi_{A'}(\vec{r})\ \mathrm{d}^3r. \tag{161}$$

Thus, to obtain a linear number of quadrature points for THC, we will develop quadrature grids which accurately and efficiently integrate the overlap metric in an auxiliary basis. We will then substitute these truncated quadratures into LS-THC. The least-squares fitting determination of the Z factor will allow further fine-tuning of the accuracy in Equation 158.

Besides considering coarse variants of standard Becke grids, we also introduce two new schemes for developing multidimensional grids, both of which build on ideas from discrete variable representations (DVR). The first radial DVR (R-DVR) scheme uses a Becke DFT-like[20] atom-centered direct-product grid, with modified radial quadratures. The second full DVR (F-DVR) scheme uses approximate simultaneous diagonalization techniques to produce a true 3-dimensional non-direct-product DVR. Within theoretical chemistry, DVR techniques have been explored thoroughly by the dynamics community, especially in the context of variational computation of ro-vibrational states. A pair of useful review articles is given in Refs. 18 and 176. DVR was technically introduced in one dimension by Harris and co-workers in 1965 (though it was not known as DVR at the time), where the technique was used to provide a collocation from a basis of appropriately chosen spectral functions to a basis of physical-space nodes at which the potential matrix is diagonal, by finding the rotation of the spectral basis that is maximally localized.[103] It was later shown that this is equivalent to a Gaussian quadrature, if the basis functions are built from orthogonal

polynomials (as is almost always done).[61] DVR was formalized and connected to the variational basis representation by Light and coworkers in 1982–1985, in a series of papers which also generalized DVR to non-Gaussian-quadrature cases, and to multiple dimensions.[177,106,175] For the case of multidimensional problems, DVR provides an additional advantage: if the multidimensional basis is built as a direct product of 1-dimensional DVR functions, many products can be removed if their underlying DVR node occurs in a region where the potential is large (because the wavefunction of interest will nearly vanish at these points), often leading to spectacular sieving.[48] For the 3-dimensional molecular problem, we are concerned with the case of a basis of local, atom-centered functions such as contracted Gaussian-type orbitals, for which a direct-product grid is often unacceptably redundant. For the development of F-DVR for the molecular problem, we follow the ideas of Carrington and coworkers, who use approximate simultaneous diagonalization of the separate components of the position operator to produce an elegant non-direct-product DVR basis.[55] This simultaneous diagonalization procedure has been applied in other contexts in theoretical chemistry by a number of authors, including applications in NMR,[310] Green's functions,[234,236] scattering problems,[235] and maximally-localized Wannier functions.[99] Carrington and coworkers have also used simultaneous diagonalization in another context, by obtaining efficient one-dimensional basis sets for direct-product bases via simultaneous diagonalization of the position and Hamiltonian operators.[56,57] Additionally, Avila and Carrington have used Smolyak quadrature methods to develop efficient non-direct-product quadrature grids in a pathway distinct from simultaneous diagonaliztion.[12]

Note that in LS-THC methods, we often apply a DF auxiliary basis in a separate context from grid determination, to reduce the scaling of the formation of the E matrix of Equation 150. To distinguish between the two, we will refer to the auxiliary basis used for the E matrix as the DF auxiliary basis, and to the auxiliary basis used for R-DVR or F-DVR grid formation as the DVR auxiliary basis. Typically, the DVR

auxiliary basis is chosen to have 2-3× the number of functions as the corresponding DF auxiliary basis.

4.3 Theory

Below, we briefly review the pedagogical mechanics of DVR by building up from 1-dimensional Gaussian quadrature (GQ) to 1-dimensional discrete variable representation (1-DVR) to N-dimensional discrete variable representation (N-DVR). We then use these ideas to develop radial discrete variable representation (R-DVR) and full discrete variable representation (F-DVR) molecular grids for use in LS-THC applications.

We wish to produce efficient, minimal quadratures for quartic products of primary basis functions,

$$(pqp'q') \approx w_P R_p^P R_q^P R_{p'}^P R_{q'}^P. \tag{162}$$

As an initial simplification, we will assert that the quadratic product space is sufficiently redundant to allow for an accurate representation in an auxiliary basis, thus reducing the problem to,

$$(pqp'q') \Leftrightarrow (AA') \approx w_P R_A^P R_{A'}^P. \tag{163}$$

The task now is to determine $\mathcal{O}(N_p)$ nodes and weights ($\vec{r}_P$ and w_P, respectively) for the accurate integration of (AA'), under a given choice of DVR auxiliary basis $\{\chi_A\}$.

The notation for the DVR auxiliary basis sets is as follows,

- $\{\chi_A\}$-raw auxiliary basis set, generally not orthonormal.
- $\{\psi_A\}$-orthonormalized auxiliary basis set, typically chosen with respect to some arbitrary prescription, such as the definition of generating polynomials or a canonical orthogonalization procedure. When using canonical orthogonalization, we apply a relative eigenvalue cutoff of 10^{-10} for all results shown in this work.

- $\{\xi_P\}$-DVR function set, a unitary transformation of the orthonormalized auxiliary basis set which most closely resembles a finite set of Dirac delta functions. This basis exactly spans $\{\psi_A\}$, and is orthonormal.

As we never refer to the primary basis set in this section, we will drop the "auxiliary" label, and just refer to the basis set, orthonormal basis set, and DVR function set, respectively. Also, in the context of THC, the quadrature is usually denoted with the index P, as in $\{< \vec{r}_P, w_P >\}$. However, within a DVR framework, there is a one-to-one correspondence between the functions in the basis $\{\psi_A\}$ (with labels A to D) and the grid nodes (with labels P). These labels are interchangeable, but we choose to switch from A-D labeling to P labeling to emphasize the critical step in the DVR procedure when the finite-basis representation of the position operator (the "DVR function" basis $\{\xi_P\}$) is introduced.

4.3.1 1-Dimensional Gaussian Quadrature

The overriding goal of this paper is to produce accurate quadratures for overlap matrices (AA') with a number of grid points N_P that is proportional to the number of basis functions N_A. For a general class of basis functions $\{\chi_A\}$, no exact quadrature limited to $\mathcal{O}(N_A)$ points exists, as there are N_A^2 integrands. Thus, we will instead ask for the "best" quadrature with the restriction that $N_P = N_A$. In one-dimension and for the special class of basis functions built on orthogonal polynomials, it is well-known that an exact quadrature for (AA') can be obtained with $N_P = N_A$, using the formalism of Gaussian quadrature (GQ). The process of constructing an arbitrary GQ will provide important inspiration for the construction of similar approximate grids in non-polynomial bases.

A GQ may be constructed for any basis set of the form,

$$\psi_A(x) = P_A(x)v(x), \tag{164}$$

where $v(x)$ is a weight function, and $P_A(x)$ is a polynomial of up to degree A (A is

numbered from 0 to $N_A - 1$ in this context). The polynomials are chosen to satisfy orthonormality,

$$(AA') = \int_{\mathbb{R}} \psi_A(x)\psi_{A'}(x)\ \mathrm{d}x = \delta_{AA'}. \tag{165}$$

Note that we could have started from a non-orthonormal class of polynomial-based functions $\{\chi_A(x)\}$ (e.g., simple monomials) and then used an orthogonalization procedure to build $\{\psi_A(x)\}$.

There are many conceptual algorithms for the production of Gaussian quadratures. The most relevant to this discussion is the Golub-Welsch algorithm,[91] which simply states that the nodes of the Gaussian quadrature are the eigenvalues of the position operator within the orthonormal basis $\{\psi_A(x)\}$,

$$X_{AB}Q_{BP} = Q_{AP}x_P, \tag{166}$$

where,

$$X_{AB} \equiv (A|\hat{x}|B) = \int_{\mathbb{R}} \psi_A(x)x\psi_B(x)\ \mathrm{d}x. \tag{167}$$

The eigenvectors Q_{AP} define a rotation of the basis $\{\psi_A(x)\}$ to a "DVR function" basis $\{\xi_P(x)\}$ which most closely resembles a set of Dirac delta functions (the eigenfunctions of the position operator in the complete basis), under the finite support of the chosen basis,

$$\xi_P(x) = Q_{AP}\psi_A(x). \tag{168}$$

For the polynomial basis of Equation 164, the DVR function basis possesses a remarkable "interpolating zeros" property,

$$\xi_P(x_Q) = \delta_{PQ}\xi_P(x_Q). \tag{169}$$

That is, at each quadrature node, all of the DVR functions are exactly zero except for one. This property allows one to uniquely pick out the part of any integrand in $\xi_P(x)\xi_P(x)$ by sampling at x_P, and is critical to exactness of the quadrature. Exploiting the orthonormality of the DVR functions, for any integrand lying entirely

within the span of the finite product space (i.e., $z(x) = \sum_{PQ} Z_{PQ}\xi_P(x)\xi_Q(x)$), we have,

$$\int_{\mathbb{R}} z(x)\ \mathrm{d}x = \sum_{PQ} Z_{PQ} \underbrace{\int_{\mathbb{R}} \xi_P(x)\xi_Q(x)\ \mathrm{d}x}_{\delta_{PQ}} = \sum_P Z_{PP}. \tag{170}$$

Thus, we need only determine the part of the integrand Z_{PP} that lies in $\xi_P(x)\xi_P(x)$. This step is trivial, due to the interpolating zeros property in the DVR basis, as,

$$z(x_P) = Z_{QR}\xi_Q(x_P)\xi_R(x_P)\delta_{PQ}\delta_{PR} = Z_{PP}[\xi_P(x_P)]^2. \tag{171}$$

leading to,

$$Z_{PP} = \frac{z(x_P)}{[\xi_P(x_P)]^2}. \tag{172}$$

Therefore, the positive-definite weights of the GQ are determined as,

$$w_P = \frac{1}{[\xi_P(x_P)]^2}. \tag{173}$$

The mathematics above are somewhat abstract and can be clarified with an example. Consider the nondimensional Hermite functions (solutions to the 1-dimensional quantum harmonic oscillator),[94]

$$\psi_A = H_A^{(0)}(x)\exp(-x^2/2). \tag{174}$$

Here $H_A^{(0)}(x)$ is the A-th physicists' Hermite polynomial,[2] multiplied by an extra normalizing factor of $(\sqrt{\pi}2^A A!)^{-1/2}$. The orthonormal Hermite functions from $A = 0$ to $A = 6$ are shown in the top panel of Figure 8. The position operator in the orthonormal Hermite basis is simply the tridiagonal (with zero diagonal),

$$X_{AB} = \int_{\mathbb{R}} \psi_A(x)x\psi_B(x)\ \mathrm{d}x = \sqrt{\frac{A+1}{2}}(\delta_{A+1,B} + \delta_{A,B+1}). \tag{175}$$

Diagonalizing this matrix to produce Q_{AP} and x_P yields the quadrature nodes and DVR functions, shown in the bottom panel of Figure 8. This is, in fact, the well-known Gauss-Hermite quadrature rule. Close inspection of the nodes indicates that

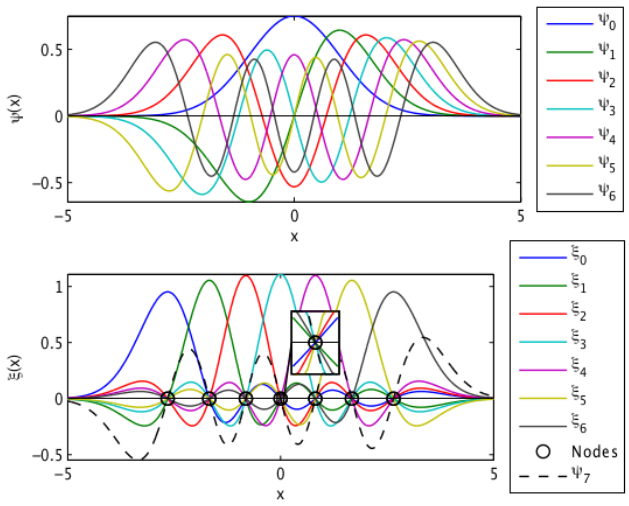

Figure 8: GQ scheme for the Hermite functions ψ_0 to ψ_6. Top: orthonormal spectral basis set $\{\psi_A(x)\}$. Bottom: DVR function set $\{\xi_P(x)\}$, Because this is a Gaussian quadrature, the DVR basis is a true interpolating set, with the exact property $\xi_P(x_Q) = \delta_{PQ}\xi_P(x_Q)$. These interpolation nodes correspond to the zeros of the next basis function outside of the set, $\psi_7(x)$.

the interpolating zeros property is indeed obtained. In fact, the interpolating zeros happen to occur exactly at the zeros of the next Hermite function outside of the finite set $\{\psi_A(x)\}$, as seen in the bottom panel. Demanding this property is a popular alternative to the Golub-Welsch algorithm, but is not as relevant to our later consideration of quadratures in more exotic bases than polynomials. A very important consideration for GQ is that this set of interpolating functions was constructed with a unitary transformation, i.e., the orthonormality is maintained in the set of interpolating (DVR) functions.

Another important way to view this process is depicted in Figure 9, which depicts a DVR function for the case where the Hermite functions range from $A = 0$ to $A = 10$. The DVR functions provide the best-possible approximations under the finite support of the chosen basis to the Dirac delta functions that are the true eigenfunctions of the complete-basis position operator. We then switch from spectral space $\{\xi_P(x)\}$ to position space x_P by conceptually replacing $\xi_P(x)$ with the Dirac delta function

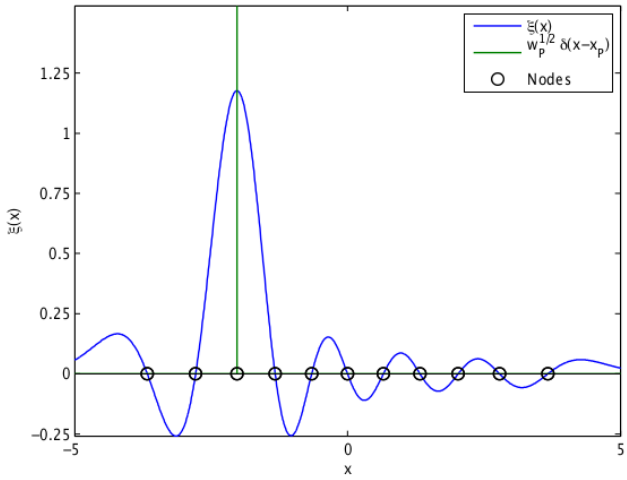

Figure 9: Conceptual representation of the Gaussian quadrature or discrete variable representation scheme, in this case $\xi_2(x)$ and its Dirac delta partner, corresponding to the Hermite functions from ψ_0 to ψ_{10}. The eigenfunctions and eigenvalues of the position operator are determined in the orthonormal basis. The eigenfunctions are finite spectral approximations to Dirac delta functions, and the eigenvalues provide the discretization of the otherwise continuous position operator. To move from spectral space to physical space, the DVR functions are replaced in a one-to-one fashion by weighted Dirac delta functions placed at the eigenvalues of the finite position operator.

$\sqrt{w_P}\delta(x - x_P)$. For the case of a polynomial basis set, with the interpolating zeros property, this last substitution is amazingly not an approximation. In more exotic basis sets, the interpolating zeros property will be lost and the substitution will be a physically-motivated approximation.

4.3.2 1-Dimensional Discrete Variable Representation

The first of two problems (see following section for the second) at this point is that the molecular auxiliary basis set $\{\chi_A(\vec{r})\}$ does not have a simple polynomial-based form of Equation 164, even for a spherically symmetric single atom. Even in the 1-dimensional case, a completely general (i.e., non-polynomial) basis set $\{\chi_A(x)\}$ implies that no exact N_A-point quadrature exists for the N_A^2 elements of the overlap matrix, as the equations are grossly overdetermined.

Starting from GQ but moving to a general basis set, we could maintain the interpolating zeros property by breaking the orthonormality in the DVR functions (implying Q_{AP} is not a unitary transformation), or we could retain the orthonormality in the DVR functions but break the interpolating zeros property. Following the DVR literature, we do the latter, choosing a unitary transformation Q_{AP} that approximately retains the interpolating zeros property. It turns out that simply applying the Golub-Welsch algorithm in the general orthonormal basis $\{\psi_A(x)\}$ accomplishes this goal admirably, for well-behaved classes of functions. The rationale is that in the complete basis set, the eigenfunctions of the position operator will converge to true Dirac delta functions, regardless of the underlying construction of the basis. In the finite basis, this property is approximately retained, and the eigenfunctions of the finite-basis position operator approximately resemble Dirac delta functions. Thus, for a completely generic raw basis $\{\chi_A(x)\}$, we first extract an orthonormal representation of the basis $\{\psi_A(x)\}$, e.g., by canonical orthogonalization, and then determine the approximate quadrature by Equations 166, 167, and 173.

A concrete example of this process is DVR in a basis of unit-length particle-in-a-box wavefunctions,[94] i.e.,

$$\psi_A(x) = \sqrt{2}\sin(A\pi x),\ x\ \in\ [0,1]. \tag{176}$$

Here A ranges from 1 to N_A. The particle-in-a-box basis $\{\psi_A(x)\}$ for $N_A = 6$ is shown in the upper panel of Figure 10. The application of the Golub-Welsch algorithm is shown in the bottom panel. It is apparent that the interpolating zeros property is approximately, but not exactly retained. Therefore, we expect that the quadrature will approximately reproduce the overlap metric, but will incur error (so-called "alias") proportional to the square of the magnitude of the near-zero DVR functions at each grid node. The error of this DVR quadrature in reproducing the overlap metric (AA') is shown in Figure 11. It is apparent that the quadrature performs remarkably well, considering that only 12 parameters were available to integrate 36 elements of the

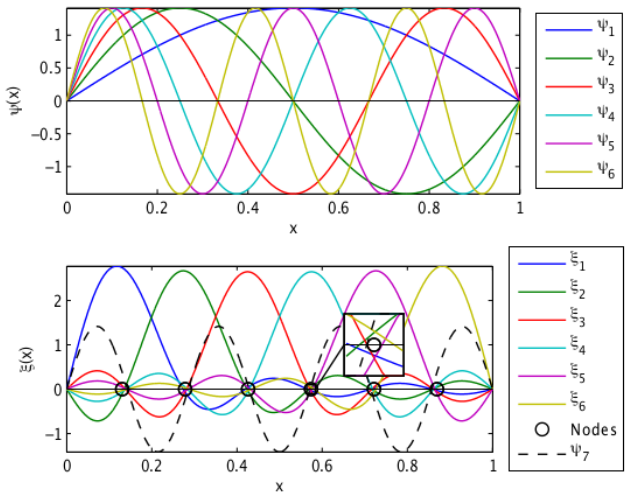

Figure 10: DVR scheme for the particle-in-a-box wavefunctions ψ_1 to ψ_6. Top: orthonormal spectral basis set $\{\psi_A(x)\}$. Bottom: DVR function set $\{\xi_P(x)\}$. Because this is a DVR and not a GQ, the DVR basis is a not a true interpolating set, and only exhibits the approximate interpolation property $\xi_P(x_Q) \approx \delta_{PQ}\xi_P(x_Q)$. These interpolation nodes correspond approximately to the zeros of the next basis function outside of the set, $\psi_7(x)$.

6-by-6 overlap matrix. The errors in each element are typically on the order of 0.05, though the maximum error is 0.182. Most of the error is committed on the diagonal, and the parity selection rule is maintained exactly. The largest error occurs at the terminal (66) overlap element, which is the most oscillatory integrand, and suffers from the loss of completeness that could be provided by ψ_7 and higher functions. Building a DVR quadrature with ψ_7 added, and then recalculating the quadrature for the overlap matrix up to (66) reduces the error of (66) to 0.025, and the maximum error (still on the diagonal) to 0.051. Therefore, we note that adding higher functions to the DVR basis provides a convenient way to increase the fidelity of the quadrature for the overlap matrix (i.e. to "dealias").

4.3.3 N-Dimensional Discrete Variable Representation

The second problem facing our extension of GQ to the molecular problem is that the molecular auxiliary basis is fully 3-dimensional and not trivially separable for the

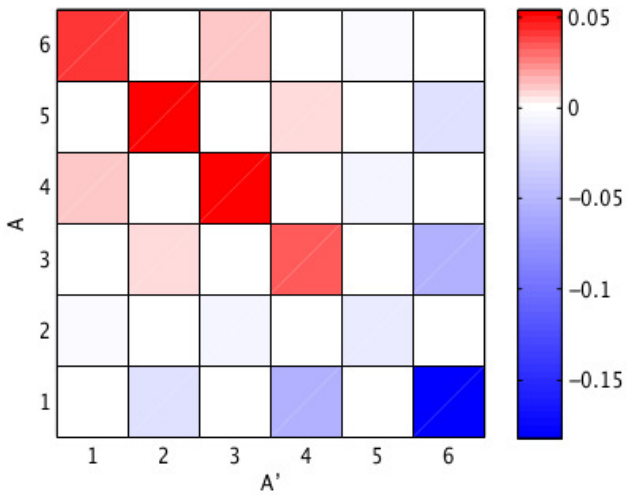

Figure 11: Error of DVR quadrature scheme for the particle-in-a-box wavefunctions ψ_1 to ψ_6. Depicted is the quadrature error $E_{AA'}$ in the overlap matrix (AA'), computed as $E_{AA'} \equiv \delta_{AA'} - \psi_A(x_P) w_P \psi_{A'}(x_P)$. Note that the exact overlap matrix is the identity $(AA') = \delta_{AA'}$.

atom-centered Gaussian functions in common use. Two primary approaches will be applied to attack this problem.

The first approach (R-DVR) applies direct-product expansions to produce molecular quadrature grids. An excellent example of this technique is the nearly-ubiquitous Becke DFT-style quadrature grid.[20] This class of grid is based on partitioning of the full molecular quadrature into atom-centered spherical quadratures, followed by direct-product separation of the atom-centered quadratures into radial and angular parts. Separate DVRs for the expected functions in the radial $\{< \rho_{P_\rho}, w_{P_\rho} >\}$ and angular $\{< \Omega_{P_\Omega}, w_{P_\Omega} >\}$ coordinates may be simply combined to produce each atomic quadrature,

$$\{< \vec{r}_P, w_P >\} \equiv \{< \rho_{P_\rho}, w_{P_\rho} > \otimes < \Omega_{P_\Omega}, w_{P_\Omega} >\}. \tag{177}$$

For the case of atom-centered Gaussian or Slater functions, the existing Lebedev-Laikov quadrature rules for the solid angle are already essentially optimal, and do not need to be modified.[171] However, standard Becke-type DFT grids use largely prescriptive quadrature rules for the radial coordinate,[20,297,204] which may or may not

reflect the distribution of radial functions in the auxiliary basis $\{\chi_A(\vec{r})\}$. Therefore, we will apply 1-dimensional DVR techniques in the radial coordinate to produce the radial discrete variable representation (R-DVR) class of quadrature grid. While this choice of grid has better physical motivation of the radial quadratures than Becke grids, it still retains the overarching direct-product flavor of the Becke grid, and will therefore necessarily be somewhat redundant in representing the span of the non-direct-product auxiliary basis, even when pruning is applied to the spherical quadratures.

The second approach (Full DVR or F-DVR) applies the Golub-Welsch algorithm in 3-dimensions by approximately simultaneously diagonalizing the Cartesian representation of the position operator. This results in true "molecule-shaped" quadrature grids with none of the direct-product structure of the Becke or R-DVR grids. These grids directly reflect the underlying structure of the full auxiliary basis $\{\chi_A(\vec{r})\}$, and may thus be more efficient than R-DVR grids.

4.3.4 Radial Discrete Variable Representation

The idea behind R-DVR is quite simple: in Becke atom-centered DFT grids,[20,205,297,204,288,171] instead of using radial quadratures based on mappings of standard Gaussian quadratures to the domain $\rho \in (0, \infty)$, we will instead use 1-dimensional DVR quadratures built from the finite support of the radial coordinate in the DVR auxiliary basis.

For our problem, the finite radial span in the raw basis consists of all integrals of the form,

$$S_{AB} = \int_0^\infty \rho^2 \chi_A(\rho)\chi_B(\rho)\ \mathrm{d}\rho. \tag{178}$$

where, for a contracted Gaussian-type basis set, the radial basis functions are,

$$\chi_A(\rho) = N_A \sum_{i_A} K_{i_A} \exp(-\alpha_{i_A}\rho^2). \tag{179}$$

Here, α_{i_A} is the orbital exponent and K_{i_A} is the contraction coefficient. The normalization is chosen so that $S_{AA} = 1$. The polynomial Cartesian products have been

absorbed into the spherical harmonics, which will be integrated in a direct-product sense by Lebedev-Laikov quadratures. Therefore, the raw DVR basis for each atom contains one function for each basis shell, with all shells being treated equally, regardless of angular momentum. One consequence of this is that severe redundancy often exists within the atomic DVR basis, corresponding to shells of different angular momenta that have nearly identical radial behavior. The elimination of much of this redundancy by a technique such as canonical orthogonalization can improve the efficiency of the R-DVR quadrature markedly. For instance an S and P shell with the same radial distribution will be combined by canonical orthogonalization, resulting in only one radial node. The Lebedev-Laikov grid placed at this radial node integrates, by design, all spherical harmonics up to its maximum degree, and can therefore integrate all products of these two functions without a need for a second radial node.

To produce the R-DVR radial quadrature at each atom, the radial overlap metric is first constructed, and canonically orthogonalized to produce the orthonormal basis, via the linear transformation W,

$$\psi_B(\rho) = W_{AB}\chi_A(\rho). \tag{180}$$

The position operator $\hat{\rho}$ is then constructed in the finite basis $\{\chi_A(\rho)\}$ as,

$$R'_{AB} = \int_0^\infty \rho^3 \chi_A(\rho)\chi_B(\rho)\ \mathrm{d}\rho. \tag{181}$$

and then transformed to the orthonormal basis $\{\psi_A(\rho)\}$,

$$R_{AB} = W_{CA}R'_{CD}W_{DB}. \tag{182}$$

This matrix is then diagonalized obtain the DVR functions and the radial nodes $\{\rho_P\}$,

$$R_{AB}Q_{BP} = Q_{AP}\rho_P. \tag{183}$$

The DVR functions are given as,

$$\xi_P(\rho) = Q_{BP}W_{AB}\chi_A(\rho), \tag{184}$$

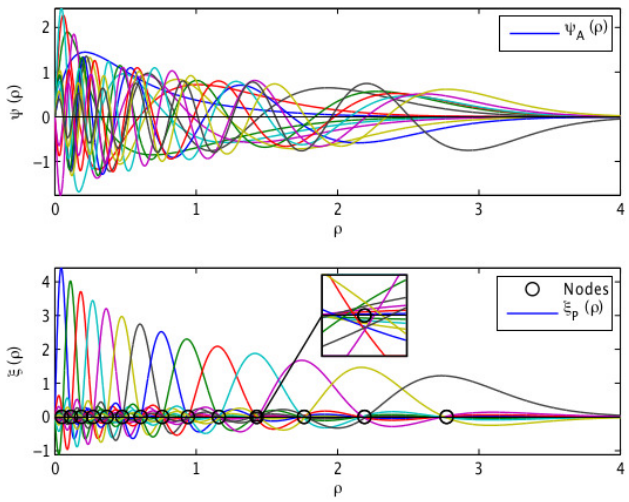

Figure 12: R-DVR scheme for Ne atom with a cc-pVDZ-RI DVR auxiliary basis set. Top panel: canonically-orthogonalized radial basis $\{\psi_A(\rho)\}$. Bottom panel: DVR function set $\{\xi_P(\rho)\}$. Note that, in both frames, the square-root of the inner-product weight is rolled into the functions as $\psi_A(\rho) = \rho\psi'_A(\rho)$, for clarity. Because this is a DVR and not a GQ, the DVR basis is a not a true interpolating set, and only exhibits the approximate interpolation property property $\xi_P(\rho_Q) \approx \delta_{PQ}\xi_P(\rho_Q)$.

and the grid weights are determined as,

$$w_P = \frac{\rho_P^2}{[\xi_P(\rho_P)]^2}. \tag{185}$$

where the inner-product weight ρ^2 is absorbed into the traditional DVR weight.

Figure 12 shows the R-DVR scheme for a neon atom in a cc-pVDZ-RI[313] DVR auxiliary basis set. The orthonormal basis $\{\psi_A(\rho)\}$ is somewhat jumbled, as it is obtained from a canonical orthogonalization procedure. However, the DVR basis $\{\xi_P(\rho)\}$ is quite regular, very nearly reproducing the interpolating zeros property.

Algorithm 4 depicts the procedure for producing the R-DVR grid for an arbitrary molecule and DVR auxiliary basis. Lines 4-12 form the heart of the R-DVR prescription by producing a physically-motivated radial DVR quadrature instead of using a standard quadrature rule as in Becke grids. All other steps are identical to those used to construct Becke grids.

Pruned selection of the Lebedev-Laikov grid at each radial node (for both the

Algorithm 4 R-DVR grid generation.

procedure R-DVR(Molecule $\{< x_i, y_i, z_i, Z_i >\}$, Molecular DVR basis $\{\chi'_A(\vec{r})\}$)
  Compute Standard Angular Orientation
  for all Atoms i in Molecule **do**
    Select $\{\chi_A(\vec{r})\} \equiv \{\chi'_A(\vec{r})\}\ \forall\ A \in i$
    $S_{AB} = \int_0^\infty \rho^2 \chi_A(\rho)\chi_B(\rho)\ \mathrm{d}\rho$
    $R'_{AB} = \int_0^\infty \rho^3 \chi_A(\rho)\chi_B(\rho)\ \mathrm{d}\rho$
    Find $W_{AB}:\ W_{CA}S_{CD}W_{DB} = \delta_{AB}$ ▷ Canonical Orthogonalization, $\epsilon_{\mathrm{rel}} \equiv 10^{-10}$
    $R_{AB} = W_{CA}R'_{CD}W_{DB}$
    $R_{AB}Q_{BP} = Q_{AP}\rho_P$
    $\xi_P(\rho_P) = Q_{BP}W_{AB}\chi_A(\rho_P)$
    $w_P = \rho_P^2/[\xi_P(\rho_P)]^2$
    Radial Quadrature for Atom is $\{< \rho_P, w_P >\}$
    for all Radial Nodes P in Radial Quadrature **do**
      Select a Spherical Quadrature $\{< \Omega_D, w_D >\}$ for ρ_P ▷ Prune based on $\rho_{\mathrm{Bragg-Slater}}$
      Add $\{< (\rho_P, \Omega_D), w_P w_D >\}$ to Atomic Quadrature
    end for
    Apply Standard Angular Orientation
    Translate Atomic Quadrature to Nuclear Center $< x_i, y_i, z_i >$
  end for
  Apply Becke Atomic Partitioning
  return Molecular Quadrature $\{< \vec{r}_P, w_P >\}$
end procedure

Becke and R-DVR grids) is accomplished by means of a heuristic but reasonable envelope function,

$$L(z \equiv \rho_P/\rho_{\mathrm{Bragg-Slater}}) = \lceil L_{\max}\left(1.0\mathrm{erf}(1.2(z-0.0)) - 0.5\mathrm{erf}(1.2(z-1.5)) - 0.5\right)\rceil .$$

Here $\rho_{\mathrm{Bragg-Slater}}$ is the Bragg-Slater radius, and $L_{\max}$ is the maximum allowed Lebedev-Laikov grid. This envelope function is peaked near the Bragg-Slater radius and remains significant in the Van der Waals region, and approaches zero both very near and very far from the nucleus. At each node, the next Lebedev-Laikov grid of equal or larger maximum angular momentum to $L(\rho_P)$ is selected. Limited hand-tuning of the scale and bias parameters inside the erf functions was performed. More complete optimization of this pruning procedure, possibly including explicitly defined Lebedev-Laikov orders for each radial node on each atom, is certainly a worthy topic of future study.[88,44]

4.3.5 Full Discrete Variable Representation

The R-DVR scheme above corrects the placement of radial grid nodes to incorporate explicit knowledge of the DVR basis set $\{\chi_A(\rho)\}$, but is still quite redundant due to the direct-product formulation in radial and angular coordinates. Here we develop F-DVR to produce a quadrature grid with intrinsic knowledge of the 3-dimensional polyatomic structure of the basis set.

If the basis were complete, any orthonormal basis representation of the three Cartesian position operators ($\hat{x}$, $\hat{y}$, and $\hat{z}$) would commute, and would therefore admit simultaneous eigenstates. These eigenstates would be Dirac delta functions in 3-dimensional space, and would form the natural position-space quadrature for this problem. However, in a finite orthonormal basis $\{\psi_A(\vec{r})\}$, the commutators are all nonzero,

$$[X, Y] \neq 0,\ [X, Z] \neq 0,\ [Y, Z] \neq 0. \tag{186}$$

Therefore, no single orthonormal transformation Q exists which diagonalizes all three finite operators, and therefore no fully rigorous DVR function set exists.

Having acknowledged that the finite position operator components cannot be simultaneously diagonalized exactly, the Simultaneous Diagonalization (SD) procedure proposes to go ahead and diagonalize anyways, at least in an approximate sense.[55] The components of the position operator are formed in the orthonormal basis $\{\psi_A(\vec{r})\}$,

$$R^i_{AB} = (A|\hat{r}_i|B),\ i \in [x, y, \ldots]. \tag{187}$$

These are the matrices X, Y, and Z in Equation 186 above. The optimal unitary transformation Q is then located such that the sum of the squares of the off-diagonal elements of the similarity transformed position operators is minimized,

$$O(Q) = \sum_{i\in[x,y,z]} \sum_{P\neq Q} \left(Q_{AP} R^i_{AB} Q_{BQ}\right)^2 \tag{188}$$

The optimal Q is presently determined by a modification of the iterative Jacobi eigenvalue algorithm[134] presented generically by Bunse-Gerstner and coworkers,[40] and later furthered by the development of closed-form expressions for the Jacobi angles by Cardoso and Souloumiac.[41] This algorithm considers the overall unitary transformation to be composed as a sequence of Jacobi rotations which minimize the sum of the squares of off-diagonal elements in all 2×2 minors of the full position matrices. Rotations in one part of the matrix affect values in other parts of the matrix, leading to the iterative nature of the algorithm. The approximately diagonalized matrices,

$$r^i_{PQ} = Q_{CP} R^i_{CD} Q_{DQ}, \tag{189}$$

are formed in place by overwriting the original R^i_{AB} matrices.

Once the optimal Q is determined, the grid nodes are taken to be the diagonal elements of the approximately diagonalized finite position operator,

$$\{\vec{r}_P\} \equiv \{< x_{PP}, y_{PP}, z_{PP} >\}. \tag{190}$$

The values of DVR functions are then determined at the grid nodes,

$$\xi_P(\vec{r}_P) = Q_{BP}\psi_B(\vec{r}_P). \tag{191}$$

These values approximately satisfy the interpolating zeros property. The grid weights are then determined in the usual manner, according to Equation 173.

Algorithm 5 depicts the procedure for forming the F-DVR grid for an arbitrary molecule and DVR auxiliary basis. Line 10 is the simultaneous diagonalization subroutine, which iteratively minimizes the sum of the squares of the off-diagonal elements of the finite-basis position operator with respect to the orthonormal elements of Q.

Algorithm 5 F-DVR grid generation.

1: **procedure** F-DVR(Molecular DVR basis $\{\chi_A(\vec{r})\}$)
2: $S_{AB} = \int_{\mathbb{R}^3} \chi_A(\vec{r})\chi_B(\vec{r})\mathrm{d}^3r$
3: $X'_{AB} = \int_{\mathbb{R}^3} \chi_A(\vec{r})x\chi_B(\vec{r})\mathrm{d}^3r$
4: $Y'_{AB} = \int_{\mathbb{R}^3} \chi_A(\vec{r})y\chi_B(\vec{r})\mathrm{d}^3r$
5: $Z'_{AB} = \int_{\mathbb{R}^3} \chi_A(\vec{r})z\chi_B(\vec{r})\mathrm{d}^3r$
6: Find W_{AB} : $W_{CA}S_{CD}W_{DB} = \delta_{AB}$ ▷ Canonical Orthogonalization, $\epsilon_{\text{rel}} \equiv 10^{-10}$
7: $X_{AB} = W_{CA}X'_{CD}W_{DB}$
8: $Y_{AB} = W_{CA}Y'_{CD}W_{DB}$
9: $Z_{AB} = W_{CA}Z'_{CD}W_{DB}$
10: Minimize $O \equiv \sum_{P\neq Q}[Q_{AP}X_{AB}Q_{BQ}]^2+[Q_{AP}Y_{AB}Q_{BQ}]^2+[Q_{AP}Z_{AB}Q_{BQ}]^2$ w.r.t. Q_{AP}
11: $x_P = Q_{AP}X_{AB}Q_{BP}$
12: $y_P = Q_{AP}Y_{AB}Q_{BP}$
13: $z_P = Q_{AP}Z_{AB}Q_{BP}$
14: $\xi_P(\vec{r}_P) = Q_{BP}W_{AB}\chi_A(\vec{r}_P)$
15: $w_P = 1/[\xi_P(\vec{r}_P)]^2$
16: Molecular Quadrature is $\{<\vec{r}_P, w_P>\} \equiv \{<(x_P, y_P, z_P), w_P>\}$
17: **return** Molecular Quadrature $\{<\vec{r}_P, w_P>\}$
18: **end procedure**

4.4 Results

4.4.1 Computational Details

R-DVR and F-DVR grid routines were coded in a development version of the open-source PSI4 quantum chemistry package,[298] alongside our existing Becke DFT quadrature grids. Below we compare several representative grids from each of the Becke, R-DVR and F-DVR methods. Thus, we start with a summary of technical details such as the choice of pruning, radial grid scaling, and grid orientation.

In choosing grids, especially of the Becke or R-DVR type, there are many tunable degrees of freedom, such as maximum Lebedev grid order (Becke and R-DVR), pruning scheme (Becke and R-DVR), radial quadrature (Becke), DVR basis (R-DVR and F-DVR), and canonical orthogonalization cutoff (R-DVR and F-DVR). For the purposes of this qualitative demonstration, we have made no attempt here to exhaustively search these, but have instead adopted a convention which qualitatively attempts to produce Becke, R-DVR, and F-DVR grids with roughly the same number of points, with the grid density designed to be maximal in the valence region. We start by adopting aug-cc-pVTZ-RI[313] as the DVR auxiliary basis $\{\chi_A\}$, which completely defines the F-DVR grid. This basis also defines the radial part of the R-DVR grid, which has 14 radial nodes for hydrogens and 28 radial nodes for first-row elements, before canonical orthogonalization is considered. A relative canonical orthogonalization cutoff of 1.0×10^{-10} is used in both the R-DVR and F-DVR procedures, and often cuts off a few nodes in the R-DVR procedure, corresponding to two radial functions with a similar exponent that would ordinarily be orthogonal due to belonging to shells of two different angular momenta. Generally this reduces the number of R-DVR radial nodes to 11 for hydrogens and 19 for first-row atoms. We have therefore chosen Becke grids with the same count of 11 and 19 radial nodes for hydrogens and first-row atoms, respectively. The Becke radial quadrature applied is the standard Treutler/Ahlrichs mapping of the Gauss-Chebyshev quadrature, scaled

by the Bragg-Slater radius of the atom.[297] For both R-DVR and Becke quadratures, a 7-th order Lebedev-Laikov grid is used, which uses 26 points to cover the solid angle. It is well known that strict direct-product atomic grids lead to excessive crowding of gridpoints for small radial distances. Thus, we follow the usual practice where spherical grids are pruned so that lower-order grids are used near the nuclei and in the tails, and the highest-order grids are only used in the valence regions. This pruning uses a heuristic approach based on the proximity of the radial point to the Bragg-Slater radius, producing envelopes analogous to those in Ref. 88. With this pruning applied, the Becke and R-DVR grids have 2-2.5$\times$ the number of nodes as the corresponding F-DVR grid. This ratio cannot be made significantly smaller without severely compromising completeness in either the radial or spherical coordinate of the Becke or R-DVR grids. Both Becke and R-DVR grids use the Treutler/Ahlrichs modification of Becke's fuzzy Voronoi partitioning of the atomic quadratures. Finally, the Becke and R-DVR grids are aligned according to the rules of Gill and co-workers, as is standard practice for Becke grids.[88,140]

For further analysis, all grids and molecular geometries are available as text and XYZ files in the supplementary materials, along with additional high-quality views of the grids produced for a variety of molecules. Grid and basis size parameters for the largest system of each class studied here are presented in Table 3, with more-detailed size information available in the supplementary materials.

The computational demands of the Becke and R-DVR methods are quite similar. The determination of the radial quadrature at each atom in R-DVR involves only two eigendecompositions of matrices of dimension $\mathcal{O}(10)$ on each side. Like Becke grids, R-DVR grids are dominated by the atomic partitioning of the spherical quadratures, which scales as $\mathcal{O}(N_P)$, if sparsity is used properly. By contrast, F-DVR is currently considerably more involved. At present, F-DVR involves the direct eigendecomposition of the overlap matrix to form the conditioned canonically-orthogonalized basis

$\{\psi_A\}$, followed by the iterative, Jacobi eigenvalue algorithm. Both of these steps are $\mathcal{O}(N_P^3)$. Besides the floating point operation (FLOP) count, these steps may also be rate-limiting for large systems due to the memory requirement, which peaks at $4N_P^2$ in the simultaneous diagonalization step. We have not endeavored to optimize the F-DVR method, but expect that it can be made significantly more efficient by exploiting sparsity.

The simultaneous diagonalization algorithm developed in Refs. 40 and 41 is applicable directly with minimal modifications for the F-DVR grid construction. The "shaking" trick applied in Ref. 40 to break spurious fixed-points is only needed in the first iteration in systems with extremely high symmetry (for instance, an atom), and the more-elaborate four-scale shaking trick of Ref. 55 is not needed at all. The algorithm typically converges to the point that all further motions of the nodes are negligible in approximately 15 Jacobi sweeps.

4.4.2 Qualitative Grid Features

Here, the qualitative features of the various grid generation methods is discussed using a few representative molecular systems. Figure 13 compares the Becke, R-DVR, and F-DVR grids for a single H_2O molecule. Figure 14 provides the same comparison for an anthracene molecule. Further analysis of F-DVR grids is provided in Figures 15 and 16, which depict the F-DVR grids for $(H_2O)_3$ and a three-residue alanine helix. The coloring used at each node is roughly representative of the order of magnitude of the raw quadrature weight, where red indicates a small weight and blue represents a large weight. Note that larger weights typically occur further from the nuclei, which is indicative of the ρ^2 inner-product metric that would be used in spherical integration. Additional views of these and several other systems are available in the supplementary data, along with the grid coordinates and molecular geometries.

The quasi-direct product nature of the Becke and R-DVR grids is clearly visible

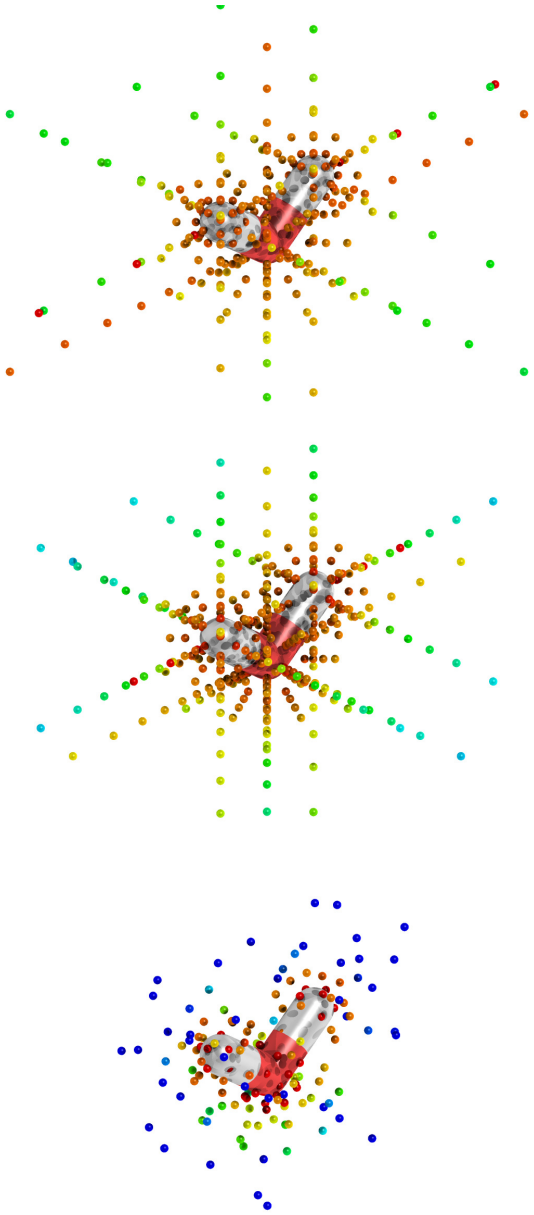

Figure 13: Comparison of physical-space grids obtained with various generation methods for the H_2O molecule. Top panel: Becke grid with 19/11-node Treutler/Ahlrichs radial quadratures (on heavy/hydrogen atoms, respectively) and a pruned 7-th order Lebedev-Laikov spherical quadrature, for a total of 442 nodes. Middle panel: R-DVR grid within an aug-cc-pVTZ-RI DVR auxiliary basis, with a pruned 7-th order Levedev-Laikov spherical quadrature, for a total of 486 nodes. Bottom panel: F-DVR grid within an aug-cc-pVTZ-RI DVR auxiliary basis, with a total of 198 nodes. Gridpoints are colored according to the magnitude of the quadrature weight on a scale from red to blue in order of increasing weight.

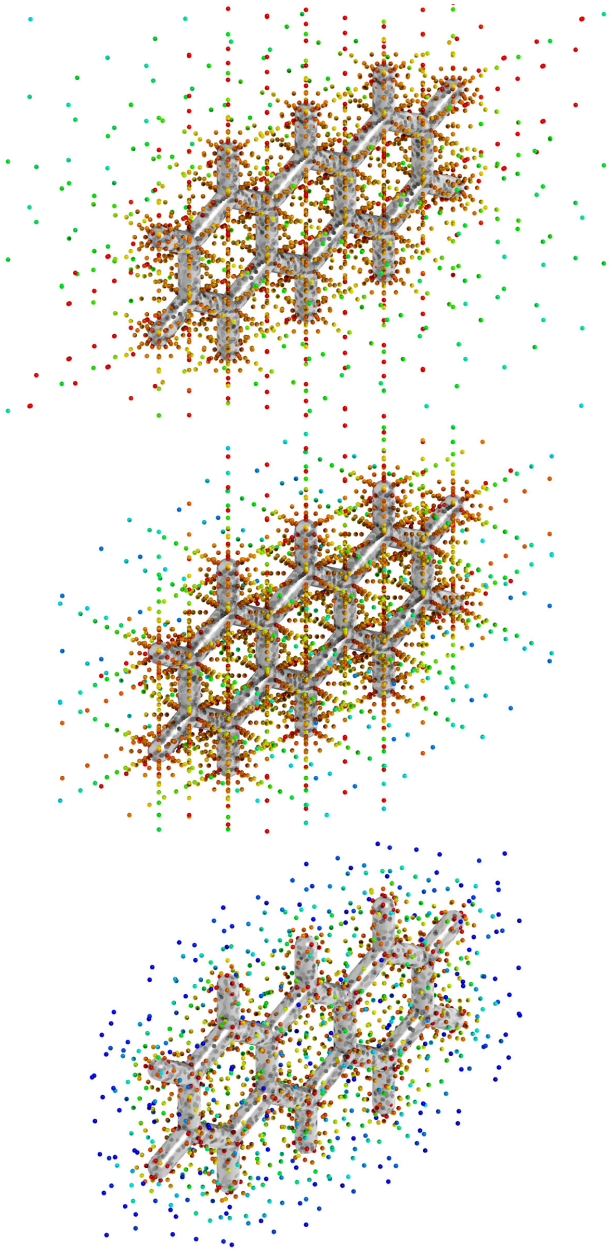

Figure 14: As in Figure 13, for the anthracene molecule. Top panel: Becke grid with 19/11-node Treutler/Ahlrichs radial quadratures (on heavy/hydrogen atoms, respectively) and a pruned 7-th order Lebedev-Laikov spherical quadrature, for a total of 3992 nodes. Middle panel: R-DVR grid within an aug-cc-pVTZ-RI DVR auxilairy basis, with a pruned 7-th order Levedev-Laikov spherical quadrature, for a total of 4300 nodes. Bottom panel: Full DVR grid within an aug-cc-pVTZ-RI DVR auxiliary basis, with a total of 1944 nodes.

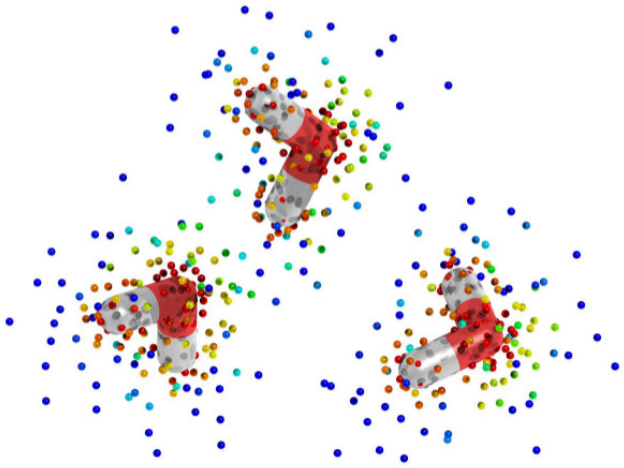

Figure 15: Full DVR grid for $(H_2O)_3$ within an aug-cc-pVTZ-RI DVR auxiliary basis. This grid contains 594 nodes.

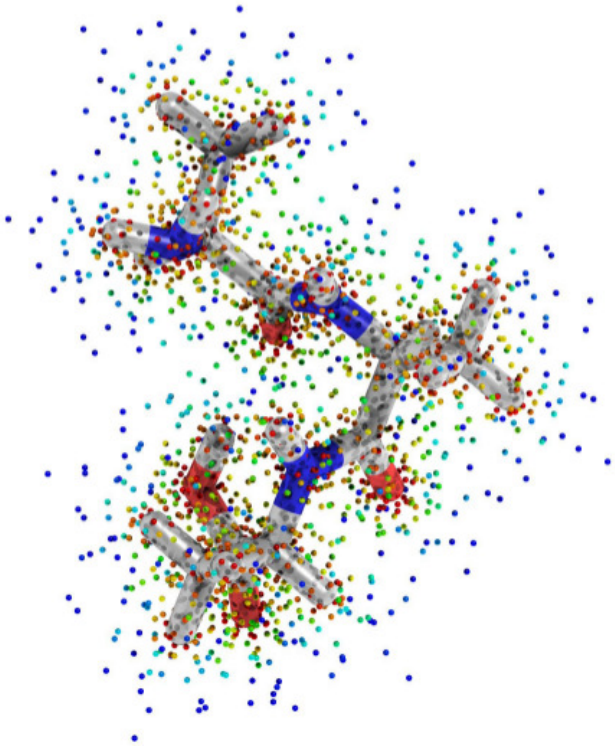

Figure 16: As in Figure 15, for the $(Ala)_3$ helix within an aug-cc-pVTZ-RI DVR auxiliary basis. This grid contains 2478 nodes.

in Figures 13 and 14, where the regular structure around each atom arises from the product of spherical and radial quadratures. This overt redundancy is somewhat ameliorated by the pruning scheme, which is evident in the larger spherical grids employed in the valence regions. The atomic partitioning is also visible, e.g., in the attenuation of the weight for a radial quadrature passing from one hydrogen through the other in the H_2O example. Additionally, both grids retain the Abelian point group of the underlying molecule, due to the choice of orientation and the octahedral symmetry of the Lebedev-Laikov grids. After noting these similarities, the differences between the Becke and R-DVR quadratures are also palpable: the Treutler/Ahlrichs radial quadrature tends to contain points both much closer to and much further away from the nucleus than R-DVR. Indeed, examination of the grid coordinate files shows that even with small 11- or 19-node radial grids, the Treutler radial quadratures extend significantly beyond the views shown, and also terminate several orders of magnitude closer to the nuclei than in R-DVR. It is quite plausible that tuning the Becke quadrature could fix this behavior, e.g., by changing the shape parameter α in the Treutler/Ahlrichs quadrature, or picking another mapping altogether. However, R-DVR is already accomplishing this goal, with explicit guidance from the DVR auxiliary basis set.

The F-DVR grids are another story altogether. As seen in Figures 13 to 16, these grids intrinsically conform to the shape of the molecule, and exhibit strong indicators of knowledge of the underlying basis set. In particular, there are no superfluous points, either near the periphery of the molecule, or in situations analogous to the extinguishing of points in partitioned atomic quadratures like Becke and R-DVR. Also notable, the F-DVR nodes seem to exhibit a shell-like structure, corresponding to the discrete set of exponents in the underlying basis. However, this shell structure is accomplished without the need for large numbers of points at each radial spacing to account for the spherical harmonics. Additionally, close inspection of the structure

for bonding regions shows that the grids exhibit nodal structure, avoiding regions at the bond midpoints (see, for instance, the front views of the anthracene F-DVR grid in the supplemental information).

While the F-DVR grids appear to capture the qualitative shape of the basis in an exceedingly efficient matter, several points merit future study. The first is that these grids do not retain any point-group symmetry of the underlying molecule. This is a generic and often correctable problem when working with discretizations of physical space; for instance, Becke quadratures with Lebedev-Laikov grids have been explicitly chosen to be complete in all Abelian point groups, but do not contain the tetrahedral point group. While the LS-THC fitting procedure will likely reduce this error to an acceptable level for many applications, possible contamination in gradient methods remains a valid concern. Projection or grid symmetrization techniques may be needed to remove this error in future studies. A related concern is the bifurcation of the Jacobi eigenvalue algorithm in the simultaneous diagonalization step: in the first Jacobi sweep, a large number of symmetries are broken based solely on the order in which the Jacobi rotations are performed. This implies that two slightly different geometries may produce markedly different F-DVR grids, potentially leading to oscillations on potential energy surfaces. A procedure to use the solution to a F-DVR grid from a previous computation to seed a qualitatively similar grid at a new geometry is under development in our groups. Finally, as mentioned in the computational details section, the effort currently required to obtain a converged F-DVR grid is considerable. However, sparsity considerations seem to offer the hope of a much faster algorithm in the near future.

4.4.3 Quantitative Grid Fidelity

The practical fidelity of the various grid schemes is investigated by comparing LS-THC-DF-MP2 to conventional MP2 for linear alkanes, linear alkenes, linear acenes,

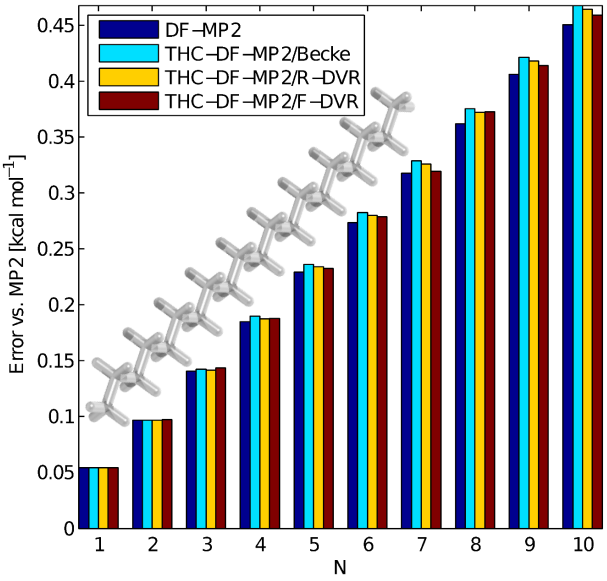

Figure 17: Energy error (kcal/mol) compared to conventional MP2 for DF-MP2 (blue bars), LS-THC-DF-MP2/Becke (turquoise bars), LS-THC-DF-MP2/R-DVR (yellow bars), and LS-THC-DF-MP2/F-DVR (red bars) for linear alkanes $C_{2N}H_{4N+2}$, with up to 20 carbon atoms. See Table 3 for size parameters.

alanine helices, and water clusters. These results are depicted in Figures 17 to 21, respectively. Each figure depicts the errors of LS-THC-DF-MP2 compared to conventional MP2 for Becke, R-DVR, and F-DVR grids in a cc-pVDZ[70] primary basis set, with the cc-pVDZ-RI[313] auxiliary basis used for the DF step, and the grids selected with an aug-cc-pVTZ-RI[313] DVR auxiliary basis as discussed above. The standard frozen core approximation was used, wherein the 1s electrons on heavy atoms are not included in the MP2 computation, as is standard practice with split-valence basis sets. The DF-MP2 error is included in each case to give a sense of the magnitude of error induced by the grid fitting. The sizes of the primary and auxiliary bases and the number of grid points for each of the grid generation methods are shown for the largest system of each class in Table 3.

For the F-DVR grids, we have found that "balancing" the quadrature weights to ensure that $S_{PP} = 1$, significantly improves the condition number of the grid metric

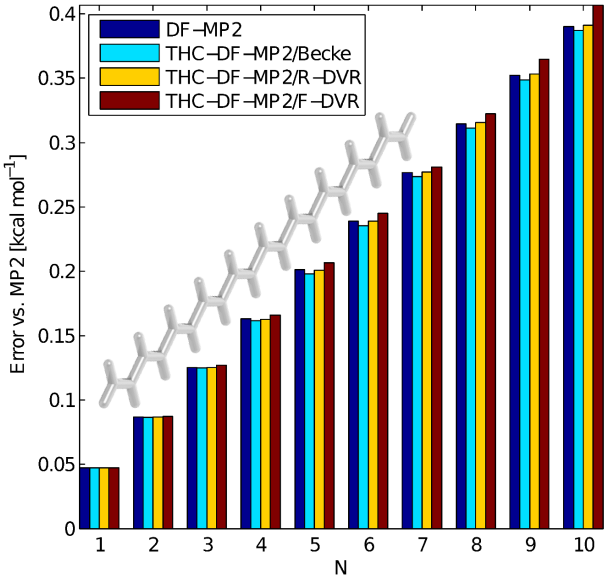

Figure 18: As in Figure 17, for linear alkenes $C_{2N}H_{4N}$, with up to 20 carbon atoms. See Table 3 for size parameters.

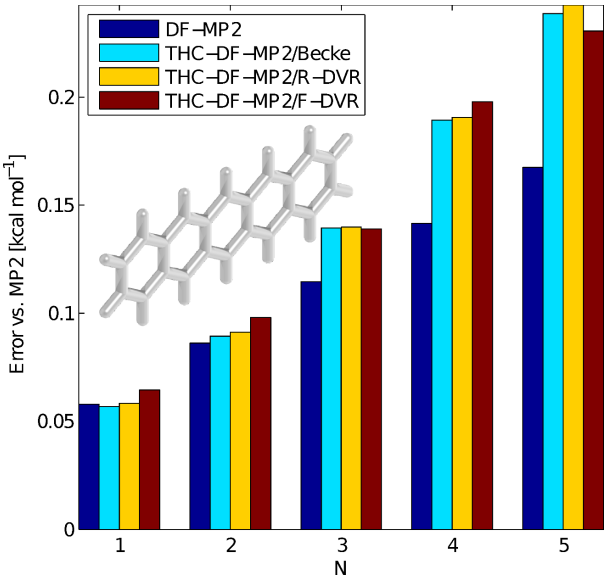

Figure 19: As in Figure 17, for linear acenes from benzene ($N = 1$) to pentacene ($N = 5$). See Table 3 for size parameters.

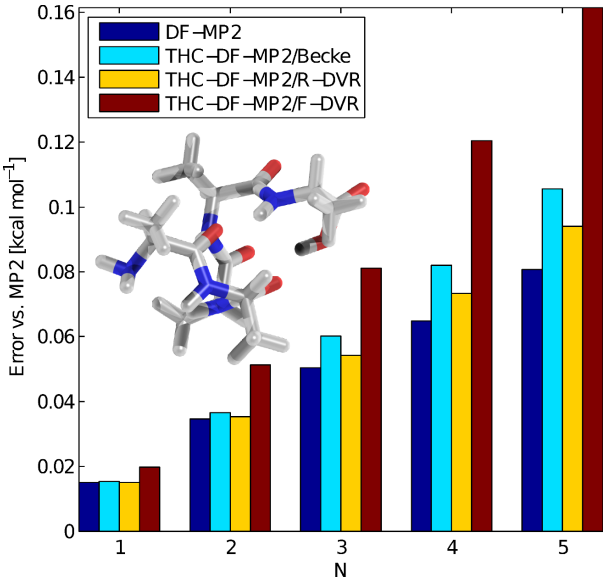

Figure 20: As in Figure 17, for helical alanine polypeptides, from $(\mathrm{Ala})_1$ to $(\mathrm{Ala})_5$. See Table 3 for size parameters.

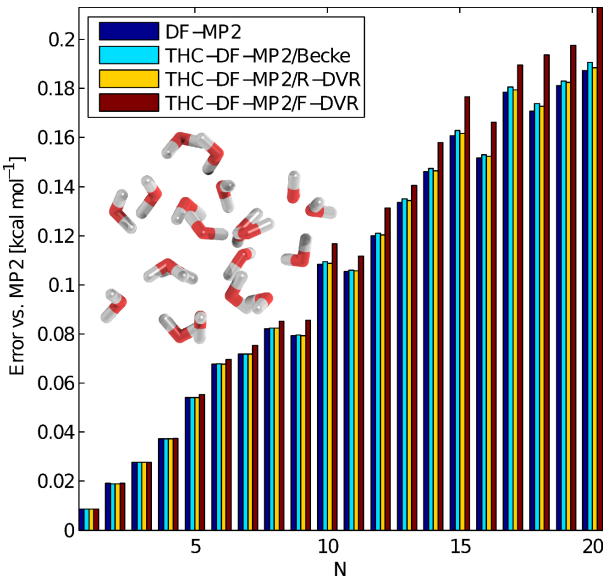

Figure 21: As in Figure 17, for water clusters from H_2O to $(H_2O)_{20}$. See Table 3 for size parameters.

Table 3: Number of primary basis functions N_p, DF auxiliary basis functions N_A, and LS-THC grid points N_P, for the largest system of each class encountered in this study. The primary basis set is cc-pVDZ, the DF auxiliary basis set is cc-pVDZ-RI, and the DVR auxiliary basis set is aug-cc-pVTZ-RI. The various grids are constructed as discussed in the Computational Details section.

System	N_p	N_A	N_P^{Becke}	$N_P^{\text{R-DVR}}$	$N_P^{\text{F-DVR}}$
$C_{20}H_{42}$	490	1708	9084	9524	4052
$C_{20}H_{22}$	390	1428	6644	7084	3132
Pentacene	378	1428	6064	6548	2976
Alanine Pentamer	499	1834	8442	9316	3998
$(H_2O)_{20}$	480	1680	8840	9720	3960

matrices in LS-THC, and improves the accuracy of LS-THC with F-DVR grids in some cases. The balancing is accomplished in the LS-THC phase of the procedure, when the desired *pq* quadratic product space is known. For each orbital space occuring in the quadratic product space, the F-DVR collocation is replaced by the balanced collocation,

$$X_p^{\prime P} \equiv v_P X_p^P : \; v_P = \frac{1}{\sqrt{X_p^P X_p^P}}. \tag{192}$$

This technique is used for all F-DVR results shown here. Unfortunately, the R-DVR and Becke grids employed here have vanishing weights as points from one atomic grid enter the Voronoi volume of another atom. This prevents balancing from being employed for R-DVR and Becke grids, as such such balancing would involve scaling the grid metric matrix by values arbitrarily near infinity. In practice, the eigenstructure of the grid metric matrix with the default R-DVR and Becke grids already reflects the true physical significance of the grid nodes, so the lack of balancing is not a problem for these two grid generation methods.

These five classes of systems all show largely the same qualitative trends, though the quantitative results do differ somewhat from plot to plot. In all cases, the grid fitting error (additional error over the underlying DF error) is explicitly zero for the first

few systems, regardless of the grid type invoked. This is because the equality in Equation 152 holds until the molecule gets large enough. At some N, the quadratic growth in the MP2 hole-particle orbital product space overcomes the number of points, and the grid fitting error rises linearly from this point. This seems unavoidable, but the results hown here demonstrate that the slopes of these grid fitting error terms are quite small.

In a quantitative sense, all three grid classes show excellent performance in LS-THC applications, with the grid fitting error being of the same order of magnitude as the DF error for the worst cases seen here, and far smaller in most cases. Somewhat surprisingly, the R-DVR and Becke grids show essentially the same error/efficiency characteristics, despite the fact that the latter were not optimized for LS-THC applications. This points to the both the robustness of the LS-THC fitting procedure and to the remarkable efficiency of existing Becke methodology. Despite the fact that the R-DVR grids produce no quantitative improvement over properly-chosen Becke grids in the present work, we are confident that if additional compression is attainable in atom-centered grids (e.g., by non-heuristic pruning or more-optimal radial node placement), R-DVR will discover it automatically.

The F-DVR grids attain similar accuracy to R-DVR or Becke grids, but with less than half as many points. We have experimented with reducing the Becke and R-DVR grids to the size of the underlying F-DVR grids by lowering the size of the radial DVR auxiliary basis or Becke radial quadrature and/or the size of the spherical Lebedev-Laikov grids, and have found that any such reduction severely compromises the accuracy of the R-DVR grid (to significantly worse than the DF errors encountered here). Thus, it appears that the non-direct-product nature of the F-DVR grids affords them with unusually high fidelity in LS-THC applications.

4.5 Conclusions

This manuscript has investigated three schemes for the construction of molecular quadratures for use in LS-THC representation of the ERI. The most straightforward method is the adoption of standard Becke-style grids typically encountered in DFT applications. Additionally, this work has introduced the R-DVR and F-DVR grid generation methods, which are based on physical understanding of the LS-THC problem as requiring efficient approximate quadratures for the overlap matrix in an auxiliary basis. Appealing to the DVR literature of the dynamics community, these classes of grids determine the molecular quadrature through consideration of the eigenstructure of the position operator in the finite DVR auxiliary basis. R-DVR diagonalizes the finite-basis representation of the radial position operator for each atom, and uses this explicit, one-dimensional DVR as the radial quadrature in a Becke-like direct-product grid. F-DVR approximately simultaneously diagonalizes the finite-basis representation of the full Cartesian position operator, and obtains a non-direct-product grid that efficiently encapsulates knowledge of the molecular geometry.

All three grid classes are shown to produce compact grids with LS-THC-DF-MP2 grid fitting errors of the same order or smaller as the underlying DF errors, with minimal tuning of the grid construction parameters. R-DVR and Becke schemes are both quite accurate, but both contain at least twice as many nodes as the corresponding F-DVR grid within the same DVR auxiliary basis. To this point, the R-DVR and Becke schemes have resisted all efforts to produce smaller grids with similar fidelity as F-DVR, indicating that the F-DVR grids provide a more compact representation of the relevant portions of physical space in the molecular environment.

The R-DVR and F-DVR results shown in this manuscript are quite encouraging, but also indicate that a significant amount of work remains, particularly for F-DVR. The R-DVR grids are easy to form (essentially the same cost as a Becke grid), directly encapsulate explicit knowledge of the span of the underlying basis set, are remarkably

accurate relative to F-DVR or bare DF, and do not suffer from the loss of Abelian symmetry or the discontinuous nature of the F-DVR grids. We therefore recommend the R-DVR scheme for production-level use until several more developments are completed with F-DVR. However, the results of this work suggest that F-DVR is uniquely efficient in encapsulating the necessary portions of physical space for LS-THC applications, and may soon become the grid scheme of choice, provided that a few improvements can be made. In particular, the F-DVR grids currently require an expensive iterative Jacobi procedure to form, for which sparsity considerations must be adopted to produce an acceptably efficient algorithm. Additionally, modifications must be made to ensure that the F-DVR grids produce smooth potential energy surfaces. Also, it would be highly desirable to produce F-DVR-like grids which preserve at least some of the point-group symmetry of the molecular geometry, e.g., for gradient or vibrational frequency computations. Finally, both F-DVR and R-DVR would surely benefit from a thorough optimization of the DVR auxiliary basis sets and other free parameters. Several initiatives are underway in our groups to accomplish these important developments.

Besides the obvious practical utility to LS-THC applications, it strikes us as exceedingly elegant that DVR can construct a wholly reasonable physical-space grid without ever "looking" at the structure of the molecule in position-space.

CHAPTER V

TENSOR HYPERCONTRACTION FOR COUPLED CLUSTER SINGLES AND DOUBLES

The following is adapted from Ref. 223.

5.1 Abstract

We apply orbital-weighted least-squares tensor hypercontraction decomposition of the electron repulsion integrals to accelerate the coupled cluster singles and doubles (CCSD) method. Using accurate and flexible low-rank factorizations of the electron repulsion integral tensor, we are able to reduce the scaling of the most vexing particle-particle ladder term in CCSD from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$, with remarkably low error. Combined with a T_1-transformed Hamiltonian, this leads to substantial practical accelerations against an optimized density-fitted CCSD implementation.

5.2 Introduction

Coupled cluster singles and doubles (CCSD)[239] has been a cornerstone method in electronic structure theory since its introduction over thirty years ago, e.g., as a critical portion of the "gold-standard" CCSD(T) method.[240] Unfortunately, CCSD is an extremely expensive method, scaling as $\mathcal{O}(o^2v^4)$ operations and requiring either explicit storage or integral-direct manipulations of a tensor scaling as $\mathcal{O}(v^4)$ elements, where o and v are the number of occupied and virtual orbitals in the system, respectively. Due to this combination of high importance and high expense, CCSD seems a prime target for the deployment of rank reduction methods. However, CCSD has proven to be remarkably resistant to formal simplification and practical acceleration by such techniques.

Among the most successful rank reduction methods in electronic structure are the density fitting (DF)[68,76,299] and Cholesky decomposition (CD)[24,254,151] compressions of the electron repulsion integral (ERI) tensor. While these methods have produced substantial formal and practical accelerations of mean-field and perturbation theory methods, they are unable to reduce the complexity of the rate-limiting term in CCSD.[245] Another interesting class of methods are the pseudospectral (PS)[80,84,187] and related chain-of-spheres (COS)[208] quadrature representations of the ERI tensor, which do possess the flexibility to formally reduce the complexity of the rate-limiting term in CCSD. However, these methods require a very large number of quadrature points to achieve sub-mH accuracy, and so have not yet achieved practical deployment in CCSD. Recently, we have introduced the tensor hypercontraction (THC)[121,216] representation for the ERI tensor, which provides a compelling combination of flexibility and accuracy. In this work, we will use the THC ERIs to reduce the complexity of the most vexing term in CCSD from $\mathcal{O}(o^2v^4)$ to $\mathcal{O}(o^2v^2M)$ (where M is the dimension of the THC quadrature grid), resulting in practical and accurate accelerations of the method as a whole.

Note that there exist several formal schemes to wholly reduce the scaling of CCSD below $\mathcal{O}(N^6)$, which we do not achieve in this work. Some authors have proposed factorizations of the amplitude tensors which might achieve scalings as low as $\mathcal{O}(N^5)$,[25,28] and we have recently shown that THC factorization of the ERIs *and* amplitudes can produce an accurate CCSD implementation scaling as $\mathcal{O}(N^4)$.[122] However, there are significant prefactors affiliated with such proposals. Another quite impressive area is the family of local correlation methods,[259,100] which have recently managed to produce CCSD(T)-quality results on systems as large as proteins.[255,252] The combination of THC and local correlation methodology lies outside the focus of the present study, but is certainly an important future development.

5.3 Theory

In this work, indices p, q, r, s (size n) represent active molecular orbitals, indices i, j, k, l (size o) represent active occupied virtual orbitals, indices a, b, c, d (size v) represent active virtual orbitals, indices A, B, C, D (size A) represent density fitting auxiliary functions, and indices P, Q, R, S (size M) represent tensor hypercontraction grid points. Often, $o \ll v \sim n < A < M$. Additionally, we will use a generalized Einstein summation convention in which repeated indices are asserted to be summed over. We also assume a restricted, closed-shell Hartree-Fock reference throughout.

5.3.1 LS-THC ERIs

The THC representation of the ERI tensor is,[121]

$$(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q. \tag{193}$$

Here $(pq|rs) \equiv \int_{\mathbb{R}^6} \mathrm{d}^3 r_1 \ \mathrm{d}^3 r_2 \ \phi_p(\vec{r}_1)\phi_q(\vec{r}_1) r_{12}^{-1} \phi_r(\vec{r}_2)\phi_s(\vec{r}_2)$ is the chemists' ERI. In the least-squares THC (LS-THC) method,[216] the factor $X_p^P \equiv \sqrt[4]{w_P}\phi_p(\vec{r}_P)$ is the weighted collocation matrix, where the quadrature grid nodes and weights are $\{< \vec{r}_P, w_P >\}$. The quantity Z^{PQ} is determined by least squares fitting to the target integrals, with orbital weights $\{\xi_p\}$ (explained shortly),

$$O(Z^{PQ}) = \frac{1}{2} \left\| \xi_p \xi_q \xi_r \xi_s \left[(pq|rs) - X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q \right] \right\|_2^2. \tag{194}$$

Thus THC is seen to arise from attempting a quantization of the spatial coordinates via $\vec{r}_1 \to \{\vec{r}_P\}$ and $\vec{r}_2 \to \{\vec{r}_Q\}$, and then replacing the singular grid operator $1/r_{12} \to \{1/r_{PQ}\}$ with a renormalized grid operator Z^{PQ}. This understanding of the THC representation has been shown to be exact within a linear number of grid points for polynomial basis sets,[218] and has proven to be an accurate approximation in several areas of molecular electronic structure theory.[217,119]

If conventional ERIs $(pq|rs)$ are decomposed by LS-THC, the determination of the factor Z^{PQ} requires $\mathcal{O}(N^5)$ operations, with a somewhat high prefactor. This can

be reduced by applying THC to DF ERIs,

$$(pq|rs) \approx (pq|A)(A|B)^{-1}(B|rs) = b^C_{pq} b^C_{rs}. \tag{195}$$

Here $(A|B)$ and $(pq|A)$ are two- and three-center ERIs, respectively. The quantity $b^C_{pq} \equiv (pq|A)(A|C)^{-1/2}$ is the symmetrically-fitted three-index tensor. If LS-THC is then invoked on these density fitted integrals [denoted in this work as LS-THC(DF)], the least-squares fittings in the bra and ket become decoupled, and Z^{PQ} has a profitable intermediate representation,

$$Z^{PQ} = L^P_C L^Q_C \tag{196}$$

where,

$$L^P_C = \left[S^{PQ}\right]^{-1} X^Q_{pq} \xi^2_p \xi^2_q b^C_{pq}. \tag{197}$$

Here $X^P_{pq} \equiv X^P_p X^P_q$ is the joint collocation, and $S^{PQ} \equiv X^P_{pq} \xi^2_p \xi^2_q X^Q_{pq}$ is the grid-basis Gramian matrix. The factor Z^{PQ} can be now computed in reduced $\mathcal{O}(N^4)$ effort.

Another plausible THC-type factorization of the ERI tensor (newly explored here) involves applying LS-THC quadrature to only the bra of the ERI, and leaving the ket indices pinned, i.e., $(pq|rs) \approx X^P_p X^P_q A^P_{rs}$. This factorization topologically resembles the PS representation, except that here it is the orbital-pair electrostatic potential A^P_{rs} that is determined by least-squares, rather than the physical-to-spectral projector $Q^P_p \equiv [X^Q_p X^Q_q]^{-1} X^P_q$ of standard PS. As such, the new THC-inspired method is considerably more accurate than raw PS. We refer to this variant of LS-THC as "partial THC," or LS-PTHC. If the ERI tensor is first represented by DF [LS-PTHC(DF)], the factorization reduces to,

$$(pq|rs) \approx X^P_p X^P_q L^P_C b^C_{rs}. \tag{198}$$

The factor L^P_C is identical to that in Equation 197.

Another new feature introduced in this work is the presence of the weights $\xi_p \xi_q \xi_r \xi_s$ in Equation 194. These weights allow us to obtain higher absolute accuracy in the

THC representation for certain integrals. One intuitive choice for the weights in the context of post-MP2 correlation methods involves the use of the unrelaxed MP2 difference density matrix γ'_{pq}. The eigenvectors U^r_p of γ'_{pq} are the usual unrelaxed MP2 natural orbitals, and the eigenvalues η'_r are the differences in occupation numbers of the natural orbitals between the SCF and MP2 wavefunctions. η'_r provides a guide for how much each natural orbital is depleted from or accumulated into in the correlation process. Therefore, if we choose $\xi_r \equiv \sqrt{|\eta'_r|}$ and perform the THC fitting in the natural orbital basis, we will weight the THC fitting to produce more absolute accuracy in integrals which are likely to be more active in the CCSD correlation process. This choice for the weights is denoted by W-LS-THC or W-LS-PTHC in this work, and is contrasted to our previous choice of $\xi_p \equiv 1$, which will be denoted by LS-THC or LS-PTHC.

5.3.2 THC-Accelerated CCSD

The most rate-limiting term in CCSD and related methods[270] [such as configuration interaction singles and doubles (CISD), coupled cluster doubles (CCD), configuration interaction doubles (CID), and third-order Mller-Plesset perturbation theory (MP3)] is the "particle-particle ladder diagram," A^{ab}_{ij}, which scales as $\mathcal{O}(o^2v^4)$ and involves a $\mathcal{O}(v^4)$ integral tensor,

$$A^{ab}_{ij} = t^{cd}_{ij}\hat{v}^{cd}_{ab}. \tag{199}$$

Here, we have invoked the T_1-transformation of the Hamiltonian[152] by setting $\hat{v}^{qs}_{pr} \equiv x_{p'p}y_{q'q}(p'q'|r's')x_{r'r}y_{s's}$, where $x_{p'p} \equiv \delta_{p'p} - t^a_i\delta_{p'i}\delta_{pa}$, and $y_{p'p} \equiv \delta_{p'p} + t^a_i\delta_{p'a}\delta_{pi}$. For CCSD, this embeds all commutators involving T_1 into an effective Hamiltonian, resulting in a drastically simpler set of residual equations (essentially CCD for the T_2 amplitudes, plus a few simple diagrams for the T_1 amplitudes). Moreover, this approach results in a formulation of CCSD with at most $\mathcal{O}(o^3v^3)$ operations on $\mathcal{O}(o^2v^2)$ tensors, with the exception of the above diagram. As a result, A^{ab}_{ij} is often wholly

rate-limiting, particularly in larger basis sets.

As was first noted in the context of CID and MP3,[189,190] this term can be reduced to $\mathcal{O}(3o^2v^2M)$ via integrals represented with PS-like topology,

$$A_{ij}^{ab} \approx \left[X_a^P \left[A_{bd}^P \left[Y_c^P t_{ij}^{cd}\right]\right]\right]. \tag{200}$$

Here $X_a^P \equiv X_p^P x_{pa}$ and $Y_a^P \equiv X_p^P y_{pa}$ are the T_1-transformed collocations. In this work, A_{bd}^P is a PS-like intermediate formed from LS-THC(DF) or LS-PTHC(DF) integrals, and the underlying L_A^P operators are fit in the full active MO space to provide fidelity in the T_1-transformed integrals. For LS-THC(DF), $A_{bd}^P \equiv Z^{PQ}[X_b^Q Y_d^Q]$, which requires $\mathcal{O}(v^2M^2)$ overhead. For LS-PTHC(DF), $A_{bd}^P \equiv L_C^P \hat{b}_{bd}^C$, where $\hat{b}_{bd}^C \equiv b_{pq}^C x_{pb} y_{qd}$, which requires lower $\mathcal{O}(v^2AM)$ overhead. We symmetrize the result, $A_{ij}^{ab} \leftarrow 1/2(A_{ij}^{ab} + A_{ji}^{ba})$. Note that blocking over the index P provides an immediate route to avoid storage of large intermediates such as $T_{ij}^{Pd} \equiv Y_c^P t_{ij}^{cd}$; only T_2-sized quantities are required [and $\hat{b}_{bd}^C$ for LS-PTHC(DF)]. Full LS-THC(DF) and LS-PTHC(DF) algorithms for A_{ij}^{ab} are presented in the supplemental material.

Conventionally, A_{ij}^{ab} can be evaluated in $\mathcal{O}(o^2v^4/4)$ operations involving $\mathcal{O}(v^4/2)$ integrals by using symmetric and antisymmetric combinations of t_{ij}^{cd} and $\hat{v}_{ab}^{cd}$.[267] Avoiding explicit storage or I/O operations involving $\hat{v}_{ab}^{cd}$ is of critical importance, particularly when the T_1-transformation is invoked. Integral-direct algorithms in the atomic orbital basis provide one avenue for such acceleration.[152] Additionally, DF algorithms provide an alternative means to accelerate the T_1-transformation of the Hamiltonian, and reduce I/O operations by generating $\hat{v}_{ab}^{cd}$ in batches, albeit with computationally expensive $\mathcal{O}(v^4A/2)$ floating-point overhead.[60]

5.3.3 Computational Details

LS-THC(DF) and LS-PTHC(DF) implementations of A_{ij}^{ab} have been added to an optimized DF-CCSD implementation in a development version of the `PSI4` package.[298] All computations below use the cc-pVTZ basis set,[70] with a DF-RHF reference in the

cc-pVTZ-JKFIT auxiliary basis,[312] and all ERIs in DF-CCSD are represented by the cc-pVTZ-MP2FIT auxiliary basis.[313] The THC grids for these computations are pruned Becke-style grids[20] extended from our R-DVR grids[217] and optimized for LS-THC(DF)-MP2/cc-pVTZ in a separate study.[154] All 1s electrons on first-row atoms are frozen in the CCSD computations. All molecular geometries, THC grids, and benchmark results are available in the supplemental material.

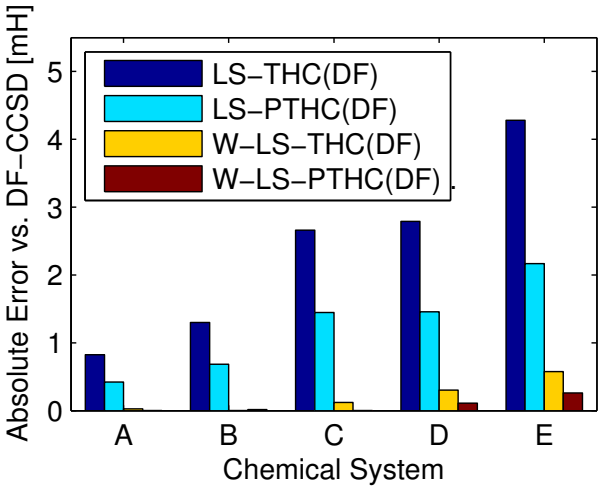

Figure 22: Absolute errors vs. DF-CCSD/cc-pVTZ for (A) $C_{12}H_{26}$ alkane (B) $C_{12}H_{14}$ alkene (C) tetracene (D) 3-alanine (E) $(H_2O)_{12}$.

5.4 Results and Discussion

Figure 22 depicts the absolute error profiles of the four THC-type CCSD algorithms vs. DF-CCSD, in a cc-pVTZ basis set, for the largest members computed of our standard THC test sets. The performance for LS-THC(DF)/LS-PTHC(DF) is reasonable (at worst 4.28 mH absolute error), while that of W-LS-THC(DF)/W-LS-PTHC(DF) is excellent (at worst 0.58 mH absolute error). In all cases, the LS-PTHC variant has roughly half the absolute error as the LS-THC variant, as expected. The W-LS-PTHC(DF) errors are remarkably low, reaching at most 0.02 mH absolute error for the three hydrocarbon test sets, with absolute errors of 0.11 mH and 0.26 mH for

3-alanine and $(H_2O)_{12}$, respectively.

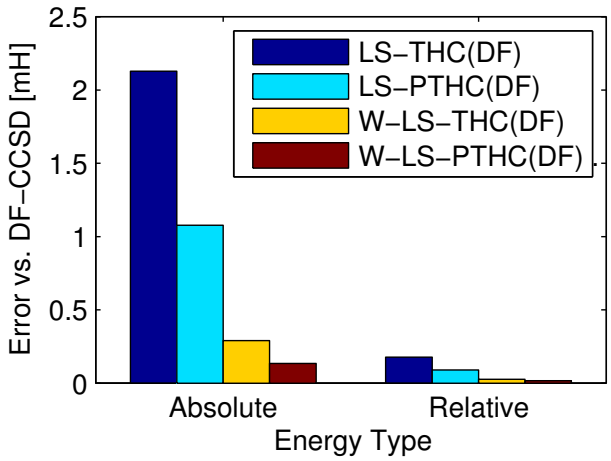

Figure 23: Maximum errors for absolute and relative energies vs. DF-CCSD/cc-pVTZ for eight low-lying $(H_2O)_6$ clusters.

It is also worth noting that THC seems to possess excellent error cancelation properties. For instance, Figure 23 contrasts the maximum errors in absolute and relative energies of eight low-lying $(H_2O)_6$ clusters.[17] Errors in absolute energies are up to 2.13 mH and 0.29 mH for LS-THC(DF)/LS-PTHC(DF) and W-LS-THC(DF)/W-LS-PTHC(DF), respectively, but the corresponding maximum errors for relative energies are only 0.18 mH and 0.03 mH, i.e., a tenfold reduction in relative compared to absolute error.

Figure 24 depicts the wall time for a complete CCSD/cc-pVTZ computation of tetracene for DF, LS-THC(DF)/W-LS-THC(DF), and LS-PTHC(DF)/W-LS-PTHC(DF), on a node with 6-core 3.2 GHz Intel Core i7 processor, 64 GB of DDR3 memory, and 3×3 TB 7200 RPM RAID0 scratch disks. Weighted and non-weighted THC/PTHC algorithms require the same wall times. The DF-CCSD computation requires 30.7 hours of wall time, with 19.9 hours of that spent in A_{ij}^{ab}. The corresponding values are 13.5/2.8 hours for LS-THC(DF) and 12.7/1.9 hours for LS-PTHC(DF). A_{ij}^{ab} is thus accelerated by a factor of 10.4×, leading to 2.4× overall speedup of CCSD. We expect

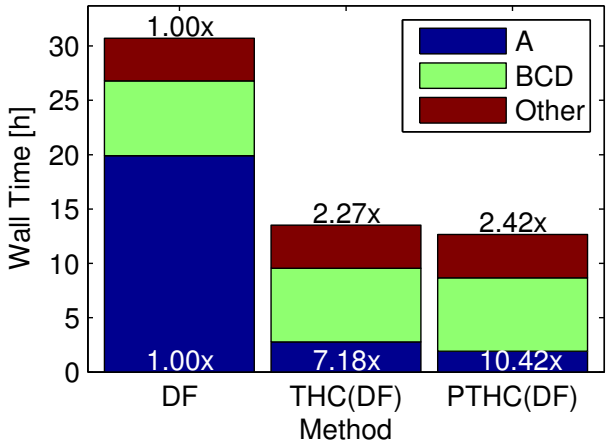

Figure 24: CCSD/cc-pVTZ wall times for tetracene. A indicates the A_{ij}^{ab} particle-particle ladder diagram, BCD indicates the remaining $\mathcal{O}(N^6)$ diagrams, and Other indicates $\mathcal{O}(N^5)$ or lower overhead. 19 CCSD iterations are performed.

this value to grow substantially with system and basis set size. Similar speedups are obtained against `MOLPRO`,[314] which uses an AO-basis algorithm for A_{ij}^{ab}.[266]

This communication has demonstrated that straightforward application of the THC representation of the ERI tensor can produce substantial computational gains in methods such as CCSD, with remarkably little error. As CCSD has been quite resistant to acceleration by DF/CD/PS methods for the past three decades, it is apparent that THC is a uniquely powerful rank reduction method in electronic structure theory. A particularly surprising observation is that the accuracy/efficiency results presented in this work use THC grids optimized for the hole-particle integrals of LS-THC(DF)-MP2: additional accuracy/efficiency may be possible by reoptimizing the THC grids for the full MO space.

CHAPTER VI

CONCLUSIONS REGARDING TENSOR HYPERCONTRACTION

Tensor hypercontraction has been established as an accurate, efficient, and markedly flexible factorization of the ERI tensor. THC compresses the ERI tensor according to the representation $(pq|rs) \approx X_p^P X_q^P Z^{PQ} X_r^Q X_s^Q$, and is the first rank reduction approach which is accurately able to decompose the order-4 ERI tensor in terms of only order-2 factor tensors. The topology of the THC representation is physically motivated, and arises from a two-stage argument. The first stage is to exploit the linear closure and spatial separability of the pair space to apply a quadrature in *both* coordinates of the ERI tensor. The second stage is to recognize that the singularities of the $1/r_{12}$ operator which plague the quadrature may be removed by replacing $1/r_{12}$ with a renormalized operator which is customized for the finite pair space present in the problem at hand. In polynomial basis sets (X-THC), these steps rely on exact linear closure and Gaussian quadrature, and produce an exact factorization. In molecular problems with atom-centered basis sets, linear closure in the pair space is only numerically obtained, and linear-scaling quadrature grids are only approximate for integrating the pair space. As a result, THC in the molecular problem is necessarily an approximation, albeit a rapidly convergent one. For the molecular problem, there are several possible recipes to generate the THC factorization. One of the most simple and robust is least-squares tensor hypercontraction (LS-THC). LS-THC is similar in spirit to X-THC, in that it applies a quadrature in both coordinates of the ERI tensor and then seeks a renormalized operator to replace the singularity-laden $1/r_{12}$. In practice, LS-THC only needs the grid-space collocations of the renormalized operator.

These collocations, which make up the Z^{PQ} tensor, are effectively determined by solving for the Z^{PQ} which minimizes the 2-norm of the error in the desired ERI tensor, under the choice of X_p^P from the quadrature grid. THC can be viewed as a one-step PARAFAC decomposition of the ERI in which only the Z^{PQ} tensor is optimized. It is shown that LS-THC ERIs with either atom-centered Becke-type quadrature grids or grids based on discrete variable representation (DVR) approaches can provide sub-mH accuracy in many electronic structure theory methods with as few as 100-200 points per atom. Moreover LS-THC demonstrates systematic cancellation of errors in computing relative properties.

There are several advantages to be had from the use of THC ERIs. One is that the order-2 factors of THC are much easier to store than the order-4 ERI tensor or the order-3 factor tensors of other factorization approaches. This should be particularly relevant on massively parallel or GPGPU architectures, where communication operations involving large tensors can quickly annihilate performance. Additionally, the uniquely flexible "unpinned" THC factorization allows for a number of new algorithms for electronic structure theory methods with lower scalings than previously observed. The use of LS-THC ERIs is shown to produce the first wholly $\mathcal{O}(N^4)$ approaches to MP2, MP3, CC2, and EOM-CC2. Moreover, LS-THC ERIs are shown to accurately and efficiently reduce the scaling of the rate-limiting particle-particle ladder terms in CCSD from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$, providing for a practical speedup of $10\times$ in this term and $2.4\times$ in the complete CCSD method. Further afield, the THC factorization has also been asserted for the opposite-spin doubles amplitudes of CI and CC theories, and is shown to result in an accurate and formally $\mathcal{O}(N^4)$ approach to CISD and CCSD.

The application of THC electronic structure theory introduces some unique challenges, many of which may be solved in the near future. An overriding requirement in

LS-THC is the specification of quadrature grids for use with each basis set and subset of the ERI tensor. Optimization techniques similar to those used to parametrize auxiliary basis sets for DF computations have been developed, and are beginning to produce markedly robust and efficient grids for a range of basis sets. Another challenge is the derivation of analytical gradients for the LS-THC ERIs - for AO-basis LS-THC, this has been completed and is currently being tested. An additional, global, challenge is the increased complexity of the equations of electronic structure theory when THC representations are invoked - the large number of factors in a THC-approximated quantity result in a hefty number of intermediate contractions that must be performed. Determining and then implementing the optimal path through a given contraction has previously been quite time consuming for the programmer. This situation is slowly being ameliorated by the introduction of tensor manipulation packages into electronic structure code stacks. A final consideration is that lower-scaling algorithms with high prefactors do not always beat higher-scaling algorithms with low prefactors - there is always a crossover point with respect to system size that must be considered. In particular, it has proven difficult to universally beat the existing state-of-the-art density fitted codes, especially at high accuracies. That being said, the tide is beginning to turn. For methods like CCSD, where DF is not flexible enough to provide much acceleration, THC is already providing clear advantages. Even for methods where DF has reigned supreme (such as MP2, CC2, and EOM-CC2), THC is beginning to provide crossovers, particularly in implementations on state-of-the-art architectures such as GPGPUs.

In the end, the THC factorization of the ERI tensor arises as a natural exploitation of the understanding of the ERI tensor as an amalgamation of five separate objects - four basis functions and one Coulomb operator, stitched together over two spatial coordinates. In exposing this underlying structure, an accurate and flexible representation for the electron-electron repulsion is realized, and is beginning to be

practically exploited to accelerate electronic structure computations of many types.

CHAPTER VII

INTRODUCTION TO PARTITIONED INTERMOLECULAR INTERACTIONS

7.1 Intermolecular Interactions

7.1.1 Background to Fuzzy Chemical Concepts

At the end of the day, the electronic Schrödinger equation (or any desired extension thereof) is believed to exactly govern all that we can know about any chemical system. To say that the even the exact results obtained by solving this equation are equivocal is an understatement of massive proportions. For example, the Schrödinger equation does not provide us with the definite energy of a chemical system, nor the position of a given electron within that system. Rather, a list of *possible* discrete energies and their probabilities is provided, along with the probability density function for finding *any* electron within a given region of space. And from this ansatz, we are expected to provide design input into the chemical technologies of tomorrow - to decide what atomic-level substitutions will yield the next drug molecule, the next organic photovoltaic device, the next catalytic material. How then are we to proceed?

The answer is the broad field of "fuzzy" chemical concepts - discrete and simplified governing precepts extracted from the continuous reality of quantum chemistry. In covalent chemistry, these concepts are well-developed and widely used - the chemical bond,[226] the Lewis structure,[173] the electronegativity tables,[225] or more-advanced heuristics such as frontier orbitals[86] and the Woodward-Hoffmann rules.[327] These concepts do not correspond to any unambiguous physical observable, and thus no instrument has ever directly observed one. And yet they resonate so strongly with the outcome of a broad range of experiments that they provide useful design information

and allow for tractable progress in the field of chemistry. The simultaneous utility and non-existence of these concepts is perhaps most elegantly stated by Frenking and Krapp,[79] who liken the concept of chemical bonding to the legend of the unicorn, "A unicorn is a mystical animal whose appearance is known to everybody although nobody has ever seen one."

There seem to exist two broad approaches to these fuzzy concepts: chemical intuition and heuristic computation. The first of these is these is the core of any organic chemistry course - the drawing of Lewis structures, the application of the curved arrow notation, the annotation of partial charges, the invocation of a mix of resonance structures. Here, the key is that no computations or numerical calculations are performed. The results fall out of a richly developed set of empirical rules, where, at each stage in the process, the chemist *asserts* that this bond or that dipole exists because the diagram on the paper looks right. The magnificent finding here is that this process works in the overwhelming majority of chemical cases, distinguishing it from being voodoo by virtue of usually being right. In fact, the only real issue with chemical intuition is that it only pertains to chemical problems which fall within the span of other exhaustively-studied problems. In delicate cases or exotic chemical environments, chemical intuition may sometimes fail. In these cases, the second approach of heuristic computation can often provide deep insights. The precept of this area is that the chemist's overall concept exists, but his/her intuition is incorrect for the particular chemical system at hand. The heuristic computation approach instead carefully develops a robust mathematical specification of the overall concept, and then lets the computer find the specific details for each system. E.g., in describing the electrons in benzene, the overall concept of localized electronic bonds is correct, but the assertion of alternating single and double bonds in the carbon ring (the Kekulé structure) is rubbish. However, if one instead performs an *ab initio* SCF computation and then asserts the overall idea that the electrons can be represented

as spatially localized objects (an orbital localization methodology), the generalized equivalent of the Lewis structure of benzene falls out immediately. Moreover, all of the salient details of this generalized Lewis structure are preserved - the result sums to the total *ab initio* density of the system, the polarization of the C-H σ bonds are encapsulated, and the π orbitals prove difficult to uniquely localize (in contrast to the σ framework), indicating that they form a non-separable chemical subsystem.

Within covalent chemistry, both the chemical intuition and the heuristic computation approaches to fuzzy chemical concepts are quite mature. With regard to noncovalent chemistry, i.e., intermolecular interactions, both the concepts and the heuristics are less-well developed. The goal of this half of the thesis is to detail our efforts to provide a robust heuristic partition of *ab initio* interaction energy components as one step along the path toward a firm set of intuitive concepts for noncovalent chemistry.

7.1.2 Noncovalent Chemistry

Broadly speaking, noncovalent chemistry is concerned with the weak residual forces leftover between molecules after the constituent electrons and protons have combined to form these molecules by covalent bonding.[203] Whereas a typical covalent bond energy is on the order of 100 kcal mol^{-1}, typical intermolecular contacts are in the range of 1 - 20 kcal mol^{-1}. As with all discrete concepts in chemistry, there is a somewhat blurry line between covalent and noncovalent interactions, e.g., as demonstrated by the longstanding debate as to whether hydrogen bonding has covalent character.[132] In general, intermolecular interactions are characterized by such forces as electrostatics, steric repulsion, polarization, and London dispersion, and also by the absence of significant charge transfer or charge sharing between the monomers in question. Such noncovalent interactions are ubiquitous in chemistry, and are the driving force in a host of important processes. The supramolecular structure of DNA and

RNA,[115,282,52,229,289,213,125] affinity of drug-ligand interactions,[32,164,168,198,113,174,155,74] and packing of organic crystals[230,247,231,29,127] are just a few examples of systems in which intermolecular interactions are key players.

The two most critical questions asked in the study of intermolecular interactions are (1) how strong is the interaction and (2) why does it exhibit that strength? There are actually several ways to phrase (1), depending on the situation. In comparing to most experiments, one wishes to calculate the binding energy ΔG_{bind} at finite temperature in whatever solvent environment is used in the experiment.[203] That is, what is the free energy required/released to bring the two monomers together to form the dimer, including the effects of zero-point molecular vibrations, solvent interactions, entropy, enthalpy, monomer deformation, and dimer interactions. Computing all of these effects with high precision is an extraordinarily difficult process, and is overwhelmingly likely to be intractable for a given chemical system. In order to make some progress, we typically start instead with the last effect under consideration: the interaction energy ΔE_{int} of the frozen monomers. For some experimental systems, especially those for which the solvent effects do not dominate the binding energy and for which the monomers are reasonably rigid, this is sufficient to provide significant insights. Even just accurate computation of the interaction energy is a substantial task, due to the sensitivity of the interaction to the underlying electronic structures of the monomers. Several systematic techniques for computing the interaction energy have been developed, including the supermolecular approach and the perturbative approach.

A systematic answer for question (2) is critical to allow for rational design of noncovalent interactions, e.g., to design new drug ligands or better organic catalysts. However, this area remains quite murky. One useful arena is the application of energy decomposition analysis (EDA)[147] to partition the *ab initio* interaction energy

into electostatics, steric repulsion (also called exchange), induction (a mixture of polarization and charge transfer), and dispersion. Many EDA-type approaches have been proposed, including Morokuma's EDA,[147] symmetry-adapted perturbation theory (SAPT),[139,291,138] absolutely-localized molecular orbital (ALMO) techniques,[13] etc. For small noncovalent interactions, EDA provides substantial insight into the nature of the interaction. However, for larger systems involving multiple functional groups on each monomers, one may find that all EDA terms play a significant role in determining the interaction energy. E.g., in phenol dimer, one expects a mix of hydrogen bonding and π stacking interactions from the hydroxyl and phenyl pieces, respectively. Unfortunately, EDA methodologies generally do not contain any quantification of the spatial origins of the interaction energy terms, but rather only their total magnitudes.

For larger and more-complicated systems, chemists often talk intuitively in terms of "contacts" between nearby functional groups on opposing monomers. These contacts are often selected intuitively on the basis of functional group composition, distance criteria, and the relative orientations of the functional groups in question. For instance, when one sees an intermolecular motif involving the topology $X-H\cdots:Y$, where X and Y are electronegative atoms and the $H\cdots Y$ distance is in the range of 1.5 to 2.5 Å, one immediately thinks "hydrogen bond." This implies a close intermolecular contact between two favorably oriented dipolar groups, contributing a fairly large attractive component to the interaction energy (typically $2-7$ kcal mol^{-1} depending on the specifics of X, Y, and the geometry). Other well-known contacts include halogen bonding, π-π stacking, CH-π, cation-π, etc. The majority of these intermolecular bonding motifs are motivated by electrostatic analysis, e.g., consideration of the local multipoles of the involved functional groups. As such, these intuitive classifications may significantly underestimate the strength of London dispersion contributions, which can be substantial even for such seemingly noninteracting motifs as

CH–CH contacts. Moreover, intuition often fails to capture a number of long-ranged or nonclassical effects. For instance, electrostatic interactions are notoriously long-ranged, and thus distance-based criteria may completely fail to identify the most important electrostatic contacts in a large noncovalent complex. Regarding nonclassical effects, one somewhat surprising finding in recent years is the importance of a nonclassical "charge penetration" term in the electrostatics to a wide variety of noncovalent interactions. Charge penetration arises due to the more-diffuse nature of the electronic density field, relative to the point-charge distributions of the nuclei. For the interaction between, for example, two neutral rare-gas atoms, the four contributions to the electrostatic interaction (electron-electron repulsion, electron-nuclear attraction, nuclear-electron attraction, and nuclear-nuclear repulsion) turn on at different rates as $r_{AB} \rightarrow 0^+$. In the end, this leads to a small but exponentially increasing electrostatic attraction between the ostensibly neutral rare gas atoms as their electronic densities begin to interpenetrate.

These considerations have led to a general sentiment that it would be very useful if there existed a robust methodology to quantify the strength of each *ab initio* interaction energy term between local pieces of two interacting monomers.[286] This lies exactly in the spirit of heuristic computation approaches - the idea is to assert the chemist's overall picture of effective pairwise contacts between local fragments of two interacting monomers, but let the computational algorithms robustly decide on the exact quantitative makeup of these fragments, as well as on the assignment of the interaction energy terms to these fragments.

7.1.3 Supermolecular Computations

The intermolecular interaction energy is defined as,

$$\Delta E_{\text{int}} \equiv E_{\text{AB}} - E_{\text{A}} - E_{\text{B}}. \tag{201}$$

That is, what is the energy required/released to bring the system to the geometry of the interacting dimer (which has energy E_{AB}) from the effective system of infinitely separated monomers frozen in their dimer conformations (which has energy E_A+E_B)? A negative interaction energy indicates a favorable noncovalent interaction.

Technically, any *ab initio* method which computes the total electronic energy can be deployed for the computation of the three terms in Equation 201, leading to the conceptually simple "supermolecular" approach. However, a number of important considerations must be accounted for in such an approach. The first is that Equation 201 involves a small energy difference (on the order of a few kcal mol^{-1}) arising from three very large total energies (on the order of hundreds or thousands of Hartree each). Therefore, the individual energy computations must be converged tightly, and any numerical approximations deployed must be carefully designed to prevent errors from overwhelming the energy difference. This is one area where the importance of favorable error cancellation is manifest - e.g., an energy computation using the DF approximation will often make a large absolute error in describing the core electrons of the dimer, but this will generally be cancelled by a corresponding error in the core electrons of the monomer computations, leading to much reduced absolute error in the interaction energy. The second consideration is the choice of one-particle basis set employed. If finite atom-centered basis sets are used, an insidious difference in the completeness of the basis set in the dimer and monomer computations is encountered, known as basis set superposition error (BSSE). The brief explanation of BSSE is that the dimer computation (performed in the dimer basis set) is effectively performed in a more complete basis set than each of the monomer computations (performed in their respective monomer basis sets), because the electrons of one of the monomers in the dimer computation can "steal" parts of the basis functions of the opposing monomer. This artificially stabilizes the dimer, leading to a too-favorable interaction energy. Improved performance can often be obtained by applying a Boys and

Bernardi "counterpoise correction," which amounts to also performing the monomer computations in the dimer basis sets.[35] However, it has been demonstrated that this formally overcorrects for the effect of BSSE. The debate on this matter continues to this day - the current prevailing wisdom is that counterpoise correction does not directly fix BSSE, but it does improve the smoothness of convergence with respect to basis set.[163]

The third major consideration in supermolecular approach is the extreme sensitivity of the interaction energy to the electronic structure of and between the underlying monomers and to the relative distances and orientations between the two monomers. For one instance, Hartree-Fock theory or uncorrected Kohn-Sham Density Functional Theory (KS-DFT)[152] does not account for the long-range correlations between electrons, and thus does not capture the dispersion interaction at all. Empirical dispersion corrections to DFT (-D approaches)[96] or second-order Møller-Plesset Perturbation Theory (MP2)[200] are essentially the first two methods to even begin to capture this interaction. Even at the level of MP2, dispersion is very poorly captured - the delicate interplay between intramolecular and intermolecular correlation (coupling) is not captured in MP2, leading to a large overestimation of the MP2-type dispersion energy for some classes of systems. In general, extremely accurate and expensive methods, such as coupled-cluster singles and doubles with perturbation triples extrapolated to the complete basis set [CCSD(T)/CBS] are required for the robust computation of interaction energies.[239,240,71] The massive expense of CCSD(T)/CBS has led to a slew of proposals for cheaper methods with higher accuracy than MP2 for the characterization of interaction energies - spin-component scaled MP2 (SCS-MP2),[95] "coupled" MP2 (MP2C),[112] spin-component scaled coupled cluster (SCS-CCSD),[292] etc. One admirably simple recipe is to perform MP2 in a carefully chosen basis set (jun-cc-pVDZ),[272] which provides remarkable efficiency and accuracy vs. CCSD(T)/CBS due to cancellation between the basis set incompleteness errors and the errors rising from

the uncoupled treatment of dispersion in MP2.

There exist several EDA methodologies based on supermolecular computations - Morokuma's EDA,[147] Head-Gordon's ALMO methodology,[13] etc. However, these methods do not rigorously converge to the many-body limit of the interaction energy, as they obtain the contributions to the interaction energy as byproducts of supermolecular computations.

An approach that directly computes the interaction energy may have the ability to naturally avoid many of the issues discussed in this section. This leads to the "perturbative" approach to interaction energy computations.

7.1.4 Symmetry-Adapted Perturbation Theory

Perhaps the best-developed perturbative approach for intermolecular interactions is symmetry-adapted perturbation theory (SAPT).[139,291,138] SAPT is an extension of the older London variant of Raleigh-Schrödinger polarization theory (RSPT), with corrections added at each order to account for the fermionic antisymmetry of the dimer wavefunction. For a molecular dimer, SAPT partitions the molecular Hamiltonian as,

$$\hat{H} \equiv \hat{F}_A + \hat{F}_B + \hat{V} + \hat{W}_A + \hat{W}_B. \tag{202}$$

Here $\hat{F}_A$ is the Fock operator of monomer A, while $\hat{W}_A$ is the corresponding intramolecular Møller-Plesset fluctuation potential. $\hat{V}$ is the the intermolecular interaction operator between the particles of monomers A and B. SAPT expands the expectation value of $\hat{V}$ (the interaction energy) as a triple perturbation series in $\hat{V}$ and the two intramolecular fluctuation potentials. The zeroth-order dimer wavefunction is a non-antisymmetrized direct product of the Slater determinants of the two underlying monomers, i.e., the Hartree-Fock wavefunctions of the separated monomers. The SAPT expansion is then written as,

$$E_{\text{int}} \equiv \sum_{nlm}^{\infty} E_{\text{pol}}^{(nlm)} + E_{\text{exch}}^{(nlm)}$$

Here, the tuple (nlm) denotes the perturbation series order in $\hat{V}$ (n), $\hat{W}_A$ (l), and $\hat{W}_B$ (m). The reduced specification $E_{\text{term}}^{(n,l+m)}$ is also often encountered. Each SAPT term has two contributions - the RSPT polarization series term $E_{\text{pol}}^{(nlm)}$ and an exchange correction $E_{\text{exch}}^{(nlm)}$ to provide for antisymmetry in the dimer wavefunction. A thorough review of the development and derivation of SAPT terms may be found in Ref. 139.

Many flavors of SAPT have been developed over the past few decades. In the more-traditional "wavefunction-based" SAPT above, the simplest method is SAPT0,[138] which neglects all intramolecular correlation due to the fluctuation potentials. SAPT0 contains the leading terms representing electrostatics, exchange, induction, and dispersion, and is essentially a Hartree-Fock treatment of the first three terms, plus an MP2-type dispersion treatment. Just as with MP2, SAPT0 may significantly overbind polarizable systems (such as π-π contacts), due to the uncoupled treatment of the dispersion term. As with MP2, the deployment of SAPT0 with a carefully chosen jun-cc-pVDZ basis provides a remarkably accurate empirical correction to this problem, and yields interaction energies which are usually within $0.5 - 1$ kcal mol^{-1} of CCSD(T)/CBS, at a small fraction of the cost.[123,214] When implemented with density fitting approximations, SAPT0/jun-cc-pVDZ has been routinely deployed on systems with up to 220 atoms and 2845 basis functions.[123,122] Many flavors of higher-order wavefunction-based SAPT have also been developed. Methods such as SAPT2, SAPT2+3, and SAPT2+3(CCD) add in intramolecular correlation effects, higher-order inter- and intra-molecular perturbation series terms, and even certain infinite-order summations of particularly important classes of perturbation series terms.[138,257,224,326] SAPT2+3(CCD)/aug-cc-pVTZ provides an accuracy of a few tenths of one kcal mol^{-1} vs. CCSD(T)/CBS, and does not rely on error cancellation as in SAPT0/jun-cc-pVDZ.[214] These methods can be routinely deployed on systems with up to 30 atoms and 1100 basis functions, if density fitting and natural virtual orbital truncations are applied.[124,117,219] Another interesting class of SAPT methods

is the SAPT(CCSD) approach of Korona and coworkers,[161,157,158,159,160] which uses CCSD susceptibility functions to provide a very accurate description of the interaction energy, albeit for systems with only a handful of atoms. Finally, much effort has been applied in recent years to the development of SAPT(DFT) approaches.[123,118] These methods strongly resemble SAPT0, but replace the zeroth-order Hartree-Fock description of the monomers with a more-accurate KS-DFT description which automatically encapsulates intramonomer correlation effects..

One of the major strengths of SAPT is that it directly and rigorously computes the interaction energy. As a result, the physical components of the interaction energy fall directly out of low-order SAPT expansions,

$$E_{\mathrm{SAPT}} = E_{\mathrm{Elst}} + E_{\mathrm{Exch}} + E_{\mathrm{Ind}} + E_{\mathrm{Disp}} \tag{203}$$

The electrostatic energy is seen as the direct interaction between the unrelaxed and unsymmetrized monomers. The exchange energy arises by the application of the antisymmetry principle to the zeroth-order wavefunction, before any induction or dispersion has occurred, and may be viewed as a Pauli correction to the electrostatic energy. The induction energy arises as a response term wherein one monomer reacts to the electrostatic environment of the other monomer by rearranging its electrons, with an exchange correction to ensure that only Pauli-allowed polarizations occur. For polar and/or polarizable systems, accurate computation of the induction term requires that the monomer wavefunctions be allowed to mutually relax, i.e., that higher-order induction terms be included. In practical computations, infinite-order induction may be included at a Hartree-Fock level of theory by means of a hybrid supermolecular and perturbational approach known as a δ_{HF} correction. Finally, dispersion emerges as concerted excitations on the two monomers, corrected for the Pauli principle.

7.2 Partitioned Symmetry Adapted Perturbation Theory

At this point, a natural question is "where in the two monomers do these SAPT terms arise from?" To answer this, we assert the chemist's picture of localized contacts between the interacting monomers, and then partition SAPT so that a fraction of each SAPT term is assigned to each pairwise contact. In practice, this is accomplished via a two-phase procedure. In the first phase, the many-body perturbation series terms in SAPT are reduced to chemically sensible effective 2-body interactions between occupied particles in the two monomers. In the second phase, a localized set of occupied particles is selected to provide a robust partition to such moieties as atom- or functional-group-pairs.

Note that this approach is roughly analogous to an older approach: the "cut-and-cap" recipe. In cut-and-cap analysis, the monomers are divided into desired constituent functional groups by the chemist, following which linking covalent bonds between functional groups are cut and subsequently capped by hydrogen atoms to passivate the electronic and electrostatic structure of the resulting fragments. For each pairwise contact in the dimer, an interaction energy computation is performed for the underlying cut-and-capped fragments, in vacuum. Cut-and-cap analysis often provides useful insight into the origins of noncovalent interactions, but it does exhibit a number of problems. One is that the approach ignores the true many-body nature of the interaction energy. For instance, the exact limit of the total interaction energy is not obtained from the sum of pairwise interactions, even if *exact* interaction energies are computed for the underlying fragments. In particular, the effect of intramolecular polarization is not captured in the cut-and-cap analysis, often leading to substantial errors in the sum of isolated interaction energy computations. By contrast, the partitioned SAPT approach extracts an effective two-body interaction between each pair of fragments while respecting the many-body limit of the interaction energy.

7.2.1 Order-2 Partition of SAPT0 Terms

In general, the SAPT contributions to the interaction energy are many-body perturbation series terms. Each term may involve, two, a few, or even an infinite number of quasiparticle indices. Each index pertains to one or the other monomer, and may involve a nucleus or occupied orbital (the occupied space of the underlying monomer) or a virtual orbital (the vacuum around the monomer). The first task in partitioned SAPT approaches is to find an effective two-body interaction for each SAPT term which matches chemical intuition for what that term represents.

One particular question encountered in partitioned SAPT theory is the consideration of the physicality of the occupied vs. the virtual space. In the second-quantization treatment of the correlation energy,[109] the equations are invariant to occupied-virtual interchange, and thus both are referred to as "quasiparticles" and treated on the same footing. However, in partitioning SAPT, we are seeking to explain which portions of the generalized Lewis structures of the underlying monomers give rise to each interaction. For example, if we remove a chemical functional group from one of the monomers, we expect the major effect to be that direct interactions with that group disappear, followed by smaller effects involving the rearrangement of the electronic density in nearby groups. This consideration obviously implies that the occupied space is the relevant quantity. Thus, in all partitions below, we will focus on the occupied space indices, and sum immediately over the virtual space indices.

It is found that a chemically intuitive order-2 partition of the full SAPT0 interaction energy is easily obtained. Below, we demonstrate the general approach to this partition with three examples of increasing complexity. The list of indices used for these examples is presented in Table 4. A full partition of SAPT0, together with complete explanation and justification, is presented in Ref. 222, included further below.

7.2.1.1 Electrostatics

The electrostatic term encountered in SAPT0 is,

$$E^{(10)}_{\text{elst,r}} \equiv \iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \rho_{\mathbb{A}}(\vec{r}_1)\frac{1}{r_{12}}\rho_{\mathbb{B}}(\vec{r}_2) = 2V^B_{aa} + 2V^A_{bb} + 4v^{ab}_{ab} + \frac{Z_A Z_B}{r_{AB}}. \tag{204}$$

Here V^P_{ij} is a nuclear attraction integral of the form $(ij|P)$, and v^{jl}_{ik} is an ERI of the form $(ij|kl)$. This term involves only two indices: the protons and electrons $\mathbb{A}$ of monomer A and the protons and electrons $\mathbb{B}$ of monomer B. Thus the term is directly a two-body interaction energy, and the partition is, e.g.,

$$E^{(10),\mathbb{AB}}_{\text{elst,r}} \equiv \iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \rho_{\mathbb{A}}(\vec{r}_1)\frac{1}{r_{12}}\rho_{\mathbb{B}}(\vec{r}_2) = 2V^B_{aa} + 2V^A_{bb} + 4v^{ab}_{ab} + \frac{Z_A Z_B}{r_{AB}}. \tag{205}$$

7.2.1.2 Dispersion

The uncoupled two-body dispersion is used in SAPT0 is,

$$E^{(20)}_{\text{disp}} \equiv 4t^{rs}_{ab} v^{ab}_{rs}. \tag{206}$$

Here $t^{rs}_{ab} \equiv v^{rs}_{ab}/(\epsilon_a - \epsilon_b - \epsilon_r - \epsilon_s)$ is the uncoupled dispersion amplitude. This term involves four quasiparticles - the occupied electrons a and b and the virtual states r and s. This corresponds to the chemist's picture of dispersion as arising from two coupled excitations - an $a \rightarrow r$ excitation on monomer A with a $b \rightarrow s$ excitation on monomer B. As discussed above, we are concerned with where in the the two monomers this interaction originated from as the interaction operator was switched on. Thus, it is appropriate to retain indices for the occupied electrons and to sum over the virtual states, leading to the order-2 partition,

$$E^{(20),ab}_{\text{disp}} \equiv 4t^{rs}_{ab} v^{ab}_{rs}. \tag{207}$$

Another way to view this is the quantum interaction between the a and b electrons via fluctuations into the vacuum provided by the r and s virtual states.

7.2.1.3 Exchange-Dispersion

The exchange correction to the dispersion interaction used in SAPT0 is,

$$E^{(20)}_{\mathrm{exch-disp}}[S^2] \equiv -2t^{rs}_{ab}\tilde{v}^{ab}_{rs}. \tag{208}$$

The exchange-dispersion potential, is, in the S^2 approximation,

$$\begin{aligned}\tilde{v}^{rs}_{ab} \equiv &+1v^{sr}_{ab}+1S^s_a(2v^{ra'}_{ba'}-v^{a'r}_{ba'})+1S^s_{a'}(2v^{ra'}_{ab}-v^{a'r}_{ab})+1S^r_b(2v^{sb'}_{ab'}-v^{b's}_{ab'})+1S^r_{b'}(2v^{sb'}_{ba}-v^{b's}_{ba})\\
&+1S^r_bS^{b'}_{a'}v^{a's}_{ab'}-2S^r_{b'}S^{b'}_{a'}v^{a's}_{ab}-2S^r_bS^{b'}_a v^{a's}_{a'b'}+4S^r_{b'}S^{b'}_a v^{a's}_{a'b}+1S^s_aS^{a'}_{b'}v^{rb'}_{a'b}-2S^s_{a'}S^{a'}_{b'}v^{rb'}_{ab}-2S^s_aS^{a'}_b v^{rb'}_{a'b'}\\
&+4S^s_{a'}S^{a'}_b v^{rb'}_{ab'}+1S^r_{b'}S^s_{a'}v^{a'b'}_{ab}-2S^r_bS^s_{a'}v^{a'b'}_{ab'}-2S^r_{b'}S^s_a v^{a'b'}_{a'b}+1S^{a'}_bS^{b'}_a v^{rs}_{a'b'}-2S^{a'}_bS^{b'}_{a'}v^{rs}_{ab'}-2S^{a'}_{b'}S^{b'}_a v^{rs}_{a'b}\\
&+2S^r_{b'}S^{b'}_aV^A_{bs}+2S^s_{a'}S^{a'}_bV^B_{ar}-1S^r_bS^{b'}_aV^A_{b's}-1S^s_aS^{a'}_bV^B_{a'r}+1S^s_aS^r_{s'}V^A_{bs'}+1S^r_bS^s_{r'}V^B_{ar'}.\end{aligned} \tag{209}$$

Altogether, this term involves the four occupied particles a, a', b, and b', and could be expanded and summed to any pair of A and B contributions. However, physically, the term represents the Pauli corrections to the dispersion contributions to remove parts of the dispersion excitations which penetrate the occupied space of the opposing monomer. The two new particles a' and b' appear in order to encapsulate the collisions with the $b \rightarrow s$ and $a \rightarrow r$ excitations, respectively. Thus, the entire exchange-dispersion term may be viewed as a correction to the order-2 $a \leftrightarrow b$ interaction partitioned in the dispersion term above. This leads us to continue to focus on the occupied particles a and b, and to sum over the a' and b' particles in this term,

$$E^{(20),ab}_{\mathrm{exch-disp}}[S^2] \equiv -2t^{rs}_{ab}\tilde{v}^{ab}_{rs}. \tag{210}$$

7.2.2 Localization Techniques

Having obtained a generic order-2 partition of SAPT0 to contributions from the occupied spaces in the Hartree-Fock wavefunctions of the underlying monomers, the remaining task is to provide a sensible and robust set of localized quasiparticles for these occupied spaces. There are several possible schemes for this task, leading to a

number of possible variants of partitioned SAPT. Note that we refer to the discrete set of functions selected to span the occupied spaces as "quasiparticles" because they may be composed of a collection of more fundamental particles whose amalgamation is treated as a single unit for the purpose of the analysis. Depending on the partition scheme, a quasiparticle may be a proton or localized occupied electron, an atom, or an entire chemical functional group.

7.2.2.1 Protons and Electrons as Localized Quasiparticles

In electronic structure theory, the fundamental particles are electrons and protons. The protons are already completely localized at the nuclei, and thus there is no ambiguity in their contributions to the order-2 partition of SAPT. In the Hartree-Fock wavefunctions of the monomers, the electrons occupy a single Slater determinant which may be parametrized by an infinite number of sets of occupied orbitals. In particular, any unitary transformation of the canonical Hartree-Fock orbitals yields the same physical observables.

It has long been observed that there exists a broad class of unitary parametrizations of the occupied orbitals which resonates strongly with the concept of the Lewis structure - the concept of localized occupied orbitals. Many schemes for localized orbitals have been proposed in the literature, all based on the general idea that the maximally localized orbital set will directly mirror the Lewis structure of the molecule. Some of the more prominent schemes are reviewed below,

Boys: Approximately simultaneously diagonalizes the position operators $\hat{x}$, $\hat{y}$, and $\hat{z}$.[34] Boys localization is robust in large basis sets, scales as $\mathcal{O}(N^3)$, but mixes σ and π orbitals to produce hybrid "τ" shapes.

Pipek-Mezey: Approximately simultaneously diagonalizes the Mulliken atomic population operators $\hat{P}_A$.[228] Pipek-Mezey localization scales as $\mathcal{O}(N^3)$, preserves σ-π separation, but is unstable in large or diffuse basis sets due to the reliance

on Mulliken charges.

Emiston-Ruedenberg: Maximizes the self-repulsion of the localized electrons.[72] Edmiston-Ruedenberg localization is robust in large basis sets and preserves σ-π separation, but scales as $\mathcal{O}(N^5)$ due to reliance on the ERI tensor.

Cholesky: Uses pivoted Cholesky decomposition of the SCF one-particle density matrix to produce the local orbitals.[9] Cholesky localization is a direct $\mathcal{O}(N^3)$, but does not produce well-localized orbitals.

Intrinsic Bond Orbitals (IBO): A much-improved Pipek-Mezey localization scheme with the Mulliken populations operators replaced by intrinsic atomic orbital (IAO) population operators.[149] IBO scales as $\mathcal{O}(N^3)$, preserves σ-π separation, and exhibits robust performance in large basis sets.

We have previously used Pipek-Mezey and Boys localization schemes, but have recently come to wholly prefer the IBO scheme. IBO converges very rapidly, even in often-used diffuse basis sets such as jun-cc-pVDZ or aug-cc-pVTZ. Moreover, IBO orbitals are quite transferable. For instance, we find that a C-C σ orbital in benzene is essentially quantitatively identical to a C-C σ bond in a phenylalanine residue in a protein, *unless* a substitution or nearby dipole causes a polarization of the C-C σ bond. In the latter case, the polarization of the C-C σ bond is carried over into the IBO localized orbital, as expected. In Pipek-Mezey orbitals in diffuse basis sets, such polarizations are often carried over into a much larger number of localized orbitals, with positive and negative polarizations often contributing to the total and contaminating the analysis.

Accumulating the order-2 SAPT0 partition according to protons and localized electrons is quite easy computationally. The SAPT0 terms are written in terms of local occupied orbitals, and partial contractions over virtual indices and other occupied particles are carried out. This leads to what we have sometimes deemed

as "orbital-based SAPT." Orbital-based SAPT leads to an order-2 partition with nuclear-nuclear, electronic-nuclear, and electronic-electronic contributions. For electrostatics and induction sources, large and nearly cancelling contributions from the repulsive and attractive components make for a muddled analysis. This mandates the use of larger quasiparticles obtained by joining nuclear and electronic particles together to provide a sensible analysis.

7.2.2.2 Atoms as Localized Quasiparticles

Our first major approach to partitioned SAPT accumulated to atom-pairwise contributions (A-SAPT).[222] Here, an atom is defined as a nucleus, plus some affiliated electronic charge. The electronic charge fields are chosen to sum to the total Hartree-Fock density fields of the monomers, leading to non-spherical atoms with partial charges. For electrostatic terms (electrostatics and the source terms in induction), atom-pair contributions were directly obtained by replacing the total molecular density with the sum of atomic densities in the relevant places in the electrostatic integrals. To obtain a tractable and accurate method for the evaluation of the electrostatic energy, a grid-based resolution of the electrostatic integrals was invoked, following which an auxiliary basis was applied to regularize the $1/r_{12}$ Coulomb operator, an approach which strongly resembles PS-THC. For terms involving electronic response (exchange, induction response, and dispersion), contributions were first accumulated to the level of local occupied electrons. This order-2 partition was then reduced to the level of atom pairs via orbital atomic charges.

A key consideration in A-SAPT is the methodology used to partition the molecular electronic density into atomic contributions. A number of approaches to atomic partitioning exist,

Mulliken-Type: It is formally possible to use the Mulliken atomic population operators to produce atomic density fields, in a spectral-space approach. However,

this approach has severe basis set dependency and does not produce nearly spherical atoms.

IBO-Type: It is also possible to use the IAOs from the IBO localization methodology to construct atomic density fields, in a spectral-space approach. This removes the basis-set dependency above, and produces atoms with reasonable charges, but the charge distributions are not close to spherical.

Bader-Type: Atomic basins may be identified from topological consideration of the density field, in a grid-based approach (often referred to as atoms-in-molecules or AIM).[15] Unfortunately, these basins are not nearly spherical, and have rigid divisions, implying that very large quadrature expansions are required to produce accurate integrals over the basins.

Hirshfeld: Overlapping and nearly spherical atoms may be developed by weighting physical space according to the formula $w_{\bar{A}}(\vec{r}) \equiv \rho^0_{\bar{A}}(\vec{r}) / \sum_{\bar{A}'} \rho^0_{\bar{A}'}(\vec{r})$.[114] The spherical proatomic densities $\rho^0_{\bar{A}}(\vec{r})$ are taken from the neutral isolated atoms. Hirshfeld charges are found to be closer to neutral than expected, due to the reliance on neutral atoms for the proatomic densities.

Hirshfeld-I: A Hirshfeld approach with each proatomic density taken from a linear combination of the neutral, cationic, and anionic variants of the free atoms.[39] This is generally found to produce more sensible charges than Hirshfeld, at the cost of introducing an iterative fitting procedure for the proatomic density fields.

Iterative Stockholder Analysis (ISA): A Hirshfeld approach with unlimited freedom in the radial form of each proatomic density fields, which thus contains no empirical parameters (beyond the form of the fitting metric).[178]

For our purposes, ISA was selected due to its lack of reliance on empirical parameters. Hirshfeld-I is expected to give similar results.

A-SAPT was found to produce useful automatic decompositions of the interaction energy terms in large systems, and was able to answer some delicate questions for certain terms. For instance, A-SAPT was able to show that the previous hypothesis of the large induction energy in cation-π interactions being wholly due to the π orbitals is incorrect - more than half of the induction energy arises from C-C and C-H σ polarization. Moreover, A-SAPT is expected to prove quite useful in parametrizing the atom-pairwise intermolecular interaction terms in next-generation force fields. Rather than fitting the sum of all atom pairs to a single *ab initio* energy, it is now possible to extract an effective *ab initio* interaction energy for each atom pair and then fit individually to these pair potentials.

7.2.2.3 Functional Groups as Localized Quasiparticles

One area where A-SAPT fails is in quantifying the interactions between chemical functional groups. The root of this failure is the existence of partial charges in the concept of atoms in molecules. For instance, in considering the hydroxyl-hydroxyl interaction in phenol, the chemist is interested in how strong this interaction is relative to the corresponding interaction in water dimer, e.g., to the attractive dipole-dipole electrostatic energy. However, if one sums together the ISA O and H atoms in each phenol group, it is found that *both* hydroxyl groups carry a partial negative charge of roughly -0.2 electrons. This negative charge is due to the greater electronegativity of O vs. C, leading to the C-O σ bond being assigned more greatly to the O atom. This, in turn, leads to an unexpected repulsive charge-charge interaction between the hydroxyl groups, muddling the analysis. The point here is not that this interaction is unphysical - under the choice of partition, it is entirely correct. Rather, the point is that this contribution does not fit into the chemist's picture of interactions between

neutral functional groups with, at most, local dipoles. To provide a robust partition to the chemist's picture, we must modify our fragmentation scheme to ensure that the functional groups do not carry partial charges.

This is accomplished by the "functional group" SAPT partition (F-SAPT).[220] F-SAPT is based on the idea that the local occupied orbitals computed under a robust localization technique form the generalized Lewis structure of the molecule. Neutral functional groups can then be constructed from subsets of occupied electrons and protons. For example, in phenol, the hydroxyl group can be built from the O core orbital, the O-H σ bond, the two lone-pair orbitals, the H proton, and seven of the eight O protons. The phenyl group can be built from the C core orbitals, the C-C and C-H σ framework, the three π orbitals, the H protons, and all but one of the C protons. The hydroxyl and phenyl group are bridged by a linking σ unit, consisting of the C-O σ orbital, one of the O protons, and one of the C protons. The phenol monomer has now been divided into hydroxyl, phenyl, and link contributions which rigorously sum to the total system's density. In practice, the electrons and protons involved in each group can be automatically selected by orbital charge considerations, using a user-specified list of atoms specifying each functional group. To simplify the analysis, the link unit may be assigned 50/50 to the underlying functional groups, so long as the link unit is not significantly involved in the interaction.

The F-SAPT partition removes the charge-charge contamination of A-SAPT, but preserves the local dipoles of the functional groups. As such, it may be viewed as a cut-and-cap analysis which also respects the many-body limit of the interaction, particularly including the intramolecular polarizations of the various functional groups.

7.2.3 Partitioned Symmetry Adapted Perturbation Theory Publications

The remainder of this half of the thesis document consists of four papers detailing the development and deployment of partitioned SAPT methods. These works include,

1. The development of a general two-body partition of SAPT0 in concert with a specific localization scheme to the level of atoms (A-SAPT), adapted from Ref. 222.

2. The development of an alternative partition to chemical functional groups via considerations of local occupied orbitals (F-SAPT), adapted from Ref. 220.

3. The deployment of F-SAPT to investigate the origins of the substituent effect in sandwich benzene dimers, adapted from Ref. 221.

4. A side project using ideas from F-SAPT together with Hartree-Fock-in-Hartree-Fock embedding to yield a method capable of characterizing the magnitude and nature of intramolecular noncovalent contacts (I-SAPT), adapted from Ref. 215.

CHAPTER VIII

ATOMIC SYMMETRY-ADAPTED PERTURBATION THEORY

The following is adapted from Ref. 222.

8.1 Abstract

We develop a physically-motivated assignment of symmetry adapted perturbation theory for intermolecular interactions (SAPT) into atom-pairwise contributions (the A-SAPT partition). The basic precept of A-SAPT is that the many-body interaction energy components are computed normally under the formalism of SAPT, following which a spatially-localized two-body quasiparticle interaction is extracted from the many-body interaction terms. For electrostatics and induction source terms, the relevant quasiparticles are atoms, which are obtained in this work through the iterative stockholder analysis (ISA) procedure. For the exchange, induction response, and dispersion terms, the relevant quasiparticles are local occupied orbitals, which are obtained in this work through the Pipek-Mezey procedure. The local orbital atomic charges obtained from ISA additionally allow the terms involving local orbitals to be assigned in an atom-pairwise manner. Further summation over the atoms of one or the other monomer allows for a chemically intuitive visualization of the contribution of each atom and interaction component to the overall noncovalent interaction strength. Herein, we present the intuitive development and mathematical form for A-SAPT applied in the SAPT0 approximation (the A-SAPT0 partition). We also provide an efficient series of algorithms for the computation of the A-SAPT0 partition with essentially the same computational cost as the corresponding SAPT0 decomposition.

We probe the sensitivity of the A-SAPT0 partition to the ISA grid and convergence parameter, orbital localization metric, and induction coupling treatment, and recommend a set of practical choices which closes the definition of the A-SAPT0 partition. We demonstrate the utility and computational tractability of the A-SAPT0 partition in the context of side-on cation-π interactions and the intercalation of DNA by proflavine. A-SAPT0 clearly shows the key processes in these complicated noncovalent interactions, in systems with up to 220 atoms and 2845 basis functions.

8.2 Introduction

The quantitative determination and qualitative understanding of noncovalent interactions are critical challenges in modern chemistry. Despite their weak magnitudes relative to covalent bonds, noncovalent interactions are the crucial forces in a host of important chemical applications, such as the specificity and strength of drug-ligand interactions,[32,164,168,198,113,174,155,74] the packing of organic crystals,[230,247,231,29,127] the structure of nucleic acid complexes,[115,282,52,229,289,213,125] and many others. Unfortunately, the magnitude and nature of noncovalent forces are quite sensitive to both the electronic structure of the underlying monomers and the geometric arrangements of the constituent molecules into the supermolecular cluster. As such, intermolecular interactions are notoriously difficult to numerically quantify, even with high-accuracy *ab initio* methods. Moreover, the results of *ab initio* computations of intermolecular interactions are presently extremely difficult to analyze, as (with the exception of a few methods discussed below) no data is returned to indicate which pieces of the monomers are most responsible for a given interaction energy contribution. In this work, we present a method to directly attribute many-body interaction energy components from symmetry adapted perturbation theory to localized regions of physical space, e.g., to provide an atomic partition of the many-body interaction energy components.

Noncovalent interactions are intrinsically many-body in nature: as Bohr once told Heisenberg,[108] "We must be clear that when it comes to atoms, language can be used only as in poetry. The poet, too, is not nearly so concerned with describing facts as with creating images and establishing mental connections." This is borne out in the current state of affairs in the field. Convergent *ab initio* methods such as supermolecular methods or symmetry-adapted perturbation theory, which can be expanded to the Schrödinger limit of the interaction energy, respect the accurate many-body nature of the interaction, but cannot currently partition the interaction components into spatial progenitors such as atoms. Alternatively, non-convergent methods such as force fields or pairwise *ab initio* computations on fragments can provide a spatial partition of the interaction components to the level of atoms or fragments, but are definitionally pairwise in nature, and therefore cannot reproduce the many-body effects inherent in the Schrödinger limit of the interaction energy. In this work, we pursue an alternative paradigm, in which we allow the many-body interaction to occur naturally, and then extract the key two-body interactions from within the dressing of the many-body effects. Partial traces over all bodies except for a local quasiparticle basis for the critical two-body interaction provides a quasi-atom-pairwise picture of the interaction components, while maintaining the interaction energy magnitude of the full many-body interaction. We stress that this partition is neither an observable quantity, nor is it unique, but it is well-defined and carefully chosen to mirror and amplify chemical intuition while preserving the many-body limit in the interaction energy components.

The starting point for our approach is the methodology of symmetry adapted perturbation theory (SAPT).[139,291,138] SAPT computes the intermolecular interaction as a perturbation to noninteracting monomer wavefunctions, with the added adaption of total wavefunction antisymmetry. One powerful feature of SAPT is that it provides a decomposition of the intermolecular interaction energy in terms

of physically intuitive electrostatic, exchange, induction, and dispersion components, though there do exist other methods which can provide similar decompositions.[147,16,13] SAPT is formally completable to the Schrödinger limit of the interaction energy, if the intermolecular interaction potential and intramolecular correlation potentials are expanded to infinite-order in perturbation theory. In practice, a tractable truncation to the perturbation series must be made, resulting in the slew of SAPT flavors in the literature. If wavefunction theory is used for the SAPT expansion, many choices of truncation are available, including complete neglect of the intramolecular correlation potentials [SAPT0],[138] many-body treatment of the intramolecular correlation potentials and/or higher-order terms [SAPT2, SAPT2+, SAPT2+(3), SAPT2+3],[138,257,224,124] infinite-order treatment of certain terms [SAPT2+(CCD), SAPT2+(3)(CCD), SAPT2+3(CCD)],[326,117,219] or expansion of the monomer susceptibility functions [SAPT(CCSD), etc].[161,157,158,159,160] Additionally, much progress has been made recently in adapting Kohn-Sham density functional theory (DFT) to the SAPT environment, resulting in the closely related SAPT(DFT) and DFT-SAPT methods with excellent accuracy to cost ratios.[325,110,195,111,233] Most variants of SAPT are designed for dimer interactions, though trimer[202,182,232] and even hybrid many-monomer[135,169] flavors also exist. We focus exclusively on dimer SAPT methodology in the present work, and additionally select SAPT0 as a convenient starting point. While SAPT0 is generally not of quantitative accuracy, we have found that semiquantitative accuracy can be obtained when the method is deployed in a carefully chosen jun-cc-pVDZ basis set (cc-pVDZ on hydrogen, aug-cc-pVDZ less the diffuse functions of highest angular momentum on heavy elements).[124,212] By applying density fitting[324,67,68,299,76,245,144,312] and Laplace transformation[4,104,37,293] numerical approximations to the implementation of SAPT0, we have developed a code that can run SAPT0/jun-cc-pVDZ on systems with > 200 atoms and > 2800 basis functions,[123,118] implying that a visualization technique developed for SAPT0 might be immediately

applicable to large-scale intermolecular interactions. Moreover, the topologies of the SAPT0 and SAPT(DFT) methods are essentially identical: the methodology developed in this work is immediately and directly transferable to SAPT(DFT) simply by substituting coupled Kohn-Sham quasiparticles in place of Hartree-Fock quasiparticles.

This work focuses on a spatial partition of the SAPT0 decomposition, yielding a method we refer to as atomic SAPT0 (A-SAPT0). A-SAPT0 is one specialization of a family of possible atomic SAPT methods (A-SAPT); we envision that further studies will yield methods such as A-SAPT(DFT) and A-SAPT2+3(CCD). In this manuscript, we provide the definition, justification, implementation, and initial testing of the A-SAPT0 partition. The key focus of the paper is the identification of critical two-body interactions in the SAPT0 interaction energy components and subsequent local quasiparticle accumulation of these two-body interactions. For the two-body partition, we use an extension of the diagrammatic notation developed for SAPT by Rybek *et al.* to illustrate the nature of each interaction energy component.[257] For the local quasiparticle basis, we use atomic density fields obtained from the iterative stockholder analysis (ISA) procedure of Lillestolen and Wheatley[178,179,315] for electrostatic source terms, and local orbitals obtained from the Pipek-Mezey[228] procedure for the electronic response terms. We also discuss key technical developments required for the correct and efficient implementation of the A-SAPT0 methodology. In particular, an intermediate projection from grid-based physical-space atomic density fields to auxiliary-basis spectral-space atomic electrostatic potentials is required to regularize the Coulomb singularity in the electrostatic source terms: this step borrows heavily from our experience with the tensor hypercontraction (THC) resolution of the electron repulsion integrals.[121,216,218] Additionally, concerted application of density fitting and canonical-to-local orbital transformation yields an implementation of A-SAPT0 with similar computational cost as the underlying SAPT0 method, with

minimal formal modification. We then discuss two visualization schemes developed to quickly analyze the results of an A-SAPT0 computation. In the results section, we test the sensitivity of the A-SAPT0 partition to the grid and convergence scheme used in the ISA procedure, the choice of orbital localization metric, and the choice of coupled vs. uncoupled induction treatment. Finally, we deploy A-SAPT0 to investigate side-on cation-π interactions and the intercalation of DNA by proflavine, demonstrating the compelling utility of the A-SAPT0 method in systems with more than 200 atoms.

At this point, a brief review of existing techniques for visualizing noncovalent interactions is appropriate. Noncovalent interactions are much less intuitive than covalent bond arrangements, which can be conceptualized by Lewis structures[173] or quantitatively visualized via local orbitals,[34,72,228,9,126,149,172] electron localization functions (ELF or LOL),[21,275] atoms in molecules (AIM),[14,15] or natural bond orbital (NBO) analyses.[78] For intermolecular interactions, the sometimes-used "..." notation for hydrogen bonds is the only conceptual analog of Lewis structures, and quantitative visualization techniques are less-well developed. Of the four intermolecular forces, electrostatics is the most intuitive, and therefore, the most developed. Plots of the electrostatic potentials of the two monomers are often use to elucidate favorable arrangements of the dimer. Additional techniques exist for the identification and characterization of hydrogen bonding, including ELF analysis[3,51] and a very interesting recent utilization of experimental atomic force microscopy (AFM) techniques to directly measure the steric repulsion from the hybridized electronic density along the hydrogen bond with the AFM tip.[330] For general noncovalent interactions, the recent visualization work by Johnson *et al.* uses real-space consideration of the electronic density and its derivatives to illuminate regions where noncovalent interactions of certain topologies may be occurring.[141,50] This method is agnostic as to intramolecular vs. intermolecular noncovalent interactions, and is applicable to large

systems. However, the method is a heuristic based on correlations between a visualization of the "non-covalent interactions (NCI)" scalar field and chemical intuition. Also relevant is the semiempirical-based divide and conquer interaction energy decomposition of van der Vaart and Merz,[302] which can provide a subsystem-pairwise accounting of electrostatics/exchange, polarization, and charge transfer. This method has been used to study the genesis of large scale noncovalent interactions, such as the biomolecule/solvent interface,[206,303] protein folding,[301] protein-ligand interactions,[241] and protein-protein association.[1] However, this methodology does not allow for the completion of the Hamiltonian elements linking the subsystems, e.g., there is no accounting for the dispersion interaction. Additionally, methods analogous to A-SAPT, which provide visualizations based on energetic contributions assigned to atoms or orbitals, should also be possible with molecular mechanics (MM),[53] energy decomposition analysis (EDA),[147] or absolutely localized molecular orbital (ALMO) techniques,[13] though we are not currently aware of any examples of such visualizations. Moreover, any such visualization techniques would not be formally completable to the full Schrödinger limit, as might be possible with A-SAPT.

8.3 Theory

8.3.1 Notation

8.3.1.1 Mathematical Notation

A large number of classes of indices will be used in this work, so we provide a brief enumeration of all index types in Table 4. Primes are used to distinguish repeated indices. A generalized Einstein summation convention is used throughout: whenever an index on the right-hand side of an expression disappears, that index is asserted to be summed over. All expressions presented in this work pertain to SAPT between two closed-shell reference monomers.

Table 4: Summary of indices occurring in A-SAPT0, for monomer A and monomer B in the first two columns, respectively.

A	B	Description
μ	μ	Primary orbital basis function (dimer-centered)
C	C	-JKFIT auxiliary basis function (dimer-centered)
D	D	-RI auxiliary basis function (dimer-centered)
P	P	Becke quadrature grid points (dimer-centered)
i	i	Becke radial node, attached to some non-ghost atom A (monomer-centered)
$\mathbb{A}$	$\mathbb{B}$	All particles of monomers, including nuclei and electrons
$\mathcal{A}$	$\mathcal{B}$	All electrons of monomers
$\mathbf{A}$	$\mathbf{B}$	Atoms of monomers, including a nucleus and assigned electronic density
A	B	Nuclei of monomers
$\bar{A}$	$\bar{B}$	Electronic density assigned to atoms of monomers
a	b	Canonical occupied orbitals of monomers
r	s	Canonical virtual orbitals of monomers
$\bar{a}$	$\bar{b}$	Local occupied orbitals of monomers

8.3.1.2 Diagrammatic Notation

To support the physical understanding of the SAPT0 interaction energy components and our order-2 partition thereof, we will use an extension of the SAPT Brandow-type interaction energy diagrams developed in the work of Rybek, Jeziorski, and Szalewicz,[257] with some additional elements to denote the renormalized interaction potentials and direct overlap interactions accompanying exchange terms. These diagrams appear in Figure 25, along with a short summary of the necessary diagram elements. Each diagram will be explained when encountered below. Note that, in contrast to the standard use of diagrams in coupled-cluster theory, we are using these diagrams for pedagogical and illustrative purposes rather than for rigorous derivations.

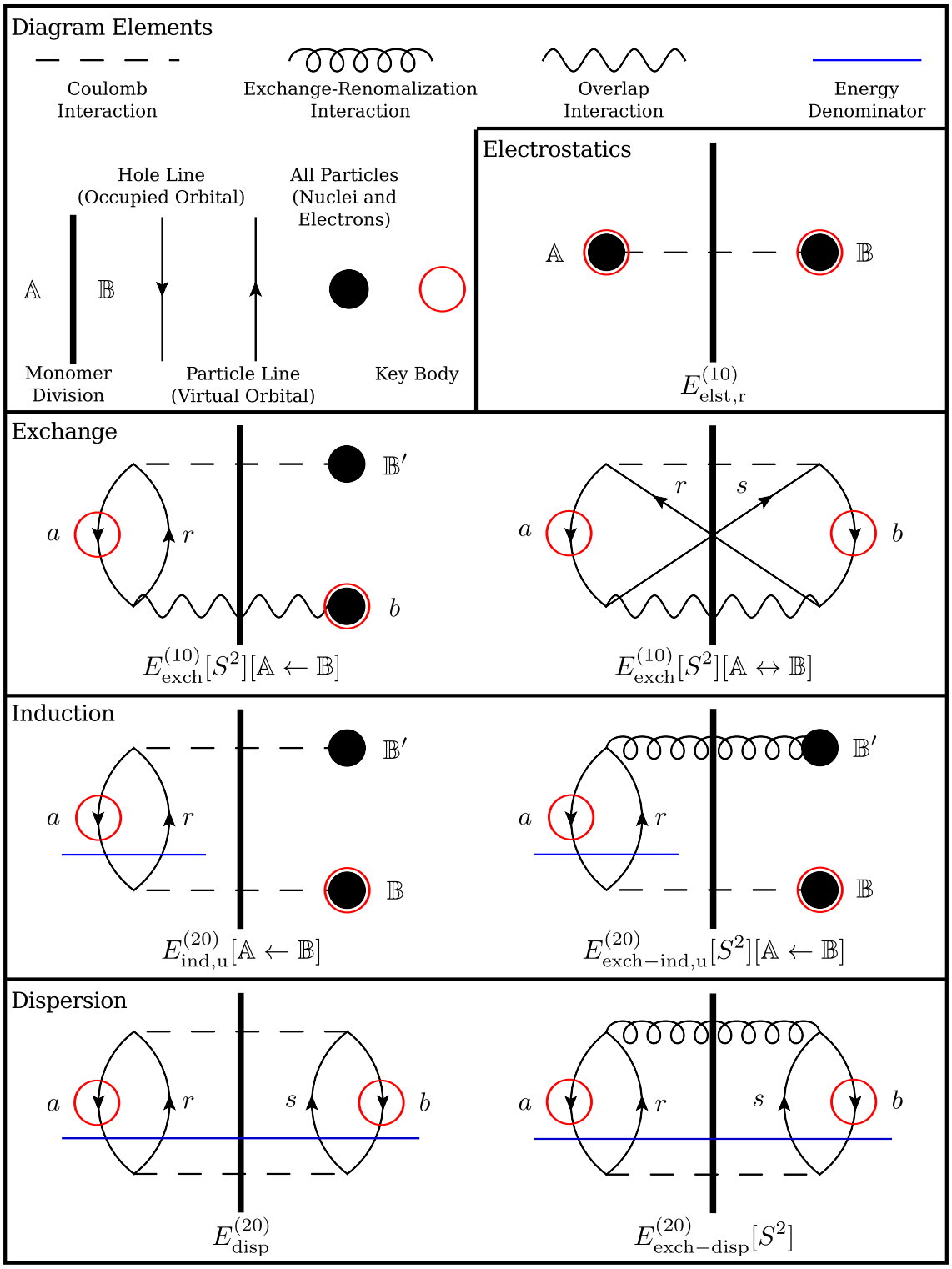

Figure 25: Relevant diagrams and key body selections for the A-SAPT0 order-2 partition.

8.3.2 SAPT0 Order-2 Partition

8.3.2.1 SAPT0 Overview

The SAPT0 interaction energy is most commonly defined as

$$E_{\mathrm{SAPT0}} \equiv E^{(10)}_{\mathrm{elst,r}} + E^{(10)}_{\mathrm{exch}}[S^\infty] + E^{(20)}_{\mathrm{ind,r}} + E^{(20)}_{\mathrm{exch,ind,r}}[S^2] + \delta^{(2)}_{\mathrm{HF,r}} + E^{(20)}_{\mathrm{disp}} + E^{(20)}_{\mathrm{exch-disp}}[S^2]. \quad (211)$$

These terms can be grouped to provide a chemically intuitive breakdown of the interaction energy in terms of electrostatic, exchange, induction, and dispersion components. Using the terms selected above, the decomposition may be chosen as,

$$E_{\mathrm{elst}} \equiv E^{(10)}_{\mathrm{elst,r}} \quad (212)$$

$$E_{\mathrm{exch}} \equiv E^{(10)}_{\mathrm{exch}}[S^\infty] \quad (213)$$

$$E_{\mathrm{ind}} \equiv E^{(20)}_{\mathrm{ind,r}} + E^{(20)}_{\mathrm{exch-ind,r}}[S^2] + \delta^{(2)}_{\mathrm{HF,r}} \quad (214)$$

$$E_{\mathrm{disp}} \equiv E^{(20)}_{\mathrm{disp}} + E^{(20)}_{\mathrm{exch-disp}}[S^2]. \quad (215)$$

A number of choices were made above, resulting in a balanced but economical decomposition. The first is that the intramonomer correlation is entirely neglected (hence the name SAPT0) and the SAPT intermolecular perturbation series is only explicitly computed to second order. Second, coupled response, denoted by the subscript r, is included in the induction terms to provide for electronic relaxation in these terms. This treatment improves the induction energy in nonpolar systems. However, in polar and/or polarizable systems, higher-order induced-multipole induced-multipole induction terms still remain important, and are compensated for by the hybrid $\delta^{(2)}_{\mathrm{HF,r}}$ correction, computed as

$$\delta^{(2)}_{\mathrm{HF,r}} \equiv E_{\mathrm{HF}} - E^{(10)}_{\mathrm{elst,r}} - E^{(10)}_{\mathrm{exch}}[S^\infty] - E^{(20)}_{\mathrm{ind,r}} - E^{(20)}_{\mathrm{exch-ind,r}}[S^2]. \quad (216)$$

This treatment pegs the sum of electrostatics, exchange, and induction to the Hartree-Fock interaction energy. Finally, $E^{(10)}_{\mathrm{exch}}$ is computed with full S^∞ treatment of the antisymmetrization operator, while the induction and dispersion counterparts are only

computed in the S^2 approximation. The former can provide substantial improvements of the steric repulsion on the exponential wall of potential energy surfaces ($E^{(10)}_{\rm exch}[S^2]$ is often observed to be roughly 5-10% too small at short intermolecular separation). The error due to the S^2 treatment of $E^{(20)}_{\rm exch-ind,r}$ is masked if the $\delta^{(2)}_{\rm HF,r}$ correction is used, while the error due to the S^2 treatment of $E^{(20)}_{\rm exch-disp}$ is usually not problematic due to the small magnitude of this term.

8.3.2.2 Electrostatics

The electrostatic term is simply

$$E^{(10)}_{\rm elst,r} \equiv \iint_{\mathbb{R}^6} {\rm d}r_1 \; {\rm d}r_2 \; \rho_{\mathbb{A}}(\vec{r}_1) \frac{1}{r_{12}} \rho_{\mathbb{B}}(\vec{r}_2) = 2V^B_{aa} + 2V^A_{bb} + 4v^{ab}_{ab} + \frac{Z_A Z_B}{r_{AB}}. \qquad (217)$$

Here $\rho_{\mathbb{A}}(\vec{r}_1)$ is the total charge density (including nuclei) of monomer A, V^B_{aa} is the electrostatic potential matrix element between the orbital pair aa and atom B [e.g., $V^B_{pp'} \equiv -(p|Z_B/r_{1B}|p')$], v^{ab}_{ab} is the standard electron repulsion integral [e.g, $v^{p'p''}_{pp'''} \equiv (pp'|p''p''')$], and Z_A is the charge of nucleus A. It is apparent that the four terms in the last expression all carry distinct physical meaning. In order, they represent the attraction of the electrons of monomer A to the nuclei of monomer B, the attraction of the electrons of monomer B to the nuclei of monomer A, the electron-electron repulsion between monomers, and the nuclear-nuclear repulsion between monomers. Each contribution is obviously two-particle in nature. It is worth noting that each of these contributions is typically $> 1000\times$ larger than the total, as the first two terms nearly cancel with the last two. This feature implies that any geometric partition must be extremely careful to provide a balanced treatment of nuclear and electronic contributions, to avoid large oscillations in the partition.

The SAPT diagram corresponding to the electrostatic term is found in the upper right corner of Figure 25. This term involves the direct interactions of the total charge fields on monomer A and monomer B. However, any reasonable visualization scheme for this term must group some electronic density together with each nucleus to avoid

visualizing extremely large, nearly cancelling quantities. The concept required here is that of an atom **A**: a nucleus A with charge density $\rho_A(\vec{r}) \equiv Z_A \delta(\vec{r} - \vec{r}_A)$ (positive), and some localized fraction of the total electronic density $\bar{A}$ defined as the atomic electronic density field $\rho_{\bar{A}}(\vec{r})$ (negative). We leave the assignment and manipulation of these atomic density fields to a later section, but for now simply assert that the total molecular density *can* be decomposed into such atom-local contributions,

$$\rho_{\mathbb{A}}(\vec{r}) = \sum_{\mathbf{A}} \rho_{\mathbf{A}}(\vec{r}) = \sum_{\mathbf{A}} \rho_{\bar{A}}(\vec{r}) + \rho_A(\vec{r}). \tag{218}$$

If this is the case, then the electrostatic term can be decomposed into an order-2 atomic partition,

$$E^{(10)}_{\text{elst,r}} \equiv \sum_{\mathbf{AB}} E^{(10),\mathbf{AB}}_{\text{elst,r}} \equiv \iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \rho_{\mathbf{A}}(\vec{r}_1) \frac{1}{r_{12}} \rho_{\mathbf{B}}(\vec{r}_2). \tag{219}$$

This choice of decomposition is denoted by the circled key bodies in the SAPT electrostatics diagram. In practice, each element of this order-2 partition contains four terms, corresponding to the matrix-element expression in the final equality of Equation 217,

$$\begin{aligned} E^{(10),\mathbf{AB}}_{\text{elst,r}} \equiv & \\ & \int_{\mathbb{R}^3} \mathrm{d}r_1 \; \rho_{\bar{A}}(\vec{r}_1) \frac{1}{r_{1B}} Z_B + \int_{\mathbb{R}^3} \mathrm{d}r_2 \; Z_A \frac{1}{r_{2A}} \rho_{\bar{B}}(\vec{r}_2) \\ & + \iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \rho_{\bar{A}}(\vec{r}_1) \frac{1}{r_{12}} \rho_{\bar{B}}(\vec{r}_2) + \frac{Z_A Z_B}{r_{AB}}. \end{aligned} \tag{220}$$

As with the total electrostatic term, the first two terms of the atomic partition nearly cancel the last two terms, mandating extremely accurate computation of each term.

8.3.2.3 Exchange

In the S^2 approximation, exchange is a one- and two-particle electronic response,[201]

$$E^{(10)}_{\text{exch}}[S^2] \equiv \underbrace{-2 S^b_a S^r_b V^{\mathbb{B}'}_{ar}}_{E^{(10)}_{\text{exch}}[S^2][\mathbb{A} \leftarrow \mathbb{B}]} \underbrace{-2 S^a_b S^s_a V^{\mathbb{A}'}_{bs}}_{E^{(10)}_{\text{exch}}[S^2][\mathbb{B} \leftarrow \mathbb{A}]} \underbrace{-2 S^s_a S^r_b v^{ab}_{rs}}_{E^{(10)}_{\text{exch}}[S^2][\mathbb{A} \leftrightarrow \mathbb{B}]}. \tag{221}$$

Here S_a^b is an overlap integral, i.e., $S_a^b = (ab)$, and $V_{ar}^{\mathbb{B}}$ is the electrostatic potential of monomer B, i.e., $V_{ar}^{\mathbb{B}} \equiv V_{ar}^{B} + 2v_{ab}^{rb}$. The SAPT diagrams corresponding to the first and third of these terms are presented on the upper middle row of Figure 25; the second term is obtained by reflecting the first diagram across the monomer division. As discussed by Moszynski and *et al.*,[201] the topology of this term is remarkably similar to that of the uncoupled, non-symmetrized second-order Rayleigh-Schrödinger Polarization Theory (RSPT) term,

$$E_{\text{pol}}^{(20)} \equiv \underbrace{2\frac{V_{ar}^{\mathbb{B}}}{\epsilon_a - \epsilon_r}V_{ar}^{\mathbb{B}'}}_{E_{\text{ind,u}}^{(20)}[\mathbb{A}\leftarrow\mathbb{B}]} + \underbrace{2\frac{V_{bs}^{\mathbb{A}}}{\epsilon_b - \epsilon_s}V_{bs}^{\mathbb{A}'}}_{E_{\text{ind,u}}^{(20)}[\mathbb{B}\leftarrow\mathbb{A}]} + \underbrace{4\frac{v_{ab}^{rs}}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}v_{rs}^{ab}}_{E_{\text{disp}}^{(20)}}, \tag{222}$$

where ϵ is an orbital eigenvalue. In both equations, the first contributions are single-particle excitations (generalized inductions) of monomer A in response to some field of monomer B, all under the electrostatic field of monomer B. For exchange, the generalized induction amplitude $y_a^r \equiv -S_a^b S_b^r$ is induced by the Pauli-excluded overlap with the orbitals b of monomer B. The potential driving this amplitude is the new overlap interaction line in the left-side exchange diagram. For polarization, the true induction amplitude $x_a^r \equiv V_{ar}^{\mathbb{B}}/(\epsilon_a - \epsilon_r)$ is induced by the electrostatic field of the nuclei and electrons of monomer B. Technically these contributions are three-body in nature: the body invoking the induction, the body performing the induction, and the body providing the field. For the purposes of partitioning, we will focus on the first two bodies, in the understanding that there is another level of detail being abstracted here. The second contribution in each equation is identical to the first, but involves monomer A inducing generalized inductions on monomer B, in the field of A. The third contributions are concerted two-particle excitations. In the case of exchange, the generalized doubles amplitude $u_{ab}^{rs} \equiv -S_a^s S_b^r$ encapsulates the exchange correlation between same-spin electrons. In the case of polarization, the usual doubles amplitude $t_{ab}^{rs} \equiv v_{ab}^{rs}/(\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s)$ encapsulates the Coulomb correlation between both same-spin and opposite-spin electrons (hence the extra factor of two).

With these observations, the order-2 partition of the exchange term in the S^2 approximation is obvious, and just involves delaying summation over the occupied spaces in Equation 221,

$$E_{\rm exch}^{(10)}[S^2] \equiv \sum_{ab} E_{\rm exch}^{(10),ab}[S^2] \equiv \sum_{ab} -2S_a^b S_b^r V_{ar}^{\mathbb{B}'} - 2S_b^a S_a^s V_{bs}^{\mathbb{A}'} - 2S_a^s S_b^r v_{rs}^{ab}. \tag{223}$$

This choice is denoted by the circled key bodies in the SAPT diagrams. In practice, the occupied space must be taken to be composed of localized quasiparticles for a localized partition: this will be accomplished by moving to a basis of local occupied orbitals in a later section.

Under the S^∞ treatment, exchange involves all possible tuples of permutations between particles, and is computed as

$$\begin{aligned} E_{\rm exch}^{(10)}[S^\infty] \equiv & -2v_{ab}^{ba} + 2T_a^{a'}(V_{aa'}^B + 2v_{ab}^{a'b} - v_{ab}^{ba'}) + 2T_b^{b'}(V_{bb'}^A + 2v_{ba}^{b'a} - v_{ba}^{ab'}) \\ & + 2T_a^b(V_{ab}^A + 2v_{aa'}^{ba'} - v_{aa'}^{a'b}) + 2T_a^b(V_{ab}^B + 2v_{ab'}^{bb'} - v_{ab'}^{b'b}) + 2T_a^b T_{b'}^{b''}(2v_{ab'}^{bb''} - v_{ab'}^{b''b}) \\ & + 2T_a^b T_{a'}^{a''}(2v_{ba'}^{aa''} - v_{ba'}^{a''a}) + 2T_a^{a'} T_b^{b'}(2v_{ab}^{a'b'} - v_{ab}^{b'a'}) + 2T_a^b T_{a'}^{b'}(2v_{ab'}^{ba'} - v_{ab'}^{a'b'}). \end{aligned} \tag{224}$$

Here the T blocks are made of the relevant elements of the Hermitian matrix

$$T \equiv \begin{bmatrix} T_a^{a'} & T_a^{b'} \\ T_b^{a'} & T_b^{b'} \end{bmatrix} = \begin{bmatrix} \delta_{aa'} & S_a^{b'} \\ S_b^{a'} & \delta_{bb'} \end{bmatrix}^{-1} - \begin{bmatrix} \delta_{aa'} & 0 \\ 0 & \delta_{bb'} \end{bmatrix}. \tag{225}$$

Despite the large number of terms, this contribution can be obtained via a handful of generalized Fock matrices. Unfortunately, a order-2 partition of this term is not apparent to us; this does not imply that such a partition does not exist, merely that we have been unable to locate such a partition, as yet. In practice, the difference between $E_{\rm exch}^{(10)}[S^2]$ and $E_{\rm exch}^{(10)}[S^\infty]$ is at most a few percent, in the neighborhood of the steric wall. As a convenient compromise, we will scale the order-2 exchange partition obtained under the S^2 approximation to match the S^∞ result, under the assumption of isotropic magnification at all points in physical space,

$$E_{\rm exch}^{(10),ab}[S^\infty] \approx \frac{E_{\rm exch}^{(10)}[S^\infty]}{E_{\rm exch}^{(10)}[S^2]} E_{\rm exch}^{(10),ab}[S^2]. \tag{226}$$

This scaling technique is used for all results presented in this work.

8.3.2.4 Induction

As introduced above, the second-order induction interaction is a set of single-particle excitations of monomer A in response to the electrostatic potential of monomer B, under the electrostatic field of monomer B. The excitations can either be uncoupled, in which case each electron moves independently from all the others, or coupled, in which case a concerted induction occurs to lower the energy as much as possible. In either case, the RSPT treatment above will overestimate the induction energy, as the single particle excitations of monomer A generally penetrate much too far into the Pauli-excluded space of monomer B. This is compensated by the exchange-induction correction, which can be viewed as providing a modification to the electrostatic field which damps out the induction energy lowering in cases where the single excitation is Pauli forbidden. Taken together, the induction and exchange-induction contributions can be reasonably viewed as a modified induction within the Pauli-allowed space. The diagrams depicting these two terms are shown in the lower middle panel of Figure 25.

Mathematically, the induction of monomer A in response to the electrostatic potential of monomer B is given as

$$E_{\text{ind}}^{(20)}[\mathbb{A} \leftarrow \mathbb{B}] \equiv 2x_{ar}^{\mathbb{B}} V_{ar}^{\mathbb{B}'}. \tag{227}$$

Here $[x^{\mathbb{B}}]_a^r$ is the matrix of singles amplitudes of monomer A in response to the electrostatic potential of monomer B.

In the uncoupled formalism, the singles amplitudes are simply given as

$$x_{ar}^{\mathbb{B}} \equiv \frac{V_{ar}^{\mathbb{B}}}{\epsilon_a - \epsilon_r}. \tag{228}$$

In the coupled formalism, the singles amplitudes are the solutions of the Coupled-Perturbed Hartree-Fock (CPHF) equations for monomer A,

$$\left[(\epsilon_r - \epsilon_a)\delta_{aa'}\delta_{rr'} + 4v_{aa'}^{rr'} - v_{aa'}^{r'r} - v_{ar}^{a'r'}\right] x_{a'r'}^{\mathbb{B}} = -V_{ar}^{\mathbb{B}}. \tag{229}$$

The uncoupled amplitudes arise from neglecting all ERI contributions to the CPHF equations.

After some manipulation from the standard forms encountered in the literature, the exchange-induction term in the S^2 approximation can be evaluated as a contraction with identical topology as the induction term, albeit with a modified "exchange-induction potential" $\tilde{V}_{ar}^{\mathbb{B}'}$,

$$E_{\mathrm{exch-ind}}^{(20)}[S^2][\mathbb{A} \leftarrow \mathbb{B}] \equiv 2x_{ar}^{\mathbb{B}} \tilde{V}_{ar}^{\mathbb{B}'}, \tag{230}$$

where,

$$\begin{aligned}\tilde{V}_{ar}^{\mathbb{B}} \equiv & -1v_{ab}^{br} - 1S_a^b V_{br}^A - 2S_a^b v_{a'b}^{a'r} + 1S_a^b v_{a'b}^{ra'} + 1S_{a'}^b v_{ab}^{a'r} - 2S_{a'}^b v_{ab}^{ra'} - 1S_b^r V_{ab}^B \\ & - 2S_b^r v_{ab'}^{bb'} + 1S_b^r v_{ab'}^{b'b} + 1S_a^b S_{b'}^r V_{bb'}^A + 2S_a^b S_{b'}^r v_{a'b}^{a'b'} + 1S_{a'}^{b'} S_{b'}^r V_{aa'}^B + 2S_{a'}^{b'} S_{b'}^r v_{ab}^{a'b} \\ & - 1S_{a'}^b S_{b'}^r v_{ab}^{a'b'} + 2S_{a'}^b S_{b'}^{a'} v_{ab}^{rb'} + 1S_a^{b'} S_{b'}^{a'} V_{a'r}^B + 2S_a^{b'} S_{b'}^{a'} v_{a'b}^{rb} - 1S_a^b S_{b'}^{a'} v_{a'b}^{rb'}. \end{aligned} \tag{231}$$

The computational effort of forming this modified potential is negligible compared to the solution of the CPHF equations. In fact, this potential may be easily formed by a number of generalized Fock matrix builds. See, for instance, the work of Heßelmann and *et al.*[111] or Hapka *et al.*,[101] and "unwind" their direct contractions to the exchange-induction energy. This potential is the new "exchange-renormalization" interaction line appearing in the exchange-induction SAPT diagram.

In either the coupled or uncoupled formalisms, the contractions to the induction and exchange-induction terms are identical, under the choice of singles amplitudes. The net induction for the monomer is given as

$$E_{\mathrm{ind}}[\mathbb{A} \leftarrow \mathbb{B}] \equiv E_{\mathrm{ind}}^{(20)}[\mathbb{A} \leftarrow \mathbb{B}] + E_{\mathrm{exch-ind}}^{(20)}[S^2][\mathbb{A} \leftarrow \mathbb{B}]. \tag{232}$$

The induction and exchange-induction counterparts for monomer B may be obtained by permutation of indices in the above expressions.

As with the single-particle response in exchange, we will focus on the particle that

causes the induction and the particle that undergoes the polarization in the order-2 partition, as indicated by the circled key bodies in Figure 25. This means that the particles providing the field will be summed over: the assumption here is that the particle providing the field is the same as that which causes the polarization. A critical feature of the linear response formalism comes into play here: the response of monomer A's density field to the total electrostatic potential of monomer B is equal to the sum of the responses to any series of trial electrostatic potentials which add up to the total electrostatic potential. Therefore, we may borrow techniques from the section on electrostatics, and solve for the response of monomer A to the electrostatic potential of each of monomer B's atoms. Defining the atomic electrostatic potential from the atomic density,

$$V_{ar}^{\mathbf{B}} \equiv -\iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \phi_a(\vec{r}_1)\phi_r(\vec{r}_1)\frac{1}{r_{12}}\rho_{\mathbf{B}}(\vec{r}_2), \tag{233}$$

we see that the total electrostatic potential is maintained,

$$V_{ar}^{\mathbb{B}} = \sum_{\mathbf{B}} V_{ar}^{\mathbf{B}}. \tag{234}$$

The induction amplitudes providing the response due to each atom may be most easily obtained at the uncoupled level, via

$$x_{ar}^{\mathbf{B}} \equiv \frac{V_{ar}^{\mathbf{B}}}{\epsilon_a - \epsilon_r}. \tag{235}$$

Alternatively, these amplitudes may be obtained at the coupled level by solving the CPHF equations for monomer A for each of the **B** atomic perturbations,

$$\left[(\epsilon_r - \epsilon_a)\delta_{aa'}\delta_{rr'} + 4v_{aa'}^{rr'} - v_{aa'}^{r'r} - v_{ar}^{a'r'}\right] x_{a'r'}^{\mathbf{B}} = -V_{ar}^{\mathbf{B}}. \tag{236}$$

The expense of obtaining the coupled treatment is heavy: instead of a single $\mathcal{O}(N^4)$ iterative procedure like a Hartree-Fock single point or gradient computation, we now have a $\mathcal{O}(N^5)$ iterative procedure, which is more analogous to a Hartree-Fock Hessian

computation. In either case, once the induction amplitudes are obtained, the order-2 partition is computed as

$$E_{\rm ind}^{(20)}[\mathbb{A} \leftarrow \mathbb{B}] = \sum_{a\mathbf{B}} E_{\rm ind}^{(20),a\mathbf{B}}[\mathbb{A} \leftarrow \mathbb{B}] = \sum_{a\mathbf{B}} 2x_{ar}^{\mathbf{B}} V_{ar}^{\mathbb{B}'}, \tag{237}$$

and

$$E_{\rm exch-ind}^{(20)}[S^2][\mathbb{A} \leftarrow \mathbb{B}] = \sum_{a\mathbf{B}} E_{\rm exch-ind}^{(20),a\mathbf{B}}[S^2][\mathbb{A} \leftarrow \mathbb{B}] = \sum_{a\mathbf{B}} 2x_{ar}^{\mathbf{B}} \tilde{V}_{ar}^{\mathbb{B}'}. \tag{238}$$

Note that we do not add the order-2 induction contributions from $\mathbb{A} \leftarrow \mathbb{B}$ and $\mathbb{B} \leftarrow \mathbb{A}$ together to make a single induction partition: it is more useful to visualize the response of monomer A due to monomer B separate from the response of monomer B due to monomer A.

At this point, there are one or two pieces missing from the order-2 induction partition, relative to SAPT0. In the case that the uncoupled singles amplitudes were used (as we will always do for computational efficiency, unless explicitly indicated), the partition will be missing the contributions due to the coupling. Additionally, in either case, the partition will be missing infinite-order polarization effects, which are captured in the hybrid $\delta_{\rm HF}^{(2)}$ correction in SAPT0 theory. In either case, we will assume that the missing effects are isotropic magnifications of the existing polarizations and scale the partition to match the total SAPT0 induction energy, e.g.,

$$E_{\rm ind}^{a\mathbf{B}} = S_\delta S_{\rm r}[\mathbb{A} \leftarrow \mathbb{B}] \left[E_{\rm ind}^{(20),a\mathbf{B}}[\mathbb{A} \leftarrow \mathbb{B}] + E_{\rm exch-ind}^{(20),a\mathbf{B}}[S^2][\mathbb{A} \leftarrow \mathbb{B}] \right], \tag{239}$$

where the monomer-uniform scaling S_δ to account for hybrid induction is

$$S_\delta \equiv \frac{E_{\rm ind,r}^{(20)} + E_{\rm exch-ind,r}^{(20)}[S^2] + \delta_{\rm HF,r}^{(2)}}{E_{\rm ind,r}^{(20)} + E_{\rm exch-ind,r}^{(20)}[S^2]}, \tag{240}$$

and the monomer-specific scaling $S_{\rm r}[\mathbb{A} \leftarrow \mathbb{B}]$ to account for coupling is

$$S_{\rm r}[\mathbb{A} \leftarrow \mathbb{B}] \equiv \begin{cases} \frac{E_{\rm ind,r}^{(20)}[\mathbb{A}\leftarrow\mathbb{B}] + E_{\rm exch-ind,r}^{(20)}[S^2][\mathbb{A}\leftarrow\mathbb{B}]}{E_{\rm ind,u}^{(20)}[\mathbb{A}\leftarrow\mathbb{B}] + E_{\rm exch-ind,u}^{(20)}[S^2][\mathbb{A}\leftarrow\mathbb{B}]}, & \text{Uncoupled A-SAPT} \\ 1, & \text{Coupled A-SAPT} \end{cases}. \tag{241}$$

This scaling technique is used for all results presented in this work.

Note that the scaling for uncoupled singles amplitudes can be avoided by using coupled amplitudes, albeit at increased computational cost, though we show below that the error due to scaled, uncoupled singles amplitudes is qualitatively negligible across the S22 database. The error due to scaling for $\delta^{(2)}_{\mathrm{HF,r}}$ is unknown at the present time, but can be avoided in future efforts by developing A-SAPT partitions for $E^{(30)}_{\mathrm{ind}}$, etc in higher-order SAPT.

8.3.2.5 Dispersion

As introduced in the section on exchange, dispersion is a two-particle correlation in which pairs of $a \rightarrow r$ and $b \rightarrow s$ excitations occur in a coupled manner, to lower the overall Coulomb repulsion. In SAPT0, the excitations within a monomer are each uncoupled from all the others. This sometimes leads to substantial overbinding of polarizable systems. However, adding intramonomer coupling requires a great deal of effort (either via MBPT or CCD+(ST) contributions to the dispersion amplitudes, or TD-DFT coupled susceptibility approaches). Fortunately, due to favorable error cancelation, SAPT0 in a well-chosen basis set (particularly jun-cc-pVDZ) has been shown to provide remarkable combination of accuracy and efficiency for a broad range of systems.[214] As with induction, the two-body dispersion initially fails to respect the Pauli-excluded space of monomer B; the exchange-dispersion term adds a corrective "exchange-dispersion potential" to penalize excitations which cross-penetrate each others occupied spaces. The diagrams corresponding to these processes are depicted in the lower panel of Figure 25.

The uncoupled two-body dispersion is

$$E^{(20)}_{\mathrm{disp}} \equiv 4 t^{rs}_{ab} v^{ab}_{rs}. \tag{242}$$

The uncoupled amplitudes are simply

$$t^{rs}_{ab} \equiv \frac{v^{rs}_{ab}}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}. \tag{243}$$

After some manipulation from the standard forms encountered in the literature, the exchange-dispersion term in the S^2 approximation can be evaluated as a contraction with identical topology as the dispersion term, albeit with a modified "exchange-dispersion potential,"

$$E^{(20)}_{\mathrm{exch-disp}}[S^2] \equiv -2t^{rs}_{ab}\tilde{v}^{ab}_{rs}, \tag{244}$$

where

$$\begin{aligned}\tilde{v}^{rs}_{ab} \equiv &+1v^{sr}_{ab}+1S^s_a(2v^{ra'}_{ba'}-v^{a'r}_{ba'})+1S^s_{a'}(2v^{ra'}_{ab}-v^{a'r}_{ab})+1S^r_b(2v^{sb'}_{ab'}-v^{b's}_{ab'})+1S^r_{b'}(2v^{sb'}_{ba}-v^{b's}_{ba})\\
&+1S^r_bS^{b'}_{a'}v^{a's}_{ab'}-2S^r_{b'}S^{b'}_{a'}v^{a's}_{ab}-2S^r_bS^{b'}_a v^{a's}_{a'b'}+4S^r_{b'}S^{b'}_a v^{a's}_{a'b}+1S^s_aS^{a'}_{b'}v^{rb'}_{a'b}-2S^s_{a'}S^{a'}_{b'}v^{rb'}_{ab}-2S^s_aS^{a'}_b v^{rb'}_{a'b'}\\
&+4S^s_{a'}S^{a'}_b v^{rb'}_{ab'}+1S^r_{b'}S^s_{a'}v^{a'b'}_{ab}-2S^r_bS^s_{a'}v^{a'b'}_{ab'}-2S^r_{b'}S^s_a v^{a'b'}_{a'b}+1S^{a'}_bS^{b'}_a v^{rs}_{a'b'}-2S^{a'}_bS^{b'}_{a'}v^{rs}_{ab'}-2S^{a'}_{b'}S^{b'}_a v^{rs}_{a'b}\\
&+2S^r_{b'}S^{b'}_a V^A_{bs}+2S^s_{a'}S^{a'}_b V^B_{ar}-1S^r_bS^{b'}_a V^A_{b's}-1S^s_aS^{a'}_b V^B_{a'r}+1S^s_aS^r_{s'}V^A_{bs'}+1S^r_bS^s_{r'}V^B_{ar'}.\end{aligned} \tag{245}$$

This potential can be constructed in blocks at a few times the cost of forming the v^{rs}_{ab} integrals, as detailed in the supplemental material.

Obviously, the key order-2 partition involves the occupied spaces a and b on both monomers, as indicated by the circled key bodies in the diagrams, yielding the mathematical partition,

$$E^{(20)}_{\mathrm{disp}} = \sum_{ab} E^{(20),ab}_{\mathrm{disp}} = \sum_{ab} 4t^{rs}_{ab}v^{ab}_{rs}, \tag{246}$$

and

$$E^{(20)}_{\mathrm{exch-disp}}[S^2] = \sum_{ab} E^{(20),ab}_{\mathrm{exch-disp}}[S^2] = \sum_{ab} -2t^{rs}_{ab}\tilde{v}^{ab}_{rs}. \tag{247}$$

This partition of the dispersion energy to pairs of occupied orbitals (in particular, to pairs of local occupied orbitals) strongly resembles the quasi-physical interpretation of the pair energies in local correlation methodology, and has been done before for dispersion in the context of LMP2.[133,97]

8.3.3 Quasiparticle Spatial Localization

8.3.3.1 Electrostatic Source Localization

In the above order-2 partition, the total electrostatic potentials of nuclei and electrons were encountered in the electrostatic and induction terms. These electrostatic potentials drive these two types of interactions, so we refer to them as "source terms." Due to extraordinary cancellation between charges with opposite signs, it is necessary to provide a highly localized spatial decomposition for these terms which balances the nuclear and electronic contributions. The concept we are seeking here is that of an atomic charge density field. While the qualitative decomposition of the electronic density into atomic contributions is an intuitive process, the concept has been found to be remarkably ill-defined and ambiguous mathematically, leading to myriad schemes for atomic partitioning in the literature. For the purposes of matching the SAPT total energy components, accurate application of numerical integration, and intuitive interpretation of results, we seek an atomic partitioning scheme with the following characteristics:

Normalization: the sum of atomic densities should exactly match the total molecular density, i.e., $\sum_{\mathbf{A}} \rho_{\mathbf{A}}(\vec{r}) = \rho_{\mathbb{A}}(\vec{r})$.

Smoothness: the atomic densities must decay smoothly as $r \to \infty$, jump discontinuities at bond midpoints are not allowed.

Elegance: the atomic densities should match the intuitive understanding of an atom as a nearly spherical quantity whose normalization follows electronegativity differences.

Intrinsicality: the atomic density partitioning should be rigorously defined in terms of only the molecular electronic density and the positions of the nuclei; external computations or parameters are not permitted.

These considerations lead to our selection of the iterative stockholder analysis (ISA) of Wheatley and co-workers.[178,179,315] This method first identifies the "closest" set of spherical atoms $\rho^0_{\bar{A}}(\vec{r}) = f_{\bar{A}}(|\vec{r} - \vec{r}_A|)$ to the molecular density, by minimizing the objective functional,[39]

$$O\left(\rho^0_{\mathcal{A}} \equiv \sum_{\bar{A}} \rho^0_{\bar{A}}\right) \equiv -\int_{\mathbb{R}^3} \mathrm{d}r_1 \ \left[\rho^0_{\mathcal{A}}(\vec{r}_1) - \rho_{\mathcal{A}}(\vec{r}_1) \ln\left(\frac{\rho^0_{\mathcal{A}}(\vec{r}_1)}{\rho_{\mathcal{A}}(\vec{r}_1)}\right)\right]. \tag{248}$$

The leading minus sign reflects our negative definite nature of the electronic density, which was adopted in Equation 218. Next, the spherical but approximate "proatomic" densities $\rho^0_{\bar{A}}$ (also called "shape functions") are used to generate non-spherical but exactly normalized atomic densities,

$$\rho_{\bar{A}}(\vec{r}) \equiv \frac{\rho^0_{\bar{A}}(\vec{r})}{\sum_{\bar{A}'} \rho^0_{\bar{A}'}(\vec{r})} \rho_{\mathcal{A}}(\vec{r}). \tag{249}$$

In practice, the atomic densities are determined iteratively on an atom-centered Becke grid,[20] with nodes and weights $\{\vec{r}_P, w_P\}$. Given a trial set of proatomic densities, the atomic densities can be generated by Equation 249. From a non-converged set of atomic densities, an update to the proatomic densities can be obtained by spherically averaging the atomic density,

$$\rho^0_{\bar{A}}(r_{i\bar{A}}) = \langle \rho_{\bar{A}}(\vec{r}) \rangle_{i\bar{A}}, \tag{250}$$

where $\langle \ldots \rangle_{i\bar{A}}$ indicates a spherical average over the Lebedev grid at the i-th radial node on atom $\bar{A}$. Equations 249 and 250 are then iterated to self-consistency.

Once the atomic densities are generated on the Becke grid, the remaining task is to integrate the various electrostatic source terms. Unfortunately, the grid-based decomposition of the molecular density is no longer directly representable in the space of orbital pairs, i.e., $\rho_{\bar{A}}(\vec{r}) \neq -2D^{\bar{A}}_{\mu\mu'}\phi_\mu(\vec{r})\phi_{\mu'}(\vec{r})$. This implies that the standard spectral-basis electron repulsion integrals cannot be used in electrostatic source terms. At this point, we might be tempted to use the Becke quadrature weights to resolve

the integral traces via direct numerical integration, e.g.,

$$E_{\text{elst,r}}^{(10),\bar{A}\bar{B}} \stackrel{?}{\approx} w_P \rho_{\bar{A}}(\vec{r}_P) \frac{1}{r_{PP'}} \rho_{\bar{B}}(\vec{r}_{P'}) w_{P'}. \tag{251}$$

Unfortunately, the grid-space Coulomb operator $1/r_{PP'}$ has a measure-3 singularity where $P = P'$, and extremely ill-conditioned otherwise. This was the same difficulty initially encountered in our work on the tensor hypercontraction (THC) representation for the conventional electron repulsion integrals.[121,216,218] Borrowing a trick from our efforts in THC, we can overcome this difficulty by projecting the grid-space atomic densities onto an auxiliary spectral basis set, using the robust Coulomb metric,

$$V_{\bar{A}}^{C} \equiv \langle \rho_{\bar{A}} | r_{12}^{-1} | \chi_C \rangle = \int_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \rho_{\bar{A}}(\vec{r}_1) \frac{1}{r_{12}} \chi_C(\vec{r}_2) \approx w_P \rho_{\bar{A}}(\vec{r}_P) \underbrace{\int_{\mathbb{R}^3} \mathrm{d}r_2 \; \frac{1}{r_{2P}} \chi_C(\vec{r}_2)}_{v_C^P}. \tag{252}$$

Here, the integrals v_C^P resemble the usual nuclear potential integrals, albeit with a single orbital index. This step is highly analogous to a pseudospectral projection (PS) for each atomic density.[80,81,82] In this representation, the finite, smooth support of the auxiliary basis $\{\chi_C(\vec{r})\}$ renormalizes the previously singular Coulomb operator. The electronic-electronic portion of the electrostatic energy is now resolved as

$$E_{\text{elst,r}}^{(10),\bar{A}\bar{B}} \approx V_{\bar{A}}^{C} (C|C')^{-1} V_{\bar{B}}^{C'}. \tag{253}$$

Here $(C|C')^{-1}$ is the conditioned pseudoinverse of the density fitting metric matrix,

$$(C|C') \equiv \iint_{\mathbb{R}^6} \mathrm{d}r_1 \; \mathrm{d}r_2 \; \chi_C(\vec{r}_1) \frac{1}{r_{12}} \chi_{C'}(\vec{r}_2). \tag{254}$$

This step is essentially a density-fitted (DF) electrostatics computation, but with atomic resolution. In the limit that the quadrature/ISA Becke grid $\{\vec{r}_P, w_P\}$ is complete, the results will attain the ISA-complete density-fitted DF-SAPT0 limit. As the density fitting basis is saturated, the results will approach the SAPT0 limit.

Another way to arrive at this result is to start from a density fitting resolution of the electron-electron electrostatics term,

$$E_{\text{elst,r}}^{(10),\bar{A}\bar{B}} = \langle \rho_{\bar{A}} | r_{12}^{-1} | \rho_{\bar{B}} \rangle \approx \langle \rho_{\bar{A}} | r_{12}^{-1} | \chi_C \rangle \langle \chi_C | r_{12}^{-1} | \chi_{C'} \rangle^{-1} \langle \chi_{C'} | r_{12}^{-1} | \rho_{\bar{B}} \rangle, \tag{255}$$

and then realize that the integrals $\langle \rho_{\bar{A}} | r_{12}^{-1} | \chi_C \rangle$ are not analytical, but must be done by quadrature, becoming the $V_{\bar{A}}^{C}$ quantities in Equation 252. Note, however, that the use of density fitting is actually the critical portion of the procedure: without this intermediate projection through a smooth spectral space, singularities would plague the result.

In practice, the nuclear-electronic terms are not accurately treated by an intermediate projection through a standard auxiliary basis, as there is no support for extremely tight functions which represent nuclear density fields. However, this difficulty is easily overcome, as the electronic-nuclear singularity is only measure-0 and the radial and atomic components of the Becke grids ameliorate this numerical singularity as the nucleus is approached. Therefore, the electronic-nuclear terms are easily handled by direct quadrature,

$$E_{\text{elst}}^{(10),\bar{A}B} \approx w_P \rho_{\bar{A}}(\vec{r}_P) \frac{1}{r_{PB}} Z_B. \tag{256}$$

Similar techniques are used for the formation of the atomic electrostatic potentials for the induction terms, i.e., the quantities in Equation 233. For the electronic source, the THC projection is used,

$$V_{ar}^{\bar{B}} \approx -(ar|C)(C|C')^{-1} V_{\bar{B}}^{C'}. \tag{257}$$

Here $(ar|C)$ is a three-center Coulomb integral,

$$(ar|C) \equiv \iint_{\mathbb{R}^6} \mathrm{d}r_1 \ \mathrm{d}r_2 \ \phi_a(\vec{r}_1) \phi_r(\vec{r}_2) \frac{1}{r_{12}} \chi_C(\vec{r}_2). \tag{258}$$

For the nuclear source, direct integration is used,

$$V_{ar}^{B} = -\int_{\mathbb{R}^3} \mathrm{d}r_1 \ \phi_a(\vec{r}_1) \phi_r(\vec{r}_1) \frac{1}{r_{1B}} Z_B. \tag{259}$$

These integrals are standard nuclear potential integrals. The total atomic electrostatic potential is just the summation of these two contributions,

$$V_{ar}^{\mathbf{B}} = V_{ar}^{B} + V_{ar}^{\bar{B}}. \tag{260}$$

8.3.3.2 Electronic Response Localization

In the above order-2 partition, electronic response terms involving occupied orbital labels occurred in the exchange, induction, and dispersion terms. Here "response" indicates that these terms involve a rearrangement of the electrons of the monomers due to an electrostatic, Pauli, or correlation source. To this point, the occupied space has been assumed to be spanned by canonical Hartree-Fock quasiparticles a or b, to allow for explicit forms for the uncoupled amplitudes in induction and dispersion. Unfortunately, these canonical occupied orbitals often span the entirety of each monomer, precluding a localized visualization. Fortunately, once the amplitudes are formed, we are free to transform both the amplitudes and potential integrals to any convenient quasiparticle basis which also spans the occupied space, by means of a unitary transformation. In particular, we seek to employ a set of quasiparticles which have minimal spatial extent, e.g., local orbitals,

$$\phi_{\bar{a}}(\vec{r}) = U^a_{\bar{a}} \phi_a(\vec{r}). \tag{261}$$

Here $U^a_{\bar{a}}$ is a unitary transformation from the canonical occupied orbital space a to the local occupied orbital space $\bar{a}$, as can be obtained from Pipek-Mezey,[228] Boys,[34] or several other orbital localization procedures.[72,9,126,149,172]

Using the canonical-basis amplitudes and the localizing unitary transformation, we can obtain a order-2 partition over local occupied quasiparticles. For example, for the case of $E^{(20)}_{\text{disp}}$,

$$E^{(20)}_{\text{disp}} = 4 t^{rs}_{ab} v^{ab}_{rs} = 4 t^{rs}_{ab} v^{a'b'}_{rs} \underbrace{U^a_{\bar{a}} U^{\bar{a}}_{a'}}_{\delta_{aa'}} \underbrace{U^b_{\bar{b}} U^{\bar{b}}_{b'}}_{\delta_{bb'}} = 4 \underbrace{\left[U^a_{\bar{a}} t^{rs}_{ab} U^b_{\bar{b}}\right]}_{t^{rs}_{\bar{a}\bar{b}}} \underbrace{\left[U^{\bar{a}}_{a'} v^{a'b'}_{rs} U^{\bar{b}}_{b'}\right]}_{v^{\bar{a}\bar{b}}_{rs}} = \sum_{\bar{a}\bar{b}} 4 t^{rs}_{\bar{a}\bar{b}} v^{\bar{a}\bar{b}}_{rs} = \sum_{\bar{a}\bar{b}} E^{(20),\bar{a}\bar{b}}_{\text{disp}}. \tag{262}$$

Applying directly analogous substitutions in the exchange and induction response partitions yields local occupied quasiparticle partitions for all electronic response terms.

If further abstraction from local quasiparticles to atoms is desired for electronic response terms, a fuzzy weighting provided by positive-definite local occupied orbital atomic charges may be used, e.g.,

$$E_{\mathrm{disp}}^{(20),\mathbf{AB}} = Q_{\bar{a}}^{\mathbf{A}} Q_{\bar{b}}^{\mathbf{B}} E_{\mathrm{disp}}^{(20),\bar{a}\bar{b}}. \tag{263}$$

Here $Q_{\mathbf{A}}^{\bar{a}}$ is the matrix of local orbital atomic charges, as obtained by any desired population metric. We note that the popular Mulliken and Löwdin charges are quite sensitive to basis set, so we instead advocate the use of orbital charges derived from the ISA atomic densities,

$$Q_{\mathbf{A}}^{\bar{a}} \equiv \langle \bar{a}|\hat{w}_{\mathbf{A}}|\bar{a}\rangle = \int_{\mathbb{R}^3} \mathrm{d}r_1 \; |\phi_{\bar{a}}(\vec{r}_1)|^2 \underbrace{\frac{\rho_{\bar{A}}(\vec{r}_1)}{\rho_{\mathcal{A}}(\vec{r}_1)}}_{w_{\mathbf{A}}(\vec{r}_1)} \approx w_P |\phi_{\bar{a}}(\vec{r}_P)|^2 \frac{\rho_{\bar{A}}(\vec{r}_P)}{\rho_{\mathcal{A}}(\vec{r}_P)}. \tag{264}$$

This definition of ISA orbital atomic charges is not unique, e.g., other authors have very recently used the proatomic density fields instead of the true atomic density fields for the product of orbital atomic charges.[172] However, the difference in orbital-to-atom assignment should be minor (our experience indicates that even Mulliken charges are acceptable for qualitative purposes). The last expression above provides a practical but approximate formula of computing these orbital atomic charges on the Becke integration grid. This implies that the total orbital norms (the sum of $Q_{\mathbf{A}}^{\bar{a}}$ over all $\mathbf{A}$ for each $\bar{a}$) will not exactly be unity. In practice, this error is observed to be at most of the order of 10^{-5} electrons for reasonable grid sizes, and is partially ameliorated by renormalizing the orbital atomic charges. This renormalization is used for all results presented in this work.

Note that, in contrast to the more traditional use of local orbitals to obtain computational accelerations in correlated methods,[259,260] here we first form the explicit amplitudes in the canonical orbital basis, then transform both the amplitudes and potential integrals to the local basis. As such, we obtain the canonical dispersion energy as a sum of local pair-energies, without needing to optimize the Hylleraas

functional. Such a Hylleraas formalism could be employed to obtain computational acceleration in concert with the A-SAPT partition for systems with more than a few hundred atoms (e.g., a Saebø-Pulay local A-SAPT0). However, this is outside the scope of the present work, which is concerned primarily with the existence of the A-SAPT partition, rather than its computational efficiency.

8.3.4 Algorithmic Details

8.3.4.1 DF-A-SAPT0 Implementation

The A-SAPT0 code detailed above was implemented in a development branch of the open source `PSI4` *ab initio* quantum chemistry package,[298] using density fitting techniques to accelerate the computation. Our previous SAPT0 implementation from our 2011 paper [SAPT0 ('11)][118] uses density fitting and Laplace transformation techniques to attain a tractability limit of several hundred atoms. However, the order-2 partitions of A-SAPT0 are incompatible with Laplace transformation acceleration of the exchange-dispersion terms, and many additional terms are required in A-SAPT0. To facilitate the rapid and correct implementation of A-SAPT0, we have implemented a new DF-SAPT0 code [SAPT0 ('13)]. SAPT0 ('13) more closely resembles the DF-SAPT(DFT) implementation of Heßelmann and co-workers[111] than SAPT0 ('11). The electrostatics, exchange, and induction terms are formulated in terms of effective Fock matrix builds and reuse the AO-basis three-index density fitting integrals from the underlying SCF computations via the machinery of our `LibFock` generalized Fock matrix infrastructure. The rate-limiting dispersion and exchange dispersion terms are formulated in a manner which closely resembles DF-MP2,[76] using a new factorization of the exchange dispersion potential detailed in the supplemental material. For SAPT0 ('13), several sets of generalized three-index density fitting integrals are built via a pilot version of the `LibPANACHE` tensor manipulation package. Then, for each *ab* pair, the dispersion potential, dispersion amplitudes, and exchange-dispersion potential are formed in $\mathcal{O}(N_r N_s N_D)$ effort, with the rate-limiting $\mathcal{O}(N^5)$ steps being

a series of DGEMM calls threaded over ab pairs.

Our DF-A-SAPT0 code [A-SAPT0 ('13)] is built as an extension of SAPT0 ('13). The Fock term portion of SAPT0 ('13) is called to provide the exchange and coupled induction scaling factors, and to provide a check of the numerical accuracy of the grid-based A-SAPT0 electrostatics computation. The A-SAPT0 ('13) order-2 dispersion computation is topologically quite similar to SAPT0 ('13), except that we must drive the computation over rs pairs instead of ab pairs, to obtain the canonical basis dispersion potential, dispersion amplitudes, and exchange-dispersion potential for all ab pairs for a given rs pair, to allow for transformation to the $\bar{a}\bar{b}$ basis. In practice, this introduces a slight additional overhead because the rate-limiting DGEMM calls are now only $\mathcal{O}(N_a N_b N_D)$ in size, and are not fully saturated in terms of FLOP performance. Additionally, some small overhead is incurred in the canonical-to-local transformation of amplitudes and potentials. As the same dispersion energy is returned in both algorithms, we do not additionally compute the dispersion using the SAPT0 ('13) algorithm. The grid-based ISA algorithms, atomic pseudospectral (PS) projections, and voxel grid analysis portions of A-SAPT0 ('13) are all built from the same `LibFock` infrastructure used to provide Kohn-Sham exchange-correlation potentials in our DFT codes. The additional atomic density fitting (DF) and order-2 Fock term partitions are accomplished with generalized three-index integrals from the `LibPANACHE` package, with usually minimal additional overhead. The exception is if coupled induction is required in A-SAPT0, in which case the CPHF equations must be solved for each atomic source.

The complete DF-A-SAPT0 ('13) algorithm is available in the supplemental material.

8.3.4.2 ISA Convergence Acceleration

As has been noted by other authors,[307] the standard ISA algorithm discussed above is extremely difficult to converge, particularly in later iterations and for smaller grids. The objective functional appears to be quite ill-conditioned, especially in the tails of the shape functions. Moreover, the universal guess for the shape functions of $\rho_{\bar{A}}^{0}(\vec{r}) = -1$ suggested by Wheatley and co-workers is quite far from the converged solution. In our experience, this situation does not seem to be solvable by standard quasi-Newton acceleration techniques such as Pulay's DIIS[238] (unless exceedingly large grids are used to provide a numerically exact gradient), as no exact gradient is available, and the solution must be constrained to be positive definite. To improve performance of the ISA procedure, we have implemented a superposition of atomic densities guess, much as is done in SCF computations,[304] and a form of Newton-Raphson with an approximate gradient and Hessian. Details are available in the supplemental material.

8.3.5 Caveats and Alternatives

It is critical to remember that A-SAPT contributions are not observable quantities: A-SAPT is merely one particularly defined partition of the SAPT perturbation series. Intermolecular forces are intrinsically many-body interactions between indistinguishable particles, and the only direct observable is the interaction energy, or, more properly, the binding energy. Non-observable quantities such as A-SAPT contributions (or even the SAPT terms themselves) should not be ascribed absolute physical realism, and should be used only insofar as they serve to aid our chemical understanding of intermolecular interactions.

By virtue of there being many quasiparticles in the SAPT diagrams (and many possible choices of local bases for these quasiparticles), any order-2 partition of the intermolecular interaction energy components is necessarily non-unique, e.g., we could

have focused on the localized representations of the virtual space or orbital pair space instead of localized occupied quasiparticles. In developing A-SAPT, we have considered many choices for order-2 selection of the quasiparticles and the local quasiparticle basis, and decided that focusing on the local occupied quasiparticles makes the most intuitive sense. One key feature in any sensible partition is that the deletion or addition of a functional group should be accompanied by the deletion or addition of the direct interaction energy contributions arising from the electronic and nuclear density of that functional group. The local occupied quasiparticles directly outline the density of the added or deleted functional group, and approach an asymptotic and still-localizable limit in the complete basis set. This is not the case for virtual quasiparticles, which may be infinitely diffuse in the complete basis set. This strongly suggests that the local occupied quasiparticles should be preferred to the local virtual quasiparticles. Indeed, such a selection has previously been used in the context of local correlation methodology, with the intermolecular local occupied pair energies from LMP2 or other methods being interpreted as localized contributions to the dispersion energy.[133,97]

Another approach considered in the development of A-SAPT was the use the atomic centers of the density fitting auxiliary basis functions to provide atomic assignment of the orbital pair spaces, particularly in the electrostatic energy. Such methodology has been well developed by Stone and co-workers in the context of distributed polarizabilities.[256,196] However, several difficulties were found in this approach. The first is that the most accurate Coulomb metric density fitting scheme fits the nonlocal electrostatic potential rather than the local density field, so sensible assigment based on the atomic centers of the density fitting functions is quite difficult. Misquitta and Stone have introduced parametrized constaints into the density fitting functional to force the auxiliary functions to represent only nearby portions of the pair space, but the results are quite sensitive to the magnitude of the locality

penalty.[196] The second difficulty encountered in our case was that assignment based on auxiliary function atomic centers mandates the use of a monomer-centered auxiliary basis, which is grossly inaccurate for the representation of the dimer-centered pair-space basis (to the point that the sign of the electrostatic energy was often lost). These problems with density-fitting-based partitioning led to the use of ISA atomic density fields for the electrostatic source terms.

8.4 Results

8.4.1 Computational Details

The SAPT0 and A-SAPT0 results shown below use the jun-cc-pVDZ primary basis set,[124,212] which has been shown to attain remarkable error cancellation between the uncoupled, two-body treatment of dispersion and the basis set incompleteness, leading to a semiquantitative method which is applicable to large systems. Density fitting is used throughout to accelerate the Hartree-Fock and SAPT0 computations. The jun-cc-pVDZ-JKFIT auxiliary basis is used for the Hartree-Fock computations, the electrostatics, exchange, and induction terms, and for all Fock-like terms in the exchange-dispersion potential. The jun-cc-pVDZ-RI auxiliary basis is used for all ERIs explicitly occurring in the dispersion and exchange-dispersion terms, i.e., those ERI contributions not reducible to Fock-like matrices. The frozen core treatment is applied unless otherwise noted, affecting only the dispersion and exchange-dispersion terms, and simply amounts to setting the dispersion amplitudes involving frozen core orbitals to zero: other core orbitals appearing in the exchange dispersion potential are required to account for the Pauli-forbidden space and are not deleted. The hybrid $\delta^{(2)}_{\mathrm{HF,r}}$ correction is included in all SAPT0 computations reported here, and is included in A-SAPT0 via the scaling factors discussed above.

Unless otherwise noted, for the ISA procedure we use the Treutler-Ahlrichs[297] flavor of the Becke molecular quadrature grids, with 302 spherical nodes and 100 radial

nodes per atom. The ISA algorithm uses the superposition of atomic densities guess, and is converged to $\epsilon < 10^{-7}$, using the approximate Newton-Raphson update detailed above in every third step, with the Newton-Raphson update disabled after 1000 steps. The default localization scheme is Pipek-Mezey, and core and valence localizations are performed separately. The default treatment of the A-SAPT induction is uncoupled, with scale factors to account for both the coupling and higher-order induction, as detailed in Equations 239 to 241. The majority of these prescriptions will be justified by performing a sensitivity analysis below.

8.4.2 Example Order-2 A-SAPT0 Results: HF Dimer

To provide a representative example demonstrating the order-2 atomic partition and asymptotic behavior of A-SAPT, we consider the case of anti-aligned HF dimer at various intermolecular separations R. The geometry and order-2 atomic A-SAPT0/jun-cc-pVDZ results are depicted in Figure 26. The HF monomer is defined by the DF-MP2/aug-cc-pVDZ optimal bond length of 0.9248 Å, and the dimer geometry is completely specified by placing the monomers in a side-on anti-parallel configuration at an intermolecular separation of R. Electrostatics, induction, and dispersion are found to exhibit power-law asymptotic decay, and are plotted on log-log scales. Exchange is found to exhibit exponential asymptotic decay, and is plotted on a semi-log scale. Note that the absolute values of the interaction energy components must be taken to allow for the application of the logarithm; to illustrate the sign of each contribution, attractive contributions are plotted with solid lines, repulsive contributions are plotted with dotted lines. The induction terms for (A←B) and (B←A) are identical due to the symmetry of the monomers, so only the former is plotted. Least-squares power law fits of the form $|E_{\text{int}}| = AR^N$ are performed for electrostatics, induction, and dispersion contributions, while a least-squares exponential fit of the form $|E_{\text{int}}| = Ae^{bR}$ is perfomed for exchange. In each case, the seven data points

for $R \geq 4.0$ Å are used to prevent contamination from charge penetration terms at short range. The R^2 values in all cases are at least 0.996. The R^2 values from the least-squares fits are available in the supplemental material.

For electrostatics, the F$\cdots$H and H$\cdots$F interactions are both attractive, while the F$\cdots$F and H$\cdots$H interactions are both repulsive. The sum of the former two are slighlty larger in magnitude than the sum of the latter two, so the overall contribution is attractive. The individual order-2 contributions decay as R^{-1} due to the partial charges on each ISA atom, while the sum of all order-2 contributions nearly cancels and decays as R^{-3}, as expected of an anti-aligned dipole-dipole interaction. At very short ranges, the F$\cdots$H interactions become more rapidly attractive, while the F$\cdots$F and H$\cdots$H interactions become less rapidly repulsive, indicating charge penetration contributions. These results are all compatible with chemical intuition and quite similar at long range to those that might be approximately obtained in force field computations or partitioned from an *ab initio* method by distributed multipole analysis (DMA).[283,285,284,286] However, A-SAPT0 respects the full SAPT0 limit, particularly including charge penetration.

The other three terms show similar results, except that there are no canceling contributions and the individual order-2 contributions decay at roughly the same rates as the total. For exchange, exponential decay is apparent, and is dominated by the F$\cdots$F contribution, which is consistent with the larger extent of the F atoms. Besides smaller overall magnitudes, the F$\cdots$H and H$\cdots$H exchange contributions are also seen to decay with slightly faster rates than F$\cdots$F, which is consistent with the faster decay of the tails of the H atoms. Induction is seen to exhibit R^{-6} decay in terms of both individual order-2 contributions and the total, which is consistent with dipole-induced-dipole interactions involving a permanent dipole with fixed orientation. The terms corresponding to F$\leftarrow$F and F$\leftarrow$H in which the F atom of monomer A are polarized by the partial charges on either atom of monomer B dominate this term,

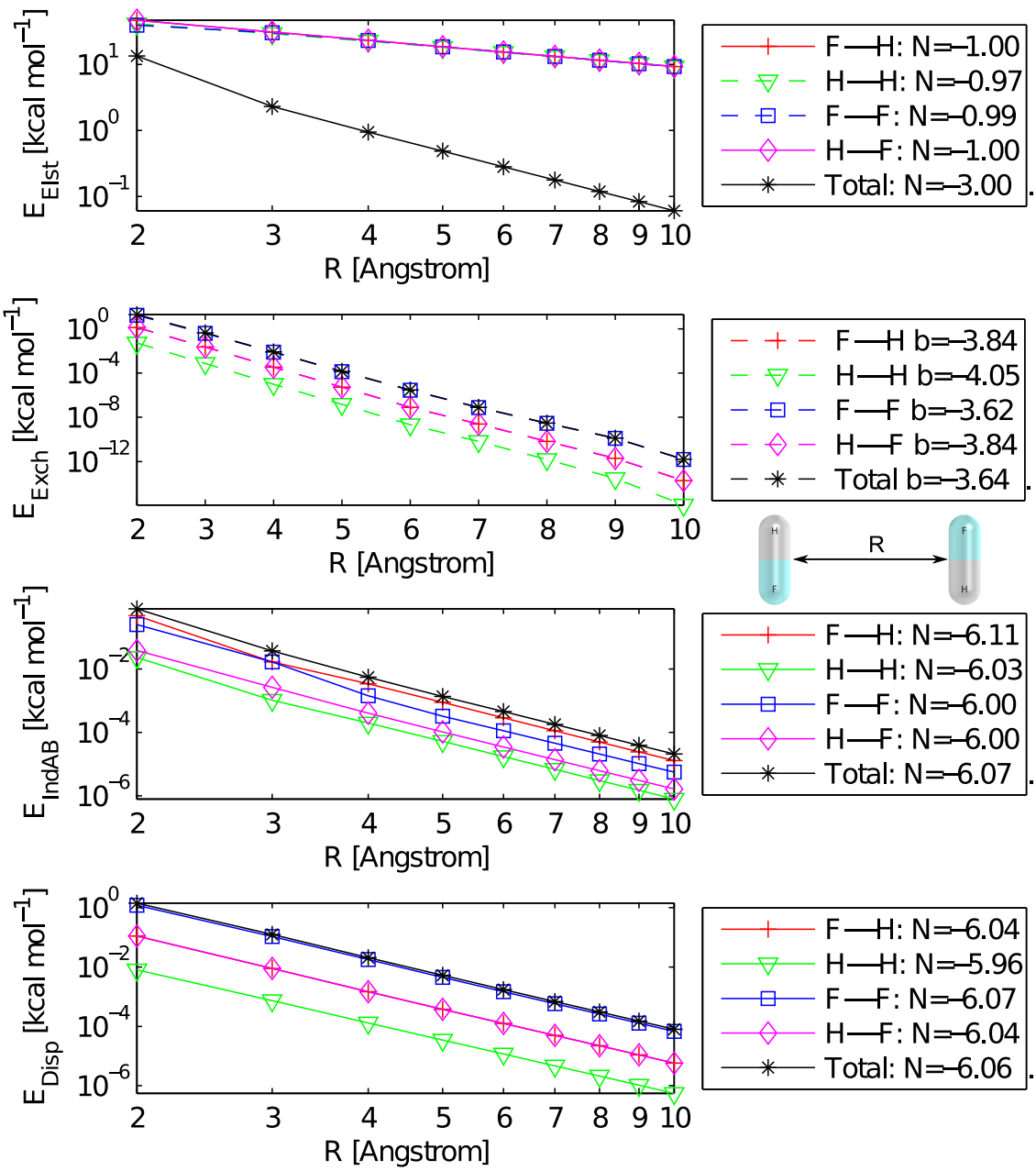

Figure 26: Order-2 atomic A-SAPT0/jun-cc-pVDZ analysis for anti-aligned HF dimer vs. the intermolecular separation R. Electrostatics, induction, and dispersion contributions are plotted on a log-log scale and fitted to a power law relationship of the form $|E_{\text{int}}| \approx AR^N$ for $R \geq 4$ Å. Exchange is plotted on a semi-log scale and fitted to an exponential relationship of the form $|E_{\text{int}}| \approx Ae^{bR}$ for $R \geq 4$ Å. Attractive contributions are denoted by solid lines, repulsive contributions are denoted by dashed lines. The HF bond distance is 0.9248 Å.

as is consistent with the higher polarizability of the F atom. Dispersion shows clear and expected R^{-6} decay behavior, and is almost wholly dominated by the F$\cdots$F interaction.

8.4.3 Order-1 Visualization Techniques

The order-2 atomic partition developed in this work has obvious applications in providing unbiased atom-pair potentials for developing next-generation force fields, and also contains a wealth of qualitative data on the important contacts in noncovalent interaction. However, the latter application is possibly plagued by an overload of data. In the HF dimer example above, each SAPT term has 4 atom-pair contributions; by the time adenine-thymine dimer is considered, this rises to 225 atom-pair contributions per SAPT term. In this section, we develop an order-1 reduction of the order-2 data to enable easier absorption of the salient details of the full order-2 A-SAPT partition.

To facilitate rapid interpretation of the A-SAPT results, we have developed a pair of visualization schemes for order-1 summations of the order-2 A-SAPT partition. The first is an atomic order-1 analysis, wherein the atoms of monomer A are colored according to the magnitude of the atomic interaction energy term contribution,

$$E_{\text{int}}^{\mathbf{A}} = \sum_{\mathbf{B}} E_{\text{int}}^{\mathbf{AB}}. \tag{265}$$

The physical interpretation of this metric is the amount to which atom **A** is attracted to or repelled from the entirety of monomer B, *in the net*.

For more detailed order-1 analysis, it is also possible to roughly assign the order-1 results directly to scalar fields in $\mathbb{R}^3$. These scalar fields can be evaluated on a rectilinear grid, and then visualized with voxel rendering techniques. The key advantage of the voxel rendering technique is for the electronic response terms, which

are assigned on the level of detail of individual orbitals,

$$\epsilon_{\text{int}}^{\mathbb{A}}(\vec{r}) = E_{\text{int}}^{\bar{a}}|\phi_{\bar{a}}(\vec{r})|^2 = \underbrace{C_{\mu}^{\bar{a}} E_{\text{int}}^{\bar{a}} C_{\mu'}^{\bar{a}}}_{D_{\text{int}}^{\mathbb{A}}} \phi_{\mu}(\vec{r})\phi_{\mu'}(\vec{r}). \tag{266}$$

We have found that this analysis provides useful insight into the origins of electronic response in exchange, induction, and dispersion, e.g., showing the breakdown of the response between σ-framework, π-framework, and lone-pair contributions. The results seem to be particularly intuitive if chemically relevant Pipek-Mezey orbitals, which usually preserve σ/π separation, are employed. For electrostatic source terms, the scalar field assignment is accomplished by weighting the discrete atomic order-1 result by the weight function of atom **A** in physical space,

$$\epsilon_{\text{int}}^{\mathbb{A}}(\vec{r}) = E_{\text{int}}^{\mathbf{A}} w_{\mathbf{A}}(\vec{r}) = E_{\text{int}}^{\mathbf{A}} \frac{\rho_{\bar{A}}(\vec{r})}{\rho_{\mathcal{A}}(\vec{r})}. \tag{267}$$

This treatment usually provides equivalent intuition as the direct atomic visualization technique.

As demonstrated below, these order-1 analyses are quite useful in extracting a qualitative picture of the intermolecular interaction from the A-SAPT partition. However, the summations inherent in the order-1 partition necessarily lose some of the specific information contained in the order-2 results. It is particularly important to realize that if two moeities (one each on monomer A and monomer B) are in close contact and both are active in the order-1 analysis, this might just be a coincidence. On the converse, a moeity in monomer A that is inactive in the order-1 analysis might have significant but cancelling contributions from two or more moeities in monomer B. Explicit consideration of the order-2 analysis can confirm or deny the presence or absence of an interaction predicted by consideration of the order-1 analysis. In general, we have found that moieties in close contact and with active order-1 contributions are very likely to have a significant order-2 pairwise interaction, but an explicit check of the order-2 partition is important to confirm this suspicion.

We are also in the process of developing an alternative order-2 partition to the level of chemical functional groups, using the two-body diagrammatic partition developed above.[220] Such a partition would combine the detailed information of an order-2 partition with roughly the level of data compression of an order-1 atomic visualization, and would provide an important complementary alternative to order-1 atomic analyses. Moreover, we emphasize that any use of A-SAPT in such applications as force-field parametrizations will necessarily use the full order-2 atom-pair data: the order-1 techniques developed here are solely for the purposes of qualitative chemical understanding.

Below, atomic visualizations are performed using the student version of MacPyMol. Voxel visualizations are performed using the open source ImageVis3D raycasting package developed at the Scientific Computing and Imaging Institute at the University of Utah.[77]

8.4.4 Example Order-1 A-SAPT0 Results: Adenine-Thymine Dimer

As a representative example of the order-1 results obtainable by A-SAPT0, in Figures 27 and 28 we depict the order-1 atomic and voxel visualizations for the hydrogen-bonded adenine-thymine complex, as taken from the S22 database.[143] The order-2 atomic data for this system are additionally available in the supplemental material. In all of our visualizations of interaction energy contributions, red indicates an attractive contribution to the interaction energy, blue indicates a repulsive contribution to the interaction energy. The saturation of the color indicates the relative magnitude of the contribution. For voxel visualizations, the color scales are the same for all systems and terms presented in this work. For atomic visualizations, the color scales are the same for all systems and terms presented in this work, with exception of electrostatics, which uses a larger scale than the other terms to prevent excessive saturation. Moreover, we use different scales for electrostatics for each monomer for

the proflavine intercalator example below, due to the presence of a charged species within this complex.

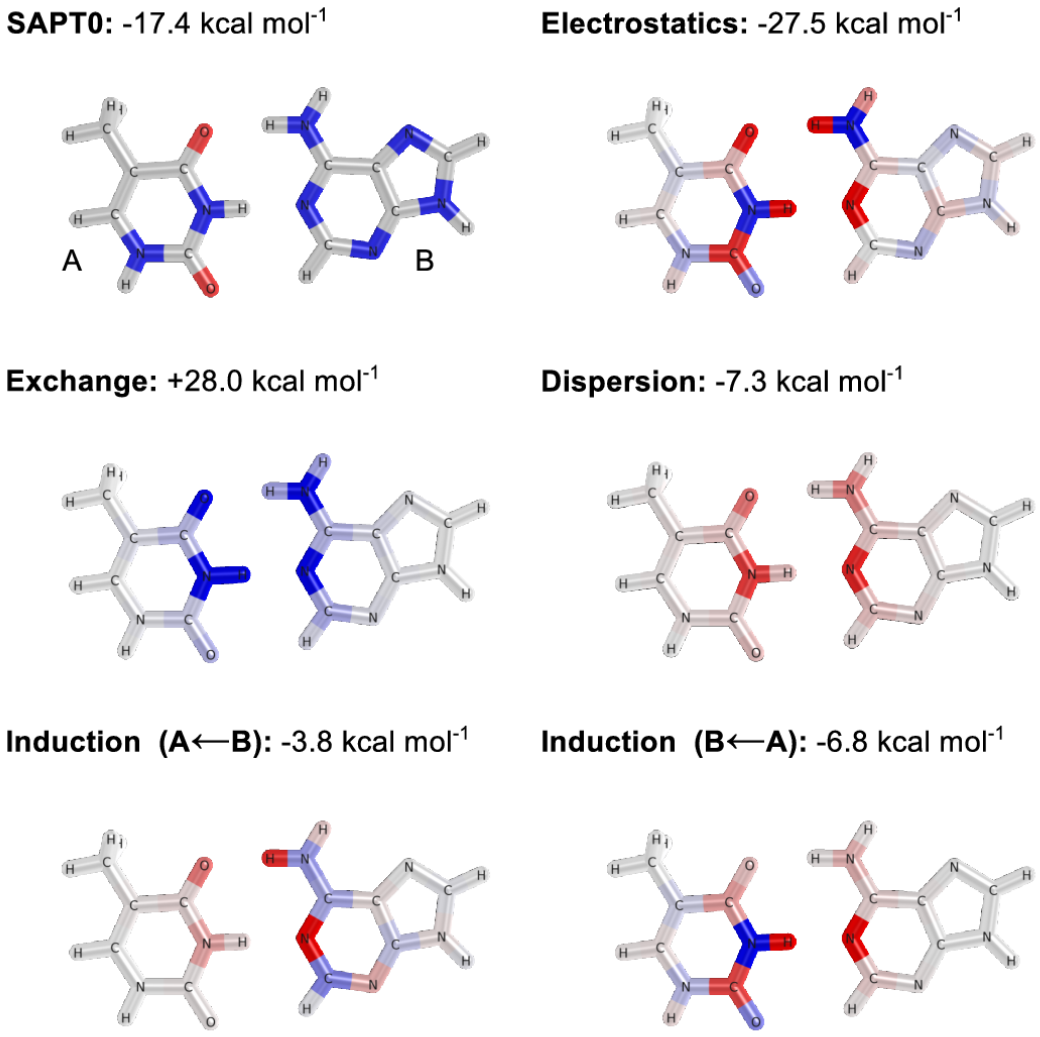

Figure 27: Order-1 atomic A-SAPT0/jun-cc-pVDZ visualization for hydrogen-bonded adenine-thymine (S22-7). Red indicates an attractive contribution, blue indicates a repulsive contribution. The colormap for electrostatics uses a larger range than the colormap for exchange, induction, and dispersion.

The adenine-thymine case demonstrates how A-SAPT mirrors and amplifies our chemical understanding of noncovalent interactions. Considering the atomic visualizations first, the quite attractive electrostatics term shows the two hydrogen bonds

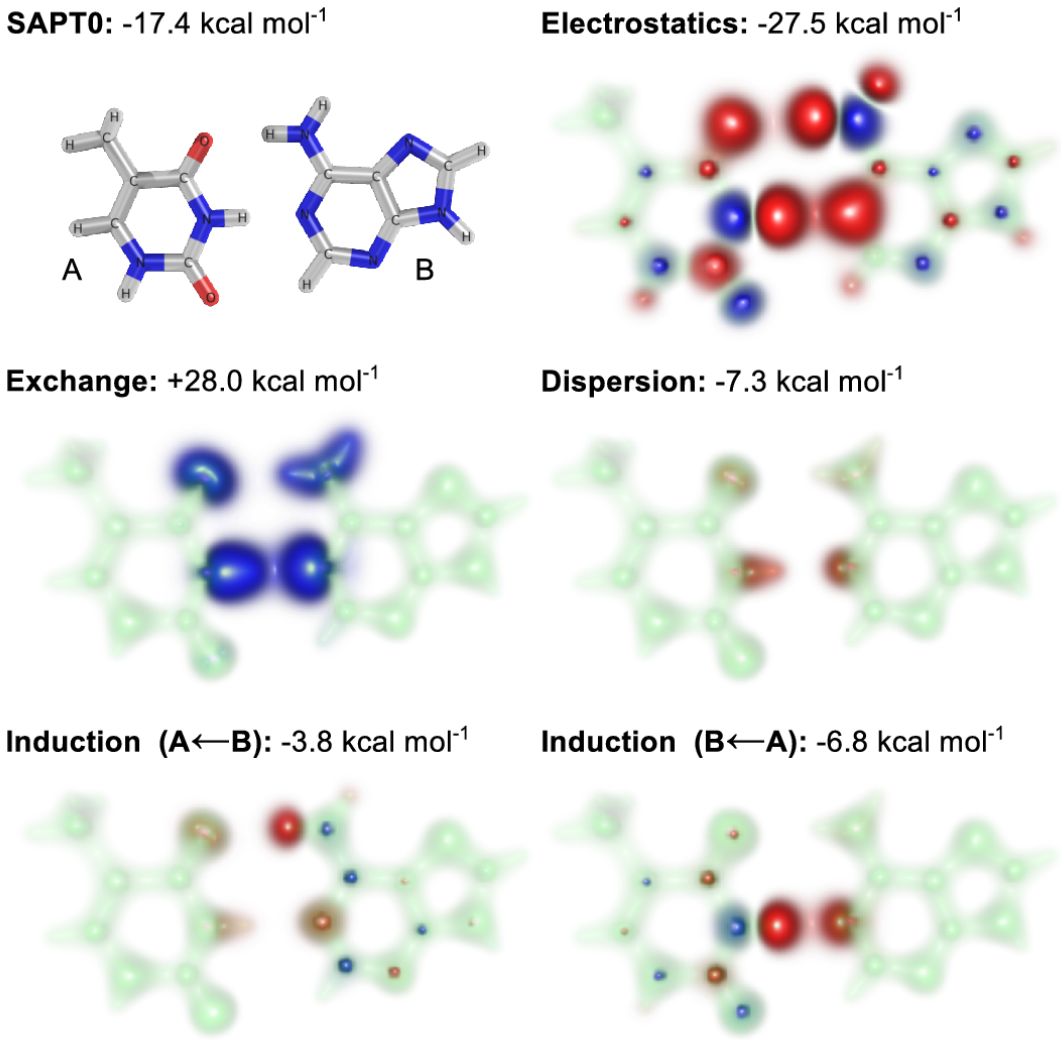

Figure 28: Order-1 voxel A-SAPT0/jun-cc-pVDZ visualization for hydrogen-bonded adenine-thymine (S22-7). Red indicates an attractive contribution, blue indicates a repulsive contribution, green is the skeleton of the molecular density (provided for visual reference). All colormaps are identical.

present in this complex: the blue-red$\cdots$red colorations of the amino$\cdots$carbonyl contact and amino$\cdots$nitrogen heteroatom contact are indicative of net favorable dipole–charge interactions between these contacts. Explicit consideration of the order-2 atomic partition in the supplemental material confirms this indication: though the atom-pairwise electrostatics contributions are generally nontrivial in magnitude due

to pairwise interactions between partial charges, the amino$\cdots$carbonyl contact and amino$\cdots$nitrogen contact are the largest contributors. The third C-H$\cdots$carbonyl contact does not light up in color, indicating that this contact does not contribute to the overall electrostatic interaction nearly as favorably as the two hydrogen bonds, as might be expected. The surprising finding is the net attractive carbon in the lower carbonyl of the thymine: this is due to the presence of three electron withdrawing atoms bound to this carbon, producing a strong partial positive charge, followed by favorable interactions between this partial positive charge and the excess electronic charge on the left-most nitrogen heteroatom of the adenine. This is an example of A-SAPT amplifying our chemical intuition by pointing to a non-obvious process in a noncovalent interaction. The other terms largely follow chemical intuition. The exchange is very short-ranged, and involves primarily the amino$\cdots$carbonyl and amino$\cdots$nitrogen heteroatom contacts responsible for the two hydrogen bonds. Dispersion shows a similar topology, though is much weaker in magnitude than exchange for this system.

A note is appropriate here regarding interpretation of the induction visualizations. For a canonical example, consider the induction A$\leftarrow$B term in Figure 27, corresponding to the adenine polarizing the thymine. In this case, the adenine lights up with both attractive and repulsive contributions which denote the portions of the adenine which drive the polarization of the thymine. By contrast, the thymine lights up with only attractive contributions, corresponding to the portions of the thymine which undergo polarization to relax the electrostatic interaction, leading to the always attractive induction interaction. For example, in the amino group$\cdots$carbonyl group interaction, the amino hydrogen (partial positive) pulls the carbonyl oxygen lone pairs to the right, while the amino nitrogen (partial negative) pushes the carbonyl oxygen lone pairs to the left. The net induction is to the right, which is favorable in the new hydrogen$\cdots$oxygen electrostatic interaction, and unfavorable in the new nitrogen$\cdots$oxygen electrostatic interaction. In the end (as is always the case), the

total induction event is net attractive, as seen in the slight red coloring of the carbonyl oxygen, and the greater attractive magnitude of the amino hydrogen vs. the repulsive magnitude of the amino carbon. Similar motifs are observed for other contacts in this visualization, and for the visualization of the induction B←A term, where the thymine polarizes the adenine.

For this system, the induction term for adenine polarizing thymine is weak (-3.8 kcal mol^{-1}), and primarily involves the amino group in the adenine polarizing the carbonyl in the thymine, and (to somewhat less extent) the nitrogen heteroatom of the adenine polarizing the amino group in the thymine. The induction term for the thymine polarizing the adenine is stronger (-6.8 kcal mol^{-1}), with the bulk of the effect attributed to the thymine amino group and lower carbonyl polarizing the nitrogen heteroatom of the adenine, with the thymine upper carbonyl polarizing the adenine amino group providing a smaller contribution.

The voxel results show roughly the same picture as the atomic results, but can provide additional detail for the origins of electronic response terms. For instance, the exchange and dispersion interactions clearly involve the amino σ orbitals and carbonyl lone pairs in the amino$\cdots$carbonyl contact at the top, and the amino σ orbitals and nitrogen heteroatom lone pair in the amino$\cdots$nitrogen heteroatom contact at the middle. This illustrates that the darker coloring of the nitrogens in the amino groups for atomic visualization of dispersion and exchange is somewhat of a visual artifact: in truth, the σ orbital is contributing strongly to the interaction, but is assigned primarily to the nitrogen atom via the ISA orbital atomic charges.

Additional atomic visualizations for the entire S22 database are available in the supplemental file, along with comparison views for other choices of parameters detailed below.

8.4.5 Parameter Selection and Sensitivity Analysis

To close the definition of A-SAPT0, we must specify the ISA grid and convergence procedure, the choice of orbital localization metric, and the treatment of coupling in the induction terms. This was accomplished by performing a limited parameter search on the hydrogen-bonded adenine-thymine dimer (S22-7) to identify a set of reasonable parameters, followed by a sensitivity study across the S22 database to determine the influence of the selected parameter set on the A-SAPT0 partitions and visualizations. The salient features are presented below, with the explicit data provided in the supplemental file.

8.4.5.1 ISA Grid and Convergence Selection

The recommended ISA grid is (302,100), converged to $\epsilon < 10^{-7}$, using the approximate Newton-Raphson update in every third step, with the Newton-Raphson update disabled after 1000 iterations. This procedure was able to converge all A-SAPT0 computations presented in this work, and produces deviations across the S22 database of less than 1 kcal mol^{-1} per atom pair for $E_{\text{elst}}^{\mathbf{AB}}$ and less than 0.1 kcal mol^{-1} per atom for $E_{\text{elst}}^{\mathbf{A}}$ and $E_{\text{elst}}^{\mathbf{B}}$ vs. results obtained in a (590,150) grid converged to $\epsilon < 10^{-9}$ using the same convergence procedure. These deviations are on the order of 1% per element of the order-2 or order-1 partitions, and produce order-1 visualizations which are indistinguishable to the naked eye. Deviations for terms other than electrostatics (either due to electrostatic source or local orbital atomic charges) are even smaller.

Without the approximate Newton-Raphson update, many choices of ISA grid and convergence criteria were unable to be converged within 2000 iterations. Moreover, the order-1 and order-2 partitions obtained without approximate Newton-Raphson acceleration showed substantially larger deviations from the reference than their accelerated counterparts, indicating that exponential slowing down of the non-accelerated procedure may be inducing premature termination, before the ISA atomic densities

are converged.

Though the accelerated ISA procedure sometimes requires up to just over 1000 iterations to converge, the ISA procedure is only quadratic scaling with the system size, so the computational burden of this portion of the A-SAPT0 analysis is not rate-limiting. Additionally, the ISA objective functional is universally convex,[179] indicating that there are no concerns about uniqueness of the solution.

8.4.5.2 Localization Metric Selection

The recommended localization metric is Pipek-Mezey. Deviations between the A-SAPT0 partitions obtained with Pipek-Mezey and Boys procedures are substantially larger than those due to the ISA grid and convergence procedure, with maximum deviations in the S22 database on the order of 1 kcal mol^{-1} per atom or atom-pair entry for both order-1 and order-2 exchange partitions. This deviation is on the order of 35% at worst. Deviations of the same order or smaller are observed for the other terms involving electronic response.

The primary source of these deviations is the differential treatment of occupied π electrons between the two metrics. Pipek-Mezey maintains σ and π separation, while Boys hybridizes the σ and π orbitals into τ orbitals. Figures 29 and 30 depict the differences in A-SAPT0 order-1 exchange and dispersion visualizations for the π-stacked indole-benzene complex (S22-14), which exhibits the largest relative deviations for these two terms. The voxel visualizations clearly show the effect of σ/π vs. τ orbital partition. This deviation is not what we consider to be an error, but rather is a form of maximum resolution of the A-SAPT0 analysis. That multiple localization metrics exist is indicative of the fact that assignment of occupied quasiparticles to localized regions of physical space is a somewhat ambiguous business. The good news is that the atomic visualizations are qualitatively quite similar to the eye, regardless of the choice of localization metric. To provide for a consistent definition for A-SAPT,

we recommend the use of the more chemically intuitive Pipek-Mezey orbitals, but we note that this choice over Boys orbitals is a matter of small increments rather than anything substantial.

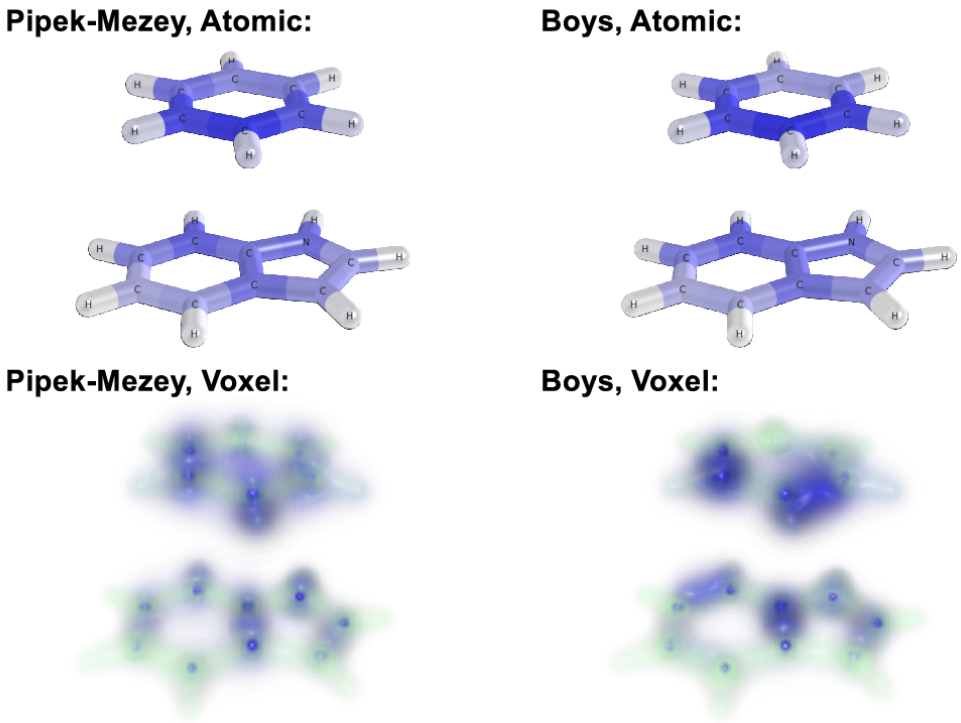

Figure 29: Comparison of order-1 A-SAPT0/jun-cc-pVDZ visualizations with Pipek-Mezey vs. Boys local orbitals for the exchange term of stacked indole-benzene (S22-14). Red indicates an attractive contribution, blue indicates a repulsive contribution, green is the skeleton of the molecular density (provided for visual reference). The colormaps are the same in the left and right panels.

8.4.5.3 Induction Coupling Selection

The recommended treatment for induction is uncoupled but scaled. Deviations between the A-SAPT0 partitions obtained with scaled uncoupled and coupled treatments exhibit maximum deviations in the S22 database on the order of 1 kcal mol^{-1} per atom or atom-pair entry for both order-1 and order-2 exchange partitions. This deviation is on the order of 30% at worst.

Figure 31 depicts the differences in the A-SAPT0 order-1 induction visualizations

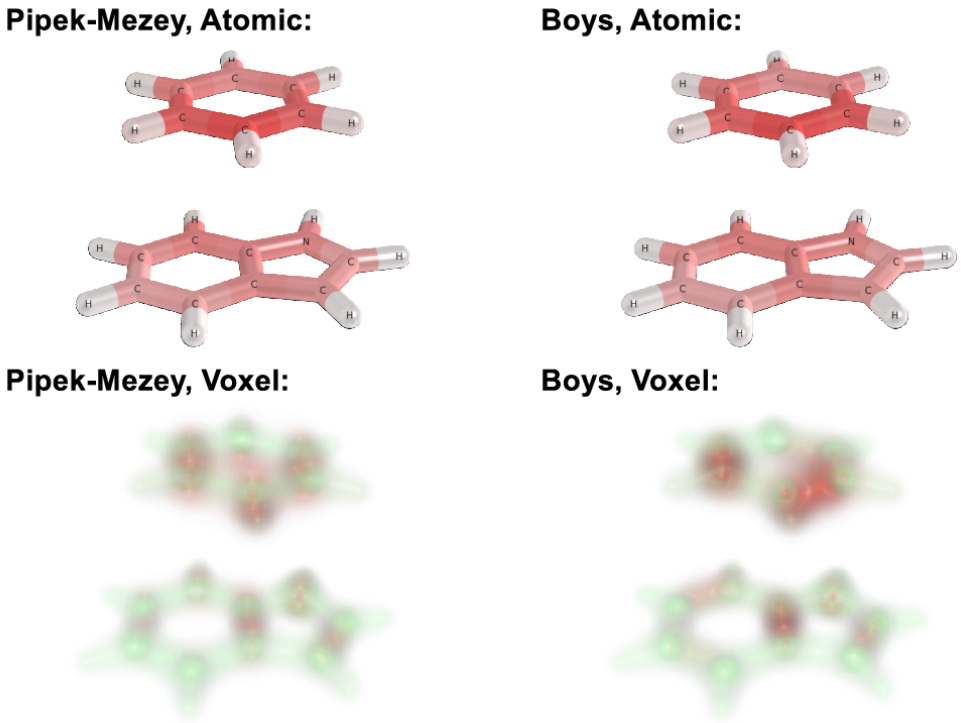

Figure 30: Comparison of order-1 A-SAPT0/jun-cc-pVDZ visualizations with Pipek-Mezey vs. Boys local orbitals for the dispersion term of stacked indole-benzene (S22-14). Red indicates an attractive contribution, blue indicates a repulsive contribution, green is the skeleton of the molecular density (provided for visual reference). The colormaps are the same in the left and right panels.

for the hydrogen-bonded adenine-thymine complex (S22-7), which exhibits the largest absolute deviation for these terms (i.e., this example has the *worst case* error for coupled induction scaling across the S22 database). The term shown is the direction in which the thymine polarizes the adenine. Visually, the results are qualitatively quite similar, with small quantitative deviations. The most obvious discrepancies are the color saturations of the carbonyl groups of the thymine. The relative magnitudes of the contributions of neighboring atoms in the source monomer are slightly altered but the overall picture of the induction event remains unchanged. Moreover, we would expect similar or larger effects to arise from consideration of the hybrid $\delta^{(2)}_{\mathrm{HF,r}}$ term, which we cannot currently account for beyond simple scaling. Due to these

considerations and the substantial computational penalty for performing a coupled induction computation for each atomic source, we advocate the uncoupled but scaled approach for production-level A-SAPT0 analysis.

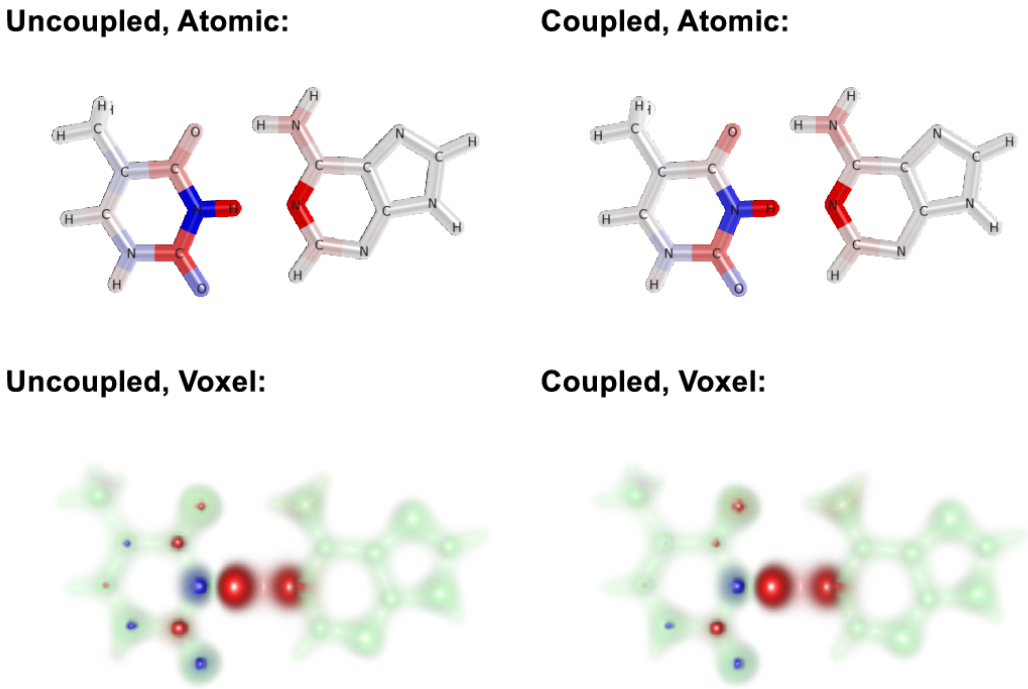

Figure 31: Comparison of order-1 A-SAPT0/jun-cc-pVDZ visualizations with uncoupled vs. coupled treatments of induction for the induction (B←A) term of hydrogen-bonded adenine-thymine (S22-7). Red indicates an attractive contribution, blue indicates a repulsive contribution, green is the skeleton of the molecular density (provided for visual reference). The colormaps are the same in the left and right panels.

8.4.6 Application to Side-On Cation-π Interactions

As an example of the level of detail provided by A-SAPT0, we consider the case of side-on cation-π interactions. In a previous study,[186] we obtained the unexpected finding that side-on cation-π interactions are energetically favorable. SAPT0 analysis (but not A-SAPT0) for benzene-Na^+ indicated that while the electrostatic and exchange terms were unfavorable, a substantial favorable induction interaction is sufficient to

stabilize the complex. At the time, it was hypothesized that this favorable induction effect was primarily the result of polarization of the π electrons of the benzene ring toward the cation. Here we revisit this study with A-SAPT0, and additionally consider cyclohexane-Na^+.

The geometries are obtained by B3LYP/aug-cc-pVDZ optimization of the benzene and cyclohexane (chair) monomers, followed by placement of the Na^+ in the ring plane, 4.0 Å from the ring center and along the line bisecting the nearest C–C bond. This treatment places the cation near the minimum energy distance for benzene-Na^+ obtained in the previous study, and ensures that the nearest C–Na^+ and H–Na^+ distances are the same to within a few thousandths of one Å for both benzene-Na^+ and cyclohexane-Na^+, providing a reasonably equal geometric footing for both complexes. The geometries are available in the supplement, along with A-SAPT0 atomic rank-1 visualizations. A-SAPT0 computations were performed with jun-cc-pVDZ and corresponding auxiliary basis sets on the hydrocarbon rings and cc-pVTZ and corresponding auxiliary basis sets on the Na^+ (for which the jun-cc-pVDZ auxiliary basis sets are not available). Additionally, frozen core is disabled for this study.

Quantitative A-SAPT0 results of this analysis are available in Table 5. For the benzene-Na^+, electrostatics and exchange are both unfavorable, while dispersion and the induction term involving polarization of the cation are both neglible. The major stabilization is provided by the induction term involving polarization of the benzene, with a value of -11.22 kcal mol^{-1}, leading to an overall favorable interaction energy of -3.31 kcal mol^{-1}. To this point, these results are equivalent to those of the previous study. However, considering cyclohexane-Na^+, the results do not match our previous hypothesis regarding the importance of π electrons: the side-on cyclohexane-Na^+ interaction is more favorable than the benzene-Na^+ interaction, with total interaction energies of -8.02 kcal mol^{-1} vs. -3.31 kcal mol^{-1}. In both cases, the exchange,

induction, and dispersion terms are all roughly the same, but the cyclohexane-Na^+ electrostatics term is much more favorable than the benzene-Na^+ electrostatics term, with values of -1.39 kcal mol^{-1} vs. +4.80 kcal mol^{-1}, respectively. The favorable cyclohexane-Na^+ electrostatics are attributed to less polar C–H bonds in cyclohexane vs. benzene, plus favorable charge penetration contributions.

Table 5: A-SAPT0/jun-cc-pVDZ analysis of side-on cation-π interactions (cc-pVTZ on Na^+). Energies in kcal mol^{-1}.

Term	Benzene-Na^+	Cyclohexane-Na^+
E_{elst}	+4.80	-1.39
E_{exch}	+3.24	+4.32
$E_{\text{ind}}[\mathbb{A} \leftarrow \mathbb{B}]$	-11.22	-10.82
$E_{\text{ind}}[\mathbb{A} \leftarrow \mathbb{B}]$ Core	-0.01	-0.01
$E_{\text{ind}}[\mathbb{A} \leftarrow \mathbb{B}]$ C–C σ	-2.25	-3.45
$E_{\text{ind}}[\mathbb{A} \leftarrow \mathbb{B}]$ C–H σ	-4.37	-7.35
$E_{\text{ind}}[\mathbb{A} \leftarrow \mathbb{B}]$ C–C π	-4.59	–
$E_{\text{ind}}[\mathbb{B} \leftarrow \mathbb{A}]$	-0.00	-0.00
E_{disp}	-0.13	-0.14
E_{int}	-3.31	-8.02

Another interesting finding is that the induction terms corresponding to polarization of the hydrocarbon rings are nearly identical between benzene-Na^+ and cyclohexane-Na^+, with values of -11.22 kcal mol^{-1} and -10.82 kcal mol^{-1}, respectively. This suggests that the π electrons in the benzene are not the only contributors to the induction event. A-SAPT0 analysis shows that C–C π, C–H σ, and C–C σ electrons all contribute significantly to the induction event in benzene-Na^+, with contributions of -4.59 kcal mol^{-1}, -4.37 kcal mol^{-1}, and -2.25 kcal mol^{-1}, respectively. In cyclohexane-Na^+, there are more C–H σ bonds, and the C–H and C–C σ electrons are less tightly held than in benzene, leading to larger contributions from the σ framework than in benzene-Na^+. These larger σ contributions are roughly the same size as the contributions from the π electrons in benzene (missing in cyclohexane),

resulting in comparable induction energies between the two systems. This quantitative analysis is directly apparent in the A-SAPT0 order-1 voxel visualizations for induction for these two systems, depicted in Figure 32. It is therefore evident that π electrons are not the whole story in side-on cation-π interactions: significant C–H and C–C σ contributions are also present. Additionally, saturated hydrocarbon rings may be even more favorable in this interaction motif, due to additional and stronger σ framework contributions to induction, and better electrostatic interactions.

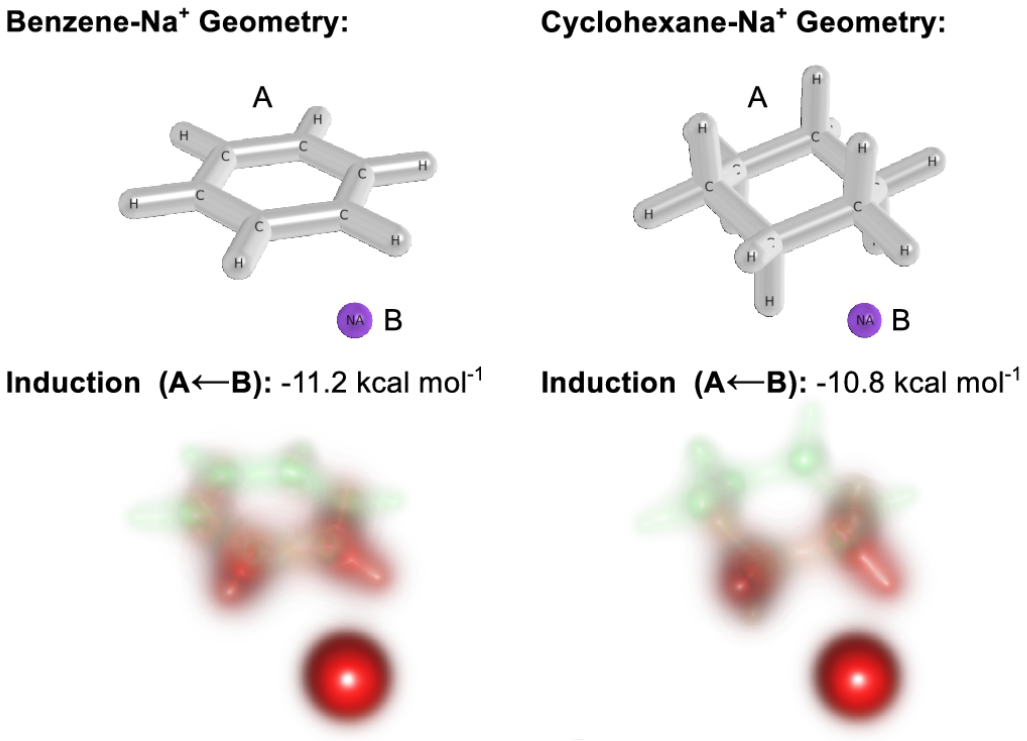

Figure 32: Order-1 voxel A-SAPT0/jun-cc-pVDZ visualizations for the major induction term in side-on cation-π interactions (cc-pVTZ on Na^+) Red indicates an attractive contribution, blue indicates a repulsive contribution, green is the skeleton of the molecular density (provided for visual reference). The colormaps are the same in the left and right panels.

8.4.7 Application to DNA-Proflavine Intercalation

As an example of the new insights and tractability afforded by A-SAPT0, we have applied the method to study the intercalation of DNA by proflavine. The A-SAPT0/jun-cc-pVDZ analysis for proflavine intercalating into a CGA step of DNA is depicted in Figure 33. This is the same case study as was used in our 2011 DF-SAPT0 paper.[118] The geometry was obtained from a crystal structure,[184] with hydrogen positions determined by constrained B3LYP/6-31G** optimizations. Each phosphate group was protonated to simulate the presence of a counterion (e.g., Na^+ or K^+), resulting in the DNA fragment obtaining a new neutral charge. The proflavine carries a net +1 charge.

In our previous study, we performed SAPT0 analysis for the total complex, as well as various fragmentations of the DNA segment to elucidate the direct interactions between the proflavine and the base-pair steps and backbone fragments. This fragmentation study yielded three conclusions: (1) roughly two-thirds of the interaction energy arises from π-stacking interactions between the proflavine and the neighboring base-pair steps (2) the next-nearest neighbor base-pair step is not significantly involved in the interaction and (3) roughly one-third of the interaction energy arises from long hydrogen bonding interactions between the proflavine and one of the two backbones (the backbone on the left in Figure 33). Using a single A-SAPT0 computation, we are now able to obtain these three conclusions without laborious fragmentation studies and additional SAPT0 computations. The first two conclusions are borne out in considerations of the exchange and dispersion A-SAPT0 plots: these two terms clearly show strong π-stacking interactions between the proflavine and neighboring base-pair steps, with minimal involvement of the backbones, and essentially zero involvement of the next-nearest neighbor base-pair step. The third conclusion is also visible, though two caveats should be discussed beforehand. The first caveat is that the proflavine species carries a net positive charge, implying that any partial

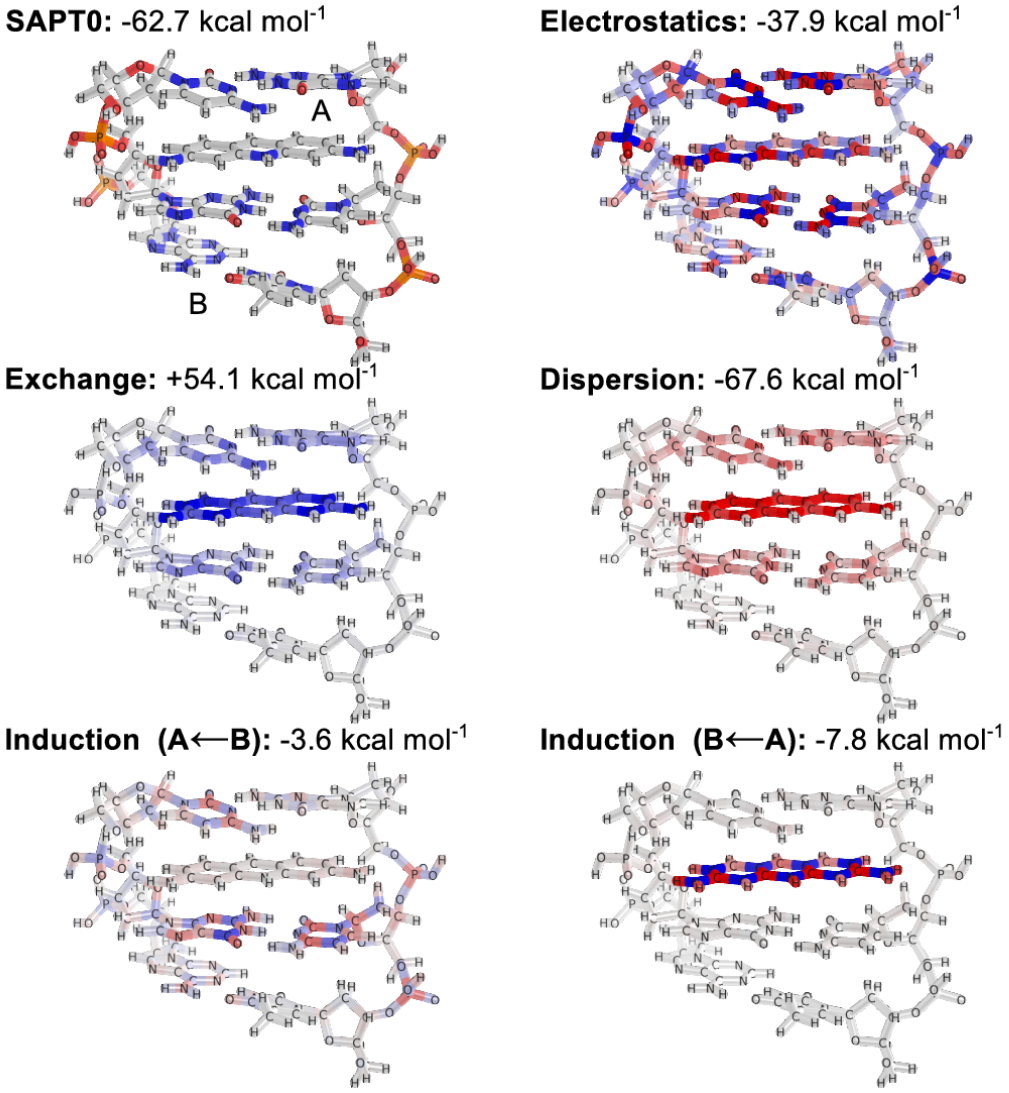

Figure 33: Order-1 atomic A-SAPT0/jun-cc-pVDZ visualization for the proflavine - DNA intercalation complex. Red indicates an attractive contribution, blue indicates a repulsive contribution. For the electrostatics only (1) a larger range in the colormap is used than for exchange, induction and dispersion and (2) a larger range in the colormap is used for monomer B than for monomer A.

charge in the DNA will light up in response to the net charge, with most of these responses canceling. To partially ameliorate this effect, we have applied a much larger color scale for the coloring of the order-1 electrostatics of the DNA atoms, and yet we still see extensive illumination of the DNA fragment, most of which will cancel

in the net. Therefore, from the perspective of the DNA, we should focus on areas of greater relative intensity. The second caveat involves the charges in the aromatic carbon network of the proflavine: it seems that withdrawal of electronic density from the carbons attached to the terminal amino groups sets up a pattern of alternating charges in the carbons in the leading edge of the proflavine (and, to a lesser extent, in the carbons in the trailing edge of the proflavine). These oscillating charges result in substantial alternating electrostatic contributions to the proflavine atoms, most of which cancels in the net. We have verified that these charge oscillations are not particular to the choice of ISA atomic density fields: charge alternations of the same topology also occur with Natural Population Analysis (NPA),[243] albeit with lower overall magnitudes (see the supplemental material for further analysis). In the end however, it is clear that the left end of the proflavine lights up far more than the right end, and that the three leftmost hydrogens (two on the amino group and one attached to the neighboring aryl carbon) are quite attractive, relative to their counterparts on the right end of the proflavine. Additionally, the neighboring phosphate lights up much more than the other three phosphates, and the nearest oxygens in the phosphate are particularly attractive. These features all point to the existence of the hydrogen bonding interaction between the left end of the proflavine and the neighboring phosphate group in the backbone. The difficulties of visualizing the electrostatics in systems with either net charges or substantial charge oscillations are outstanding challenges to be resolved in the next iteration of A-SAPT theory. Even considering these difficulties, the method is already extremely compelling in its power to directly elucidate the nature of large-scale noncovalent interactions.

Figure 34 depicts the explicit wall time breakdown for various SAPT0 and A-SAPT0 implementations for the proflavine intercalation complex, in the jun-cc-pVDZ basis. All timings were performed using a single node with a six-core Intel i7-3930K overclocked to 3.9 GHz on an ASUS P9X79 Pro motherboard, 64 GB of memory,

and 3× 3 TB 7200 rpm scratch drives in a software RAID0 configuration. Our 2011 and 2013 implementations of DF-SAPT0 require roughly the same wall times (69.9 hours and 64.1 hours, respectively), though the underlying algorithms are quite different. SAPT0 ('13) is actually slightly faster ($\sim$ 8%), due to the reuse of the three-index integrals in the -JKFIT auxiliary basis between the SCF computations and the Fock-like terms of SAPT0, and the new DFMP2-like formulation of the exchange-dispersion term. A-SAPT0 ('13) is built from SAPT0 ('13), and inherits most of the former's performance. The first deviation is in the dispersion term, which requires slightly longer ($\sim$ 11%) due to the requirement of driving the computation over $N_r N_s$ contractions of $\mathcal{O}(N_a N_b N_Q)$ effort in A-SAPT0 ('13), vs. $N_a N_b$ contractions of $\mathcal{O}(N_r N_s N_Q)$ effort in SAPT0 ('13). The latter involves a larger DGEMM, and is more efficient. The A-SAPT0 dispersion term also involves a small overhead from the canonical-to-local transformation of the amplitudes and potentials. Additionally, A-SAPT0 ('13) involves many additional steps beyond SAPT0 ('13). The worst of these is ISA, which requires an additional 12.7 hours of wall time for this system, using the convergence acceleration techniques discussed above. The atomic density fitting, atomic pseudospectral, additional atomic Fock terms, and other overhead contribute roughly 7.7 additional hours of wall time, resulting in A-SAPT0 ('13) requiring $\sim$ 37% additional effort over SAPT0 ('13), for a total wall time of 87.5 hours. This is remarkably small overhead considering the additional chemical insight provided by the A-SAPT partition: in particular, A-SAPT0 is tractable for any system for which SAPT0 is tractable. It should be noted that the high-prefactor but low-scaling steps present in A-SAPT0 produce considerably more relative overhead for smaller systems (e.g., $\sim$ 200% vs. $\sim$ 37% above), but the principal concern of tractability is not an issue for A-SAPT0.

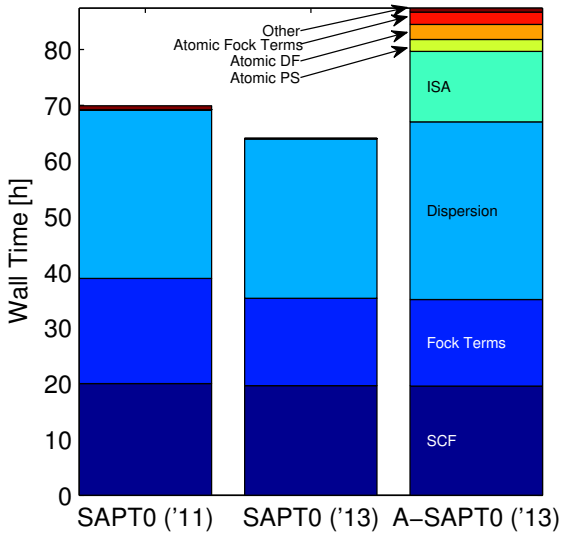

Figure 34: Wall time breakdown for various SAPT0 and A-SAPT0 implementations for the proflavine intercalation complex, in the jun-cc-pVDZ basis. SAPT0 ('11) is the DF-SAPT0 code of our 2011 paper. SAPT0 ('13) is our reference DF-SAPT0 code built on `LibFock` and `LibPANACHE`. A-SAPT0 ('13) is our DF-A-SAPT0 code, an extension of SAPT0 ('13).

8.5 Conclusions

The A-SAPT0 partition is introduced as a physically motivated scheme for the visualization of intermolecular interactions. The method is based on the partition of SAPT0 intermolecular interaction energy terms to spatially localized components. We caution the reader that any such partitioning scheme is neither a physical observable nor universally unique, but we have selected the A-SAPT partition in a way that corresponds strongly to the chemical explanation of each interaction term, while respecting the many-body SAPT limit for the interaction energy. First, an effective two-particle interaction is extracted from the diagram for each SAPT term. Next, an appropriate set of local quasiparticles is introduced to close the local order-2 partition. For terms involving the electrostatic potentials of the monomers, atomic density fields obtained from Wheatley's iterative stockholder analysis are the relevant quasiparticles. For terms involving the response of the monomer electronic density fields,

local occupied orbitals are the relevant quasiparticles. Local orbital partitions can additionally be projected onto atomic partitions via orbital atomic charges. A critical technical development in the method is the use of a doppelgänger of tensor hypercontraction to project the grid-based atomic charge densities onto a spectral auxiliary basis, to avoid problems with the Coulomb singularity and provide for an accurate, efficient treatment of the electrostatic and induction interactions. A density fitted implementation of A-SAPT0 is developed to lessen the computation cost of the method. Additionally, an approximate Newton-Raphson scheme is developed to accelerate the ill-conditioned ISA portion of the computation. Order-1 visualization schemes based on atomic and voxel assignments of the A-SAPT partition are developed to facilitate rapid and intuitive analysis of the A-SAPT data.

A-SAPT0 is found to mirror and amplify chemical intuition regarding noncovalent interactions. Important qualitative features such as hydrogen bonding and π-stacking interactions are correctly reproduced, while more subtle features of particular interaction motifs are illuminated. The sensitivity of the method to the choices of several important parameters is probed, and the results are found to be qualitatively robust against such choices. The recommended ISA grid and convergence procedure is found to be quite well converged against reference computations. The choice of orbital localization metric and coupling treatment in the A-SAPT0 induction terms are shown to produce larger deviations, but the qualitative utility of the partition and visualization is unaffected. The practical use of the A-SAPT0 partition is demonstrated in studies of side-on cation-π interactions and of the intercalation of a CGA DNA fragment by proflavine. In the former study, surprising contributions to the induction energy from the σ framework are found, overturning the previous hypothesis that favorable side-on cation-π interactions are due to the presence of π electrons. In the latter study, a single A-SAPT0 computation (on a system with 220 atoms and 2845 basis functions) is able to clearly identify three important qualitative conclusions which were

previously attainable only through a laborious fragmentation study. Additionally, this A-SAPT0 computation is shown to take only $\sim 37\%$ longer than the underlying SAPT0 computation, indicating that A-SAPT0 has the same tractability limit as SAPT0.

This introduction of A-SAPT methodology suggests several further technical studies to improve the robustness, accuracy, and utility of the method. Within A-SAPT0, the sensitivity of the result to the choice of ISA atomic density fields should certainly be investigated. It is possible that another atomic charge model such as the Gaussian ISA variant,[307] Hirshfeld-I,[39] or Atoms-In-Molecules[14,15] might provide more universal or intuitive visualizations than ISA. Additionally, more advanced visualization techniques should be developed to aid in the understanding of the electrostatics terms in systems with net charges, which exhibit physically correct but distracting electrostatic contributions of opposite signs, which largely cancel in the net. Further afield, the A-SAPT methodology should be extended to more accurate models than SAPT0. SAPT(DFT) is an obvious first choice, as the topologies of SAPT0 and SAPT(DFT) are functionally identical: such extension would essentially involve replacing Hartree-Fock quasiparticles with (coupled) Kohn-Sham quasiparticles. More advanced wavefunction-based SAPT methods, such as SAPT2+3(CCD) should also be compatible with A-SAPT, but will require the partition of more involved diagrams than SAPT0. These developments notwithstanding, A-SAPT analysis should be helpful in myriad topics ranging from the fundamental understanding of π-stacking interactions to the elucidation of intermolecular packing forces in organic crystals to the design of better catalysts and drug molecules.

CHAPTER IX

FUNCTIONAL GROUP SYMMETRY-ADAPTED PERTURBATION THEORY

The following is adapted from Ref. 220.

9.1 Abstract

Recently, we introduced an effective atom-pairwise partition of the many-body symmetry-adapted perturbation theory (SAPT) interaction energy decomposition, producing a method known as atomic SAPT (A-SAPT) [R.M. Parrish and C.D. Sherrill, *J. Chem. Phys.*, **141**, 044115 (2014)]. A-SAPT provides *ab initio* atom-pair potentials for force field development and also automatic visualizations of the spatial contributions of non-covalent interactions, but often has difficulty producing chemically useful partitions of the electrostatic energy, due to the buildup of oscillating partial charges on adjacent functional groups. In this work, we substitute chemical functional groups in place of atoms as the relevant local quasiparticles in the partition, resulting in a functional-group-pairwise partition denoted as functional-group SAPT (F-SAPT). F-SAPT assigns integral sets of local occupied electronic orbitals and protons to chemical functional groups and linking σ bonds. Link-bond contributions can be further assigned to chemical functional groups to simplify the analysis. This approach yields a SAPT partition between pairs of functional groups with integral charge (usually neutral), preventing oscillations in the electrostatic partition. F-SAPT qualitatively matches chemical intuition and the cut-and-cap fragmentation technique, but additionally yields the quantitative many-body SAPT interaction energy. The conceptual simplicity, chemical utility, and computational efficiency of F-SAPT is demonstrated

in the context of phenol dimer, proflavine^{+}-DNA intercalation, and a cucurbituril host-guest inclusion complex.

9.2 Introduction

While quantitative determination of noncovalent interactions is considered to be a critical challenge facing modern quantum chemistry, even robust qualitative chemical explanation of such interactions is difficult to attain. Consider, for just one instance, the large volume of work devoted to understanding the origins of interaction energy differences between stacked substituted benzene dimers.[128,54,276,277,249,318,248,116,316,242] Even simply producing a high-quality interaction energy for such a system is a formidable task: the extreme sensitivity of the interaction to the electronic structure of the constituent monomers and to the intermolecular orientations and distances present in the dimer mandate the use of such methods as complete basis set coupled cluster with singles, doubles and perturbative triples [CCSD(T)/CBS].[71] For more qualitative analysis, less demanding electronic structure methods will often suffice, but the qualitative assignment of an interaction energy component to a particular chemical progenitor remains a delicate art. For instance, it required many years of effort to reach the conclusion that substituent effects in stacked benzenes could be largely explained by local interactions between the substituent and nearby portion of the neighboring benzene.[318,316,242] Moreover, while this conclusion is quite useful to chemists interested in pure face-to-face stacking interactions, it does not seem to immediately transfer to slipped or T-shaped conformations of benzene dimer, or to more exotic intermolecular motifs.[11,277,249,320] At present, intuitive intermolecular chemistry is lacking the simple and broadly transferrable pictures defining covalent chemistry, i.e., there are few analogs of the guiding concepts of Lewis structures, electronegativies, bond enthalpies, frontier orbitals, etc., in intermolecular chemistry. If we cannot currently provide the applications chemist with such a rulebook with

which to design intermolecular interactions, then we should at least provide fast and accurate methods to determine why an arbitrary intermolecular motif has a particular interaction energy. To that end, the standard theoretical analysis these days is a high-quality interaction energy and possibly a symmetry adapted perturbation theory (SAPT)[139] or energy decomposition analysis (EDA)[147] partition of the noncovalent interaction into electrostatics, steric repulsion, polarization, and dispersion contributions. While such analysis can provide useful insights for small model systems,[273] this insight becomes limited as one progresses to larger systems where multiple noncovalent contacts can occur. Unfortunately, standard SAPT or EDA methods provide only a single value for each of the interaction energy components, precluding elucidation of how each intermolecular contact contributes to the total interaction energy and its components.

Recently, we have been working to produce a robust and efficient partition of the SAPT decomposition into physically-localized progenitors, i.e., to assign the interaction energy components to contributions from such chemically useful quasiparticles as protons and electrons, atoms, or functional groups. In our previous work,[222] we identified an effective two-body partition of the SAPT0 expressions[138] which both matches the chemists' intuition as to what each term represents and also respects the many-body limit of the SAPT interaction energy. In the previous approach, terms involving the response of the monomer wavefunctions (e.g., dispersion) used local occupied orbitals as the relevant quasiparticles, while terms involving electrostatic potentials (e.g., electrostatics) grouped fractions of the molecular electronic density and nuclei into atoms to prevent enormous oscillations in the partition. The terms involving local occupied orbitals could additionally be assigned to atoms via orbital atomic charges, yielding a complete spatial partition of SAPT in terms of pairs of atoms. We denoted this method atomic SAPT (A-SAPT). By automatically providing an effective atom-pairwise partition of the *ab initio* SAPT interaction energy, A-SAPT has

obvious applications in the development of next-generation molecular dynamics force fields. With respect to supporting chemical understanding of noncovalent contacts, A-SAPT can also produce automatic visualizations which clearly show the most important chemical elements in complicated intermolecular interactions. For response terms, A-SAPT can even provide both qualitative and quantitative understanding; for example, it was able to identify the origin of the induction stabilization of side-on cation-benzene and cation-cyclohexane interactions in terms of C–H σ, C–C σ, and π orbital contributions.[222] Unfortunately, as was previously reported[222] for the case of a charged proflavine$^+$-DNA intercalation complex, A-SAPT sometimes exhibits oscillating and nearly canceling contributions to the electrostatic energy, particularly in systems with charged or highly polar monomers. This complicates attempts to analyze and understand how different functional groups of atoms contribute to the electrostatic component of the interaction energy. The root of this difficulty has been identified as the very concept of atoms in molecules: σ bonds linking neighboring functional groups will necessarily be assigned preferentially to the more electronegative atom, leading to oscillating charges between the neighboring functional groups. This gives rise to large charge-charge interactions between intermolecular pairs of functional groups, obfuscating the true higher-order multipole-multipole interactions which are intuitively expected. Note that the trouble with the atoms-in-molecules picture only arises when a partition into chemically-relevant functional group contributions is attempted. For many problems, the atoms-in-molecules picture is the correct *ansatz*, e.g., as in the development of atom-pairwise force fields, for which A-SAPT is a well-posed tool.

In this work, we ameliorate the problems with respect to chemical intuition in the electrostatics term by switching from an atom-based *ansatz* to a functional-group-based *ansatz*. In the new functional-group SAPT (F-SAPT) approach, integral sets

of localized electrons and protons are assigned to chemical functional groups and linking σ bonds. All of these fragments are neutral or carry a chemically-expected net charge (e.g., the carboxylate ion functional group would carry a net charge of -1 in our analysis), preventing canceling oscillations in the electrostatic partition. To simplify the analysis, an even assignment of the link-bond contributions to their adjacent functional groups is used after the full link-bond and functional group partition is accumulated. The division of an intermolecular interaction energy decomposition into identifiable orbital or functional group contributions is somewhat reminiscent of the natural orbital EDA (NEDA) technique,[89] absolutely localized molecular orbital EDA (ALMO EDA) methodology,[146,145,13] fragment molecular orbital (FMO)[148] based pairwise interaction EDA (PIEDA),[75] or divide-and-conquer semiempirical techniques.[302] We also note a completely different but complementary approach: the use of density-based heuristics to visually indicate the presence of noncovalent interactions, including those based on electron localization functions (ELF)[21] or localized orbital locators (LOL),[275] the non-covalent interactions (NCI) scalar field,[141,50] or the recent density overlap regions indicator (DORI).[58] Finally, we note that many important developments toward spatially localizing interaction energy components to fragments of monomers have been made by Stone and co-workers, e.g., in distributed multipole analysis (DMA) of electrostatics[283,285,284,286] or distributed polarizabilities.[256,196] However, these approaches work almost wholly in terms of atomic contributions, and, as such, are much more analogous to our previous A-SAPT approach than the present F-SAPT approach.

Below, we outline the intuitive development of the F-SAPT partition by first identifying desired characteristics of the new approach and weaknesses of other partition methods such as the cut-and-cap fragmentation approach and A-SAPT. We then construct F-SAPT to satisfy all of the desired constraints, and briefly detail the straightforward modifications to the A-SAPT program needed to obtain an implementation

of F-SAPT at the level of SAPT0. As with any heuristic partition, F-SAPT can be neither derived nor proved, so we rely on practical discussion and chemical examples to justify the method. We focus on the most-difficult electrostatic term to simplify the discussion, and use phenol dimer as an illustrative example. In the results and discussion section, we apply cut-and-cap fragmentation, A-SAPT, and F-SAPT to provide functional-group pairwise partitions of phenol dimer and the proflavine$^+$-DNA intercalator (whose electrostatic oscillations under A-SAPT motivated this work). These case studies demonstrate that F-SAPT produces chemically useful partitions of the electrostatic energy while also respecting the many-body limit of the interaction energy, placing it clearly above cut-and-cap or A-SAPT methodologies. We close by discussing a complete F-SAPT analysis for an interesting cucurbituril inclusion complex.

9.3 Theory

9.3.1 Motivation: Phenol Dimer

To frame the discussion, we consider the case of phenol dimer in a hydrogen-bonded arrangement (from the S22 database),[143] depicted in Figure 35. We denote the hydrogen-bond-donating member as monomer A, and the hydrogen-bond-receiving member as monomer B. Standard SAPT0/jun-cc-pVDZ[138,124,212] analysis indicates a total interaction energy of -7.1 kcal mol^{-1}, with contributions of -9.1, $+9.9$, -0.5, -2.3, and -5.1 kcal mol^{-1} from E_{elst}, E_{exch}, $E_{\text{ind(A}\leftarrow\text{B)}}$, $E_{\text{ind(B}\leftarrow\text{A)}}$, and E_{disp}, respectively. This interaction is somewhat stronger than the -5.1 kcal mol^{-1} of water dimer (also from the S22 database), and considerably different in both geometry and interaction energy from the standard configurations of the benzene dimer. What then, are the chemical origins of this interaction? How important is the π–π interaction relative to the hydrogen bonding interaction, and are OH–π interactions relevant? Below, we will pursue several approches to answer these questions, focusing on the

critical and difficult electrostatics term.

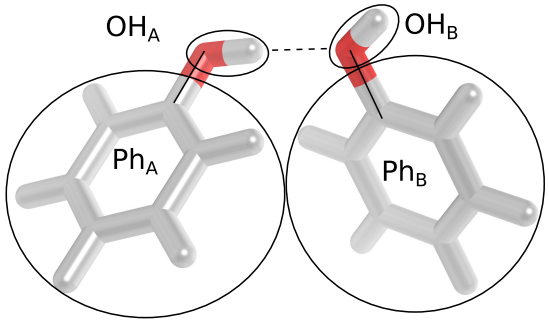

Figure 35: Phenol dimer geometry and functional group partitioning scheme. Ovals denote functional group selection. Link bonds are indicated by solid black lines. Intermolecular hydrogen bonds are indicated by dashed black lines.

To the chemist, these phenol monomers (PhOH) are intuitively defined as a phenyl group (Ph) attached to a hydroxyl group (OH) via a σ bond, as seen in Figure 35. Therefore, the most-useful partition of the $\mathrm{PhOH_A}\cdots\mathrm{PhOH_B}$ interaction would be into the pair contributions of $\mathrm{Ph_A}\cdots\mathrm{Ph_B}$, $\mathrm{Ph_A}\cdots\mathrm{OH_B}$, $\mathrm{OH_A}\cdots\mathrm{Ph_B}$, and $\mathrm{OH_A}\cdots\mathrm{OH_B}$. There are two fundamental paradigms for such partitions. The first is the family of fragmentation techniques, which *a priori* divide the monomers into relevant subsystems, and then perform isolated pairwise interaction energy computations on the subsystems. The other is the family of many-body assignment techniques, wherein a single interaction energy is computed between the two complete monomers, and the interaction energy components are *a posteriori* divided between the functional groups. A-SAPT is a member of the latter paradigm. The fragmentation approaches are compelling due to their conceptual and computational simplicity. However, the fragment approaches have three key problems that can be remedied by the many-body assignment techniques:

1. The computation of the interaction energy contributions in terms of pairs of isolated fragments ignores possibly very large intramolecular many-body coupling effects between different functional groups within the monomers. This implies

that the *ab initio* limit cannot ever be achieved for such methods, even if a convergent *ab initio* methodology such as SAPT or coupled-cluster theory is used for the isolated pairwise interaction energy computations. Note that this caveat directly applies only to explicitly pairwise interaction fragment models: higher order fragmentation models (e.g., which include trimers and perhaps even larger clusters) can approach the *ab initio* limit of the interaction energy, albeit at the cost of losing a straightforward pairwise partition of the interaction energy.

2. When covalent bonds link two functional groups, myriad problems arise in the division of the monomers into fragments. In particular, it is difficult to provide a scheme which simultaneously respects particle count, charge neutrality of adjacent functional groups, link bond polarization etc.

3. The best schemes involve cleavage of the link bonds and capping with hydrogen atoms to maintain the neutral closed-shell nature of the fragments, which can involve additional user effort to produce geometries and analyze results.

9.3.1.1 Fragmentation: Homolytic Cleavage

One particularly straightforward but ill-formed fragmentation approach is to simply cleave the link bond homolytically, assigning one link electron to the Ph group and the other to the OH group in each monomer. This effectively defines the monomers as the neutral radical species Ph· and ·OH. Pairwise interaction energy computations could then be performed between these radical species, e.g. $(\mathrm{Ph}\cdot)_\mathrm{A}\cdots(\cdot\mathrm{OH})_\mathrm{B}$. This approach has the advantages that the particle count (the total number of protons and electrons) is maintained and that the Ph· and ·OH are both neutral. However this approach is frought with peril: the electronic structures of the radical fragments are often nothing like the underlying monomer functional groups, which are stabilized by electron pairing in the link σ orbital.

9.3.1.2 Fragmentation: Heterolytic Cleavage

One possible remedy for the homolytic cleavage approach is to recognize that the pairing of electrons is vital, and to assign *both* of the electrons in the link σ bond to the more-electronegative functional group. This effectively defines the monomers as the charged closed-shell species Ph^+ and ^-OH. This approach has the advantages that the particle count is maintained and that there are no radical species involved, but has the distinct disadvantage that the fragments are charged to the maginitude of a complete electron, in an oscillating $+/-$ fashion that cancels in the net. For instance, the $^-OH_A \cdots {}^-OH_B$ electrostatic interaction is dominated by a repulsive charge-charge interaction on the order of $(-1)^2/R$, where R is the average separation between the hydroxyl groups. The contribution is a sizeable fraction of a Hartree, and completely masks the expected attractive dipole-dipole interaction that characterizes a hydrogen-bonding interaction.

9.3.1.3 Fragmentation: Cut-and-Cap

A superior approach is to homolytically cut each link bond, and then cap the unpaired electrons in each functional group with hydrogens, each of which contains a proton and an electron. This effectively defines the monomers as the neutral closed-shell species benzene (Bz) and H_2O. This approach maintains the pairing of electrons and neutrality of the functional groups, and roughly retains the anisotropy of the link bond (though the overall link bond polarity is almost always lost in the new fragments). Unfortunately, the link bonds are essentially counted twice in this approach, as the particle count is increased by two protons and two electrons for each link bond. Often, the particle count effect is assumed to be minor and ignored. Alternatively, one can subtract a hydrogen molecule placed along each link bond to maintain the particle count, e.g., defining the phenol as Bz + H_2O - H_2. This correction can significantly clutter the analysis, e.g., due to the ambiguity in assigning the H_2 contributions to

the functional groups. An additional concern is the polarization of the capped link bonds. Link bonds are usually of X–Y topology where X and Y are heavy atoms, and are much less polarized than the X–H and Y–H bonds they will be replaced with, potentially cluttering the electrostatics and induction terms. Finally, one last ambiguity arises in the placement of the protons in the capped bonds; these are often handled by constrained DFT optimizations with all nuclei but those of the capping hydrogens frozen, mandating additional computational effort.

This approach has previously been used to roughly elucidate the relative contributions of chemical functional groups to the interaction energy, e.g., in probing the hypothesis that substituent effects in benzene dimers are primarily direct, through-space effects,[318] or in illuminating the many competing contributors in proflavine$^+$-DNA intercalation.[118]

9.3.1.4 Many-Body Assignment: A-SAPT

An alternative to fragmentation is to compute the many-body interaction through a convergent *ab initio* method, while retaining a two-body quasiparticle partition identified from the many-body interaction energy diagrams. One such method is A-SAPT, which extracts the effective two-body partition from the formalism of SAPT (particularly SAPT0 in our previous work), and then uses either local occupied orbitals or iterative stockholder analysis (ISA)[178,179,315] atoms as the quasiparticle set. The invocation of the concept of atoms is critical to obtain an automatic and sensible partition of the electrostatic energy component and the electrostatic sources in the induction energy component: in this paradigm, fractions of the molecular electronic density are assigned to atoms to provide reasonable cancellation between the extremely large electronic and nuclear contributions to these terms. This atom-based assignment leads to an efficient and black-box method which produces useful visualizations identifying the active portions of complicated electrostatic interactions.

For the full technical details of A-SAPT, we refer the reader to our previous work,[222] but it is appropriate at this point to provide a self-contained description of A-SAPT, to facilitate the discussion below. The development of the A-SAPT partition is a two-stage procedure. In the first stage, the many-body interaction energy diagrams (or corresponding algebraic equations) for a given level of SAPT are examined, and an effective two-body interaction is identified for each SAPT diagram, with one body taken from each monomer. As the SAPT diagrams may contain anywhere from two to an infinite number of occupied or virtual quasiparticles (and include both electronic and nuclear contributions), this choice is not unique. However, we have found that the identification of a chemically-sensible effective two-body interaction is entirely straightforward, at least at the level of SAPT0 or DFT-SAPT. In particular, we restrict the partition to occupied quasiparticles only, summing over the virtual quasiparticles. For electrostatics and for the electrostatic sources in induction, this is an entirely rigorous choice, as no corresponding virtual quasiparticles are present. For exchange, dispersion, and the response portion of induction, the diagrams are all characterized by excitations from the occupied quasiparticles to virtual quasiparticles of one monomer, in response to some perturbation provided by the adjacent monomer. For these terms, the intuitive choice for the partition is the occupied quasiparticle from which the excitation arose, not the virtual quasiparticle to which the excitation concluded, as we are concerned with those portions of the monomers from which the interaction energy term originate. Having made the restriction to occupied quasiparticles, the choice of a specific pair of occupied quasiparticles is obvious. For instance, the dispersion contribution involves the trace of a double excitation t_{ab}^{rs} with the Coulomb integral v_{rs}^{ab}, which obviously involves only the two occupied quasiparticles a and b. The corresponding exchange-dispersion term involves additional occupied quasiparticles, but these encapsulate Pauli-forbidden collisions between the

electrons involved in the double excitation and the occupied spaces of the complementary monomers, so we continue to focus on the two occupied quasiparticles a and b.

In the second stage, an appropriate set of spatially-localized quasiparticles must be selected to provide a spatially meaningful partition of the effective two-body interation identified in the first stage. In the development of A-SAPT, our initial aim was to provide a completely black-box partition (e.g., there was to be no invocation of user-specified subsets of atoms to denote functional groups). Moreover, we also intended to provide an automatic partition of the interaction energy to atom-pairwise contributions, to facilitate the development of *ab initio* force fields. Both of these considerations mandated the selection of atoms as the proper quasiparticles, and we further selected the iterative stockholder analysis (ISA) method of Wheatley and coworkers to provide these atoms.[178,179,315] ISA is a variant of the Hirshfeld family of partitions to automatically extract a set of atomic density fields which sum to the total molecular density field. Hirshfeld-type partitions assign a spherical proatomic density to each atom, and then obtain the true atomic density contribution by weighting the total molecular density by the ratio between the proatomic density and the total promolecular density at each point in space. This results in smooth, overlapping atomic density fields. The difference between the various Hirshfeld-type partitions lies in the particular choice of spherical proatomic densities: Hirshfeld[114] uses the sphericalized densities of the underlying free atoms which are neutral, Hirshfeld-I[39] uses the sphericalized densities of underlying free atoms with charges selected to be equal to those of the true atomic densities, while ISA uses spherical proatomic densities which are wholly unconstrained in the radial coordinate. In the latter two approaches, the radial forms of the proatomic densities are determined iteratively by minimizing the information loss between the true molecular density and the promolecular density. Our selection of ISA was based on the fact that ISA makes no

a priori constraints on the proatomic densities beyond being spherical, though we note that A-SAPT could also be easily adapted to use Hirshfeld, Hirshfeld-I, or any other atoms-in-molecules scheme. Having selected ISA as the atoms-in-molecules approach, it remains to connect the effective two-body partition of the SAPT diagrams to the ISA atom-pairwise contributions. For electrostatics and for the electrostatic sources in induction, the atomic contributions are directly obtained by finding the electrostatic potential or electrostatic energy contribution of each ISA atom. For the exchange, dispersion, and response portition of induction, the atomic contributions are found by first accumulating in terms of local occupied orbitals, and then assigning fractions of each local occupied orbital contribution to atomic contributions via ISA orbital atomic charges. This procedure closes the definition of ISA-based A-SAPT0, yielding an atom-pairwise partition for all SAPT0 terms.

It seems obvious that A-SAPT should be able to provide a quantification of functional-group$\cdots$ functional-group interactions, i.e., by simply summing up the relevant atomic contributions $\mathbf{A}$ and $\mathbf{B}$ to the two functional groups $\mathscr{A}$ and $\mathscr{B}$:

$$E^{\mathscr{A}\mathscr{B}}_{\mathrm{int-term,A-SAPT}} = \sum_{\mathbf{A}\,\in\,\mathscr{A}} \sum_{\mathbf{B}\,\in\,\mathscr{B}} E^{\mathbf{AB}}_{\mathrm{int-term,A-SAPT}}. \tag{268}$$

Unfortunately, as noted in the context of the proflavine-DNA intercalation complex of our previous work, A-SAPT can produce unexpected results if quantitative chemical partitions of the electrostatic energy are attempted. The root of this problem is the very concept of atoms in molecules: regardless of the definition of the atom type (Hirshfeld,[114] Hirshfeld-I,[39] ISA,[178] GISA,[307] Bader,[14] etc.), every atom in each molecule will, in general, carry a net partial charge, determined by its arrangement in space relative to the atoms around it, and to the electronegativities of the surrounding atoms. In particular, link bonds between adjacent functional groups are almost always assigned preferentially to the functional group with the more electronegative linking atom, leading to very large charge-charge interactions between functional groups on the two monomers. In the case of phenol dimer, the problem is

obvious: the greater electronegativity of the oxygen means that the OH will carry a charge of roughly -0.2 electrons, while the Ph will carry the balance of $+0.2$ electrons (at ISA/jun-cc-pVDZ as in standard A-SAPT0 analysis). This implies, e.g., that the favorable hydrogen-bond interaction expected in the $\mathrm{OH_A}\cdots\mathrm{OH_B}$ electrostatic interaction is actually dominated by the *unfavorable* $(-0.2)^2/R$ charge-charge interaction. This charge-charge repulsion, plus the corresponding charge-charge repulsion of the $\mathrm{Ph_A}\cdots\mathrm{Ph_B}$ interaction, largely cancel the charge-charge attractions from the two $\mathrm{Ph}\cdots\mathrm{OH}$ interactions. Overall, this leads to a muddled analysis which obfuscates the desired dipole-dipole and higher interactions between neutral functional groups. Note that this phenomenon is not unique to A-SAPT: all other atom-based methods (e.g., DMA or atom-pairwise force fields) will necessarily suffer this same artifact.

9.3.1.5 Many-Body Assignment: F-SAPT

Motivated by the failure of A-SAPT to produce a chemically reasonable functional-group partition of the electrostatic energy, we now develop a paradigm which ameliorates this difficulty by directly considering chemically separable functional groups of integral charge (usually neutral) connected by neutral link bonds. This approach closely resembles the cut-and-cap fragmentation technique above, but retains the possibility of reaching the *ab initio* limit by not performing pairwise computations in isolation. Additionally, arbitrary division of the monomers into any desired patterns of functional groups can easily be obtained in postprocessing, from the results of a single SAPT-type computation on the original dimer. We refer to this new approach as "functional-group SAPT," or F-SAPT.

F-SAPT starts from the A-SAPT-like assertion that the interaction energy terms can be partitioned to pairwise effective interactions between nuclei (A or B) or localized orbitals ($\bar{a}$ or $\bar{b}$, doubly occupied), e.g.,

$$E_{\mathrm{int-term}} = E^{\bar{a}\bar{b}}_{\mathrm{int-term}} + E^{\bar{a}B}_{\mathrm{int-term}} + E^{A\bar{b}}_{\mathrm{int-term}} + E^{AB}_{\mathrm{int-term}}. \tag{269}$$

A-SAPT is already nearly in this form. For instance, the exchange and dispersion terms of A-SAPT0 are already explicitly defined in terms of local orbital pairs, and involve no partitioned nuclear contributions. Only terms involving electrostatic potentials need modification: these were previously assigned directly to atomic quasiparticle contributions in A-SAPT, to avoid extreme oscillations in the visualization. The new nuclear and orbital partition for F-SAPT0 is quite easily obtained, as the contribution of each charge source to the electrostatic potential is linear. For the electrostatics term in SAPT0 (closed-shell RHF), the partition is,

$$E_{\mathrm{elst,r}}^{(10),\bar{a}\bar{b}} \equiv 4(\bar{a}\bar{a}|\bar{b}\bar{b}) = 4\iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \phi_{\bar{a}}(\vec{r}_1)\phi_{\bar{a}}(\vec{r}_1)\frac{1}{r_{12}}\phi_{\bar{b}}(\vec{r}_2)\phi_{\bar{b}}(\vec{r}_2), \tag{270}$$

where $\phi_{\bar{a}}(\vec{r})$ is the $\bar{a}$-th local occupied orbital,

$$E_{\mathrm{elst,r}}^{(10),\bar{a}B} \equiv -2(\bar{a}\bar{a}|Z_B) = -2\int_{\mathbb{R}^3} \mathrm{d}^3 r_1 \; \phi_{\bar{a}}(\vec{r}_1)\phi_{\bar{a}}(\vec{r}_1)\frac{Z_B}{r_{1B}}, \tag{271}$$

where Z_B is the charge of the B-th nucleus, and similarly for $E_{\mathrm{elst,r}}^{(10),A\bar{b}}$, and,

$$E_{\mathrm{elst,r}}^{(10),AB} \equiv \frac{Z_A Z_B}{r_{AB}}. \tag{272}$$

For the induction term, the new partition produces separate nuclear and local orbital contributions to the total electrostatic potential,

$$V_{\bar{a}r}^{\mathbb{B}} \equiv -(\bar{a}r|\rho_{\mathbb{B}}) = -\iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \phi_{\bar{a}}(\vec{r}_1)\phi_r(\vec{r}_1)\frac{1}{r_{12}}\rho_{\mathbb{B}}(\vec{r}_2) \equiv V_{\bar{a}r}^{\bar{b}} + V_{\bar{a}r}^{B}. \tag{273}$$

where $\rho_{\mathbb{B}}(\vec{r})$ is the total charge density of monomer B. Here,

$$V_{\bar{a}r}^{\bar{b}} \equiv 2(\bar{a}r|\bar{b}\bar{b}) = 2\iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \phi_{\bar{a}}(\vec{r}_1)\phi_r(\vec{r}_1)\frac{1}{r_{12}}\phi_{\bar{b}}(\vec{r}_2)\phi_{\bar{b}}(\vec{r}_2), \tag{274}$$

and,

$$V_{\bar{a}r}^{B} \equiv -(\bar{a}r|Z_B) = -\iint_{\mathbb{R}^6} \mathrm{d}^3 r_1 \; \mathrm{d}^3 r_2 \; \phi_{\bar{a}}(\vec{r}_1)\phi_r(\vec{r}_1)\frac{Z_B}{r_{1B}}. \tag{275}$$

These nuclear and local occupied orbital sources each produce linearly-additive induction responses, encapsulated by the induction amplitudes $x_{\bar{a}r}^{\bar{b}}$ or $x_{\bar{a}r}^{B}$. As in our previous work (see Equations 18 and 19 in Reference 222), these amplitudes can be

obtained through either the uncoupled approximation (used throughout this work, with scalings as in the previous work) or through coupled-perturbed Hartree-Fock. Once the amplitudes are obtained, the full pairwise partition to the induction energy is obtained as,

$$E^{(20),\bar{a}\bar{b}}_{\mathrm{ind(A\leftarrow B)}} \equiv 2x^{\bar{b}}_{\bar{a}r}\left[V^{\mathbb{B}}_{\bar{a}r} + \tilde{V}^{\mathbb{B}}_{\bar{a}r}\right], \tag{276}$$

and,

$$E^{(20),\bar{a}B}_{\mathrm{ind(A\leftarrow B)}} \equiv 2x^{B}_{\bar{a}r}\left[V^{\mathbb{B}}_{\bar{a}r} + \tilde{V}^{\mathbb{B}}_{\bar{a}r}\right]. \tag{277}$$

Here, $\tilde{V}^{\mathbb{B}}_{\bar{a}r}$ is the exchange-renormalization potential of the exchange-induction term, defined in our previous work (see Equation 21 of Reference 222). Note that the blocks of $E_{\mathrm{ind(A\leftarrow B)}}$ corresponding to the nuclei A are zero, as the involvement of monomer A in this term is an electronic response only. Further, corresponding expressions for $E_{\mathrm{ind(B\leftarrow A)}}$ can be obtained by permutation of indices.

In practice, the above modifications are easily and efficiently implemented, using the density fitting approximation[324,67,68,299] in the -JKFIT auxiliary basis sets.[312] These terms are actually much easier to implement and evaluate than those in the atom-based A-SAPT partition, as the above F-SAPT expressions are all tractable in a spectral basis set, and do not require numerical quadrature to evaluate.

The nuclear and local orbital partition above is easy to attain, but it does not immediately yield interpretable data. One reason for this is that electronic and nuclear sources have been split, leading to enormous and nearly cancelling contributions to the partition. To overcome this, we group chemically intuitive electronic and nuclear contributions into fragments $\mathscr{A}$ and $\mathscr{B}$, including chemical functional groups and link-bond subsystems. The mathematical form for this process is given as,

$$E^{\mathscr{A},\mathscr{B}}_{\mathrm{int-term,full}} \equiv w^{\mathscr{A}}_{\bar{a}} w^{\mathscr{B}}_{\bar{b}} E^{\bar{a}\bar{b}}_{\mathrm{int-term}} + w^{\mathscr{A}}_{\bar{a}} w^{\mathscr{B}}_{B} E^{\bar{a}B}_{\mathrm{int-term}} + w^{\mathscr{A}}_{A} w^{\mathscr{B}}_{\bar{b}} E^{A\bar{b}}_{\mathrm{int-term}} + w^{\mathscr{A}}_{A} w^{\mathscr{B}}_{B} E^{AB}_{\mathrm{int-term}}. \tag{278}$$

The quantities $w^{\mathscr{A}}_{\bar{a}}$ and $w^{\mathscr{A}}_{A}$ are the orbital-to-fragment and nucleus-to-fragment weights, respectively. The orbital-to-fragment weights $w^{\mathscr{A}}_{\bar{a}}$ are always 0 or 1, as each orbital is

wholly assigned to a single fragment in this scheme. The nucleus-to-fragment weights are also 0 or 1 for any atom not included in a link bond, in which case the weight will be a fraction between 0 and 1, as described below.

The general procedure for assigning orbital and nuclear weights is as follows. First, the monomer geometry is traversed to identify bridging σ bonds between user-specified functional groups. If linking covalent bonds are found (typically by inspection of the interatomic distances and van der Waals radii, followed by orbital charge analysis) the corresponding link fragment will be assigned a single σ orbital bridging the two constituent atoms (by "assigned," it is meant that the corresponding weight $w_{\hat{a}}^{\mathscr{A}}$ is set to 1). Once this is done, the remaining orbitals are assigned to the user-specified functional groups according to the functional group on which each orbitals has the largest atomic charge. With the orbital-to-fragment weights assigned, the next task is the nucleus-to-fragment weights. To seed these weights, $w_A^{\mathscr{A}}$ is set to 1 if the nucleus A is in the user-defined functional-group $\mathscr{A}$, 0 otherwise. From this, the fraction $1/Z_A$ is removed from the weight $w_A^{\mathscr{A}}$ and assigned to the weight $w_A^{\mathscr{A}'}$, for each of the two atoms in the link fragment $\mathscr{A}'$. This effectively removes a proton from each functional group involved in the link bond, and assigns these protons to the link bond fragment, resulting in a neutral link fragment with two electrons in a σ orbital and two protons. This shuffling of two protons to each linking σ bond is critical to achieve a sensible partition: if the protons were left with the functional groups, the functional groups would usually carry net positive charges while the link bond groups would each carry a charge of -2, leading to large oscillations in the electrostatic partition. Another way to view this division of the monomer is into a set of neutral functional groups interconnected by a set of link fragments (each of which loosely resembles an extended hydrogen molecule).

In practice, much of the above weight selection can be automated, with excellent results for chemically reasonable choices of functional groups. The user first provides

the F-SAPT post-processing script with the set of atoms defining each chemically-relevant functional group (e.g., O and H to define the OH functional group and C_5H_6 to define the Ph functional group). The distances between pairs of atoms in difference are then checked against the sum of van der Waals radii of the constituent atoms, and candidates for link bonds are identified automatically (this can be specified manually in special cases). For each link bond between atoms A and A', the local orbital corresponding to the linking σ bond is easily identified as that which maximizes the bond-index-like metric $M_{\bar{a}} = Q_A^{\bar{a}} Q_{A'}^{\bar{a}}$, where $Q_A^{\bar{a}}$ are local orbital atomic charges computed at the same level as the Pipek-Mezey-like orbital localization routine. Once the link orbitals are assigned, the remaining orbitals are assigned to the chemical functional group containing the atom with the maximal orbital atomic charge $Q_A^{\bar{a}}$. The nuclear weights are then automatically assigned as above.

A key assumption in the F-SAPT analysis is that the functional groups are chemically separable entities linked by, at most, simple σ bonds. This is obvious when one considers the weights $w_{\bar{a}}^{\mathscr{A}}$, which are integral, meaning that we have chosen the partition in such a way that non-link electrons are wholly owned by a single functional group. This implies that neighboring functional groups linked by double, aromatic, etc. bonds, or featuring significant fractional charge transfer are not appropriate choices within F-SAPT. We do not believe this represents a limitation of F-SAPT, but rather a form of maximum resolution of effective quasiparticle separability in the many-body interaction: two user-defined fragments linked by more than simple σ bonds are actually acting as a single chemical entity, and should be treated as such. Moreover, F-SAPT *does* provide for polarization of the electrons within each functional group, to respond to neighboring functional groups (this flexibility is not present in fragmentation approaches, which are performed on pairs of fragments in isolation). An excellent automatic check of the validity of user-defined functional

groups is the completeness of the orbital atomic charges,

$$C_{\bar{a}} = 1 - \sum_{A \in \mathscr{A}} Q_A^{\bar{a}}. \tag{279}$$

If $C_{\bar{a}}$ is greater than a few hundredths of an electron pair, this is an indication that a significant piece of this orbital lies across more than one fragment, and that the functional group partition should be redefined.

For example, consider the F-SAPT partition scheme for phenol, depicted in Figure 36. The link σ bond is clearly identifiable from the orbital atomic charges, and is also assigned one proton from the OH oxygen (a nuclear-to-fragment weight of 1/8) and one proton from the nearest carbon on the phenyl group (a nuclear-to-fragment weight of 1/6). The hydroxyl group retains 7 of the 8 oxygen protons and the hydrogen proton, as well as the oxygen $1s$ orbital, the OH σ orbital and the two lone-pair orbitals. The phenyl group retains 5 of the 6 protons on the linking carbon, the 5 remaining carbon nuclei, the 5 hydrogen nuclei, the 6 carbon $1s$ orbitals, the 6 C–C σ orbitals, the 5 C–H σ orbitals, and the 3 π orbitals. Note that the π orbitals form a delocalized system over the carbons of the phenyl group, and cannot be unambiguously localized to subsets of these atoms. However, the π system can be localized to the level of the phenyl functional group, which functions as a single, separable chemical unit. The largest value of $C_{\bar{a}}$ for this system is 0.045 electron pairs, indicating mild polarization of one of the oxygen lone pairs toward the π system of the benzene.

To this point, the F-SAPT analysis produces a pairwise partition of the interaction energy terms between chemical functional groups *and* link bond fragments, e.g. $\mathrm{Ph_A}\cdots\mathrm{Ph_B}$, $\mathrm{Ph_A}\cdots\mathrm{OH_B}$, $\mathrm{Ph_A}\cdots\mathrm{Link_B}$, etc. contributions. We refer to this level-of-detail as a "full F-SAPT" partition. For detailed analysis, the explicit enumeration of the link fragment contributions may be important, particularly if these (often polar) bonds are significantly involved in the interaction. In these cases, the link bonds are chemical entities of their own, and should be considered separately from the chemical

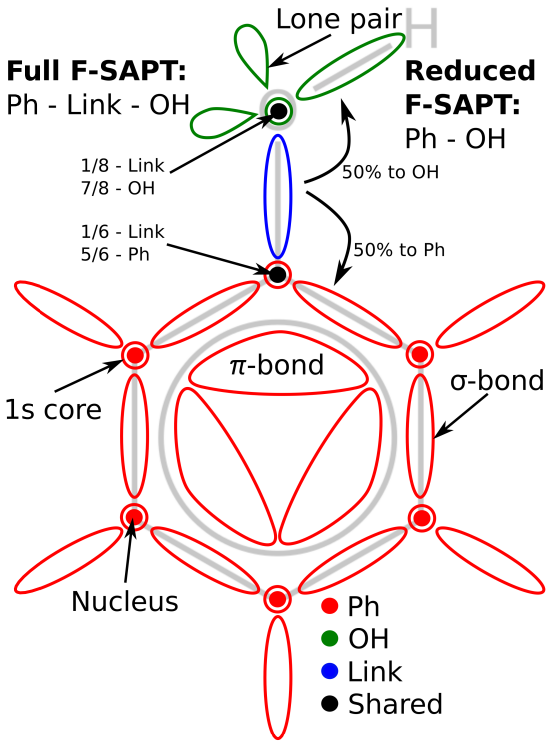

Figure 36: Phenol monomer F-SAPT partitioning scheme. In full F-SAPT, one proton from the hydroxyl O and one proton from the nearest phenyl C are assigned to the link σ-bond to produce neutral Ph, Link, and OH fragments. In reduced F-SAPT, the full F-SAPT partition for the Link fragment is split equally between the Ph and OH fragments.

functional groups. However, if the link bonds do not contribute significantly to the interaction, or if the user is only concerned with semi-quantitative trends, the analysis can be vastly simplified by assigning the link fragment contributions to the two functional groups that they bridge, i.e.,

$$E^{\mathscr{A}\mathscr{B}}_{\text{int-term,reduced}} = E^{\mathscr{A}\mathscr{B}}_{\text{int-term,full}} + f^{\mathscr{B}}_{\mathscr{B}'} E^{\mathscr{A}\mathscr{B}'}_{\text{int-term,full}} + f^{\mathscr{A}}_{\mathscr{A}'} E^{\mathscr{A}'\mathscr{B}}_{\text{int-term,full}} + f^{\mathscr{A}}_{\mathscr{A}'} f^{\mathscr{B}}_{\mathscr{B}'} E^{\mathscr{A}'\mathscr{B}'}_{\text{int-term,full}}. \tag{280}$$

Here $f^{\mathscr{A}}_{\mathscr{A}'}$ is a weight assigning all or part of the link fragment $\mathscr{A}'$ to the chemical functional group $\mathscr{A}$, which is only nonzero if $\mathscr{A}'$ and $\mathscr{A}$ share a nucleus. Unfortunately, there is no clearly unambiguous scheme to provide these weights: one could

imagine assigning all of the link bond to one of the two functional groups, assigning some fraction of the link fragment to the two functional groups based on the ratio of orbital atomic charges, or evenly dividing the link fragment between the two functional groups. We have found that, for chemically reasonable choices of functional groups, all of these schemes generally produce qualitatively identical results, with the last two producing nearly quantitatively identical results in most cases. The last two schemes also contain a bit more physical merit than the first: partitioning based on orbital charges respects the concept of how much each fragment "owns" the link σ orbital, while equal partitioning produces chemical functional groups with integral particle counts in both electrons and protons (e.g, the OH group in phenyl contains 9 electrons and 9 protons with equal partitioning). Due to the particular elegance of particle count of the last choice, and the lack of concrete definition of an orbital atomic charge, we have selected the equal partitioning scheme ($f^{\mathscr{A}}_{\mathscr{A}'} = 0.5$) for use in simpler, semi-quantitative F-SAPT analysis. We denote this choice of partition as a "reduced F-SAPT" analysis. By default, "F-SAPT" below refers to reduced F-SAPT, while consideration of full F-SAPT will be explicitly enumerated.

9.3.2 Interpretations and Limitations

In the process of review, a referee raised the admirable critique that F-SAPT is necessarily not the only possible decomposition of the SAPT interaction energy, and, moreover, raised the complaint that our choice of partition ignores the virtual quasiparticles present in the response portions of the SAPT expressions (the exchange, dispersion, and induction response terms). We provide this section to discuss the limitations and choices made in the development of F-SAPT.

One way to view F-SAPT is to assert that the monomers are composed of transferable collections of localized electrons and protons/neutrons (functional groups) which, after selection under some localization metric and prescribed set of functional

groups, are allowed to interact naturally with one another. At the level of Hartree-Fock theory (as encountered in SAPT0 here), the protons/neutrons and occupied orbitals completely parametrize the states of the noninteracting monomers, albeit with a mean-field description of the wavefunctions. If Kohn-Sham theory were invoked (as in DFT-SAPT, which we intend to apply in later efforts), the protons/neutrons and localized occupied Kohn-Sham orbitals could theoretically be used to completely parametrize the *exact* states of the noninteracting monomers, if the exact Kohn-Sham functional were used. To this point, no mention has been made of the concept of the virtual space. We now ask the same question as SAPT itself asks: how do these monomers interact with one another? Further, we will retain the labels of the localized occupied quasiparticles in computing this interaction, yielding F-SAPT. In doing so, we will arrive at the F-SAPT partition of the SAPT equations, which do contain references to the virtual space. The virtual space appears here as a mechanism to provide for the response of one monomer's wavefunction in a perturbation (usually from the other monomer). However, here we are interested only in which portion of the monomer the electron was localized before the interaction occurred (the functional group of origin), not to where it excited to after the interaction was invoked. In this view, the virtual space is not relevant as anything other than a mathematical construction to allow for response interactions, i.e., as a medium through which the occupied quasiparticles interact. Further evidence for this choice is the fact that the local occupied orbitals and nuclei directly outline the intuitive shapes of the chemical functional groups, and are localizable and largely invariant to basis set size or type. By contrast, the virtual space is intrinsically nonlocalizable, particularly in the complete-basis or plane-wave limits. All of these considerations point to the intuitive validity of our selection of local occupied quasiparticles in designing F-SAPT.

At this point, it is imperative to remember that (1) any partition of the interaction energy terms into spatially-localized contributions will not correspond to any set of

observable quantities and (2) any partition of interaction energy terms will necessarily not be unique. Such partitions are intended to aid our chemical understanding of noncovalent contacts, but should not be mistaken for absolute physical realism. We have designed F-SAPT to correspond to chemical intuition while using only minimal prescription (it requires only a localization metric and a human-specified set of functional groups, together with the picture of interactions between localized occupied quasiparticles discussed above), hopefully resulting in a robust, reliable partition. See Section II.E on page 10-11 of our A-SAPT paper[222] for an additional discussion of the related ambiguities present in A-SAPT.

9.4 Results

9.4.1 Computational Details

All results presented in this work use A-SAPT0 and F-SAPT0 implementations in a development version of the `PSI4` package,[298] largely according to the methodology discussed in our original A-SAPT0 paper. The dimer-centered jun-cc-pVDZ orbital basis set is used for all computations.[124,212] The density fitting approach is used to accelerate all Hartree-Fock[312] and SAPT0[111,123,118] computations, with the jun-cc-pVDZ-JKFIT auxiliary basis set used for the Hartree-Fock wavefunctions, electrostatics, exchange, induction, and one-particle contributions to the exchange-dispersion terms, while the jun-cc-pVDZ-RI auxiliary basis set is used for the two-body contributions to the dispersion and exchange-dispersion terms. The frozen core treatment is invoked in the dispersion term, freezing $1s$ electrons on first-row atoms and $1s$, $2s$, and $2p$ electrons on second-row atoms. For A-SAPT, atomic assignment of the electronic density is accomplished through the ISA procedure, using large Treutler-Ahlrichs grids with 100 radial points and 302 angular points.[20,297] Scalings for infinite-order exchange, induction coupling, and higher-order induction are applied as recommended in our previous work.

In contrast to our previous paper, we use intrinsic bond orbital (IBO) methodology developed by Knizia to provide local occupied orbitals and orbital atomic charges.[149] In our previous work, these were provided by Pipek-Mezey local orbitals[228] and a form of ISA orbital charges, respectively. IBO is an extremely elegant Pipek-Mezey-like approach that is essentially invariant to basis set quality (a major problem with Mulliken-charge-based Pipek-Mezey). IBO methodology has very recently been shown[136] to be essentially equivalent to (albeit numerically simpler than) the quasi-atomic minimal basis set (QUAMBO) methodology of Ruedenberg and coworkers.[183] As in Knizia's work, we use the fully contracted cc-pVTZ core and valence functions[70] to define the minimal atomic orbitals, and use the fourth-power metric in the localization procedure. Separate localizations are performed for the core and valence orbitals (after the unified set of intrinsic atomic orbitals are generated), to prevent mixing of core and valence spaces in the frozen-core dispersion treatment. Minimal AOs are placed only on the active monomer: ghost atoms do not receive minimal AOs.

For cut-and-cap analyses, constrained optimizations for capping hydrogens are performed at the B970-D2/aug-cc-pVDZ[22,96] level in `Q-Chem 4.1`,[269] with monomer-centered basis sets. SAPT0/ jun-cc-pVDZ is used for the pairwise cut-and-cap electronic structure computations, as is consistent with the A-SAPT and F-SAPT analyses. The H_2 particle-count corrections are not applied in this work.

Dimer geometries, cut-and-cap fragment geometries, and complete cut-and-cap, A-SAPT, and F-SAPT analyses for all systems and SAPT terms are provided in the supplemental material.

9.4.2 Phenol Dimer Example

Figure 37 depicts various decompositions of the electrostatic energy for the framing phenol dimer example of the theory section, partitioned by the cut-and-cap fragmentation method, and by the A-SAPT and reduced F-SAPT many-body partitioning

techniques. The most immediate feature of this illustration is that A-SAPT provides a significantly weaker hydrogen bonding contribution than one might expect (vs. say that predicted by cut-and-cap analysis), due to the obfuscating influence of partial charges on the ISA atom-based fragments. The dipole-dipole terms manage to overcome the $\delta^- \cdots \delta^-$ charge-charge repulsion to produce a net attraction between the two hydroxyl groups, but the magnitude of this attraction is much weaker than expected. In fact, the A-SAPT $\mathrm{Ph_A} \cdots \mathrm{OH_B}$ interaction is significantly more attractive than that of $\mathrm{OH_A} \cdots \mathrm{OH_B}$, due to the spurious $\delta^+ \cdots \delta^-$ charge-charge attraction in the former. The $\mathrm{Ph_A} \cdots \mathrm{Ph_B}$ electrostatic interaction is also poorly described by A-SAPT, with substantial repulsion arising from the $\delta^+ \cdots \delta^+$ charge-charge term.

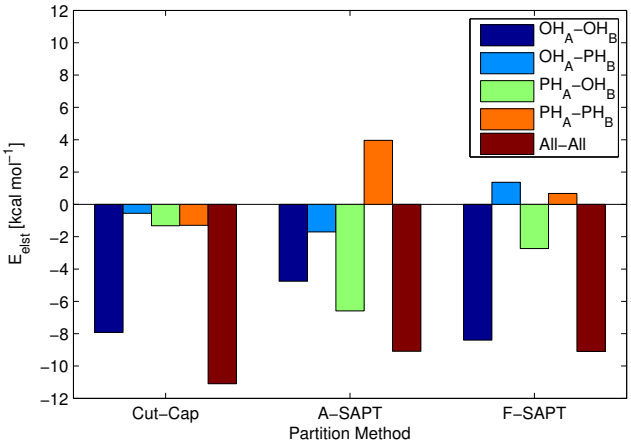

Figure 37: Phenol dimer SAPT0/jun-cc-pVDZ electrostatic energy partitioned by Cut-Cap, A-SAPT, and reduced F-SAPT schemes.

The second key feature of this figure is the strong similarity between the cut-and-cap partition and the reduced F-SAPT partition. Both techniques clearly indicate the $\mathrm{OH_A} \cdots \mathrm{OH_B}$ hydrogen bonding, with electrostatic contributions of -7.9 kcal mol^{-1} and -8.4 kcal mol^{-1} respectively. The next strongest contributor is $\mathrm{Ph_A} \cdots \mathrm{OH_B}$ interaction, with electrostatic contributions of -1.3 kcal mol^{-1} and -2.7 kcal mol^{-1}, respectively. Note that the total SAPT0/jun-cc-pVDZ electrostatic energy for this system is -9.1 kcal mol^{-1}, while the cut-and-cap electrostatic energy is -11.1 kcal

mol^{-1}, indicating that isolated pairwise computations induce an error of 2.0 kcal mol^{-1} in cut-and-cap analysis. As the differences between the reduced F-SAPT and cut-and-cap electrostatic partitions are at most 2.0 kcal mol^{-1}, we can conclude the F-SAPT and cut-and-cap analyses agree to within the maximum resolution of the latter method.

The link bond contributions to the full F-SAPT electrostatic partition are not entirely negligible, but never exceed a magnitude of 1.5 kcal mol^{-1}, e.g., much smaller than the total $OH_A \cdots OH_B$ interaction. Moreover, partitioning the link bonds by IBO orbital charges vs. equal assignment results in a maximum deviation of only 0.34 kcal mol^{-1} in the reduced F-SAPT partition.

The results obtained for this case study are quite encouraging. F-SAPT produces chemically intuitive results for the electrostatic contributions for this system (in contrast to A-SAPT) and also returns the many-body SAPT0 limit (in contrast to cut-and-cap). Additionally, the ambiguities associated with link-bond assignment in reduced F-SAPT are far smaller in magnitude than the key contributions to the electrostatics partition. We have observed similar correspondence in many other chemical systems and in the other (easier) SAPT terms.

9.4.3 Proflavine Intercalation Example

We now turn our attention to a particularly problematic and motivating case encountered in our A-SAPT paper: the intercalation of a proflavine cation into a double-stranded CGA trinucleotide fragment of DNA, including the sugar-phosphate backbone. The geometry and chemical fragmentation pattern for this system are depicted in Figure 38. The geometry[184] is the same as that used in our 2011 DF-SAPT0 paper[118] and our A-SAPT paper.[222] Note that we have protonated each phosphate group to simulate the presence of a counterion, resulting in a neutral DNA monomer.

The proflavine cation carries a +1 net charge. The DNA fragment is labeled according to the nucleotide on the left side of Figure 38, going from 5′ to 3′ (top to bottom). The proflavine has intercalated between the C and G steps in this ordering, causing significant distortion of the DNA structure. We label the base pairs as BP1 (C), BP2 (G), and BP3 (A), from top to bottom. The sugars are labeled according to their side and neighboring base pair, e.g., SL1 is the left-hand sugar bonded to BP1. The phosphates are labeled by their side and the two base pairs they link, e.g., PL12 is the left-hand phosphate bridging BP1 and BP2, via SL1 and SL2. Sugar–base pair functional groups are separated by cutting glycosidic C–N bonds, while sugar-phosphate functional groups are separated by cutting C–O bonds.

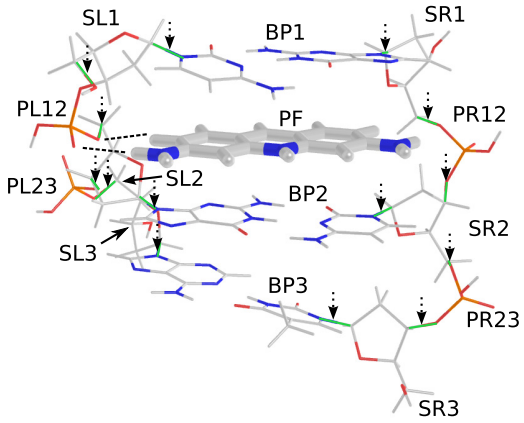

Figure 38: Proflavine$^+$-DNA intercalation complex geometry and functional group partitioning scheme. Link bonds are indicated by solid green lines and emphasized by downward-pointing dotted arrows. Intermolecular hydrogen bonds are indicated by dashed black lines.

This is an interesting test case due to the complex interplay between various contributions to the intermolecular interaction: in our 2011 paper,[118] we performed a coarse SAPT0/jun-cc-pVDZ cut-and-cap analysis (to the level of base pairs and backbones), and determined that (1) roughly 2/3 of the interaction strength comes from π-stacking between the proflavine and neighboring base pairs, (2) the next-nearest

neighbor base pair contributes negligibly to the interaction, and (3) roughly 1/3 of the interaction strength comes from interactions with the closer left-side backbone, possibly including long hydrogen bonds forming between the proflavine and left upper phosphate. In our A-SAPT paper, the first and second findings were easily visible in order-1 atomic visualizations of the dispersion and exchange terms, which clearly showed the nearest-neighbor π-stacking interactions, and negligible interactions with the next-nearest neighbor base pair. However, the third finding was much harder to elucidate, due to the sizeable oscillations in the order-1 atomic electrostatics visualizations induced by the uneven assignment of bonds linking adjacent atoms of the DNA fragments, under the strong driving field of the proflanine cation. By differentially increasing the color scales in the order-1 atomic visualizations, or looking through the extensive order-2 atom-pairwise data, it was apparent that PL12 was the most active phosphate in the electrostatic interaction, but this was, at best, a qualitative conclusion.

Here we revisit this problem by quantitatively partitioning the electrostatic energy to base pair, sugar, and phosphate contributions via cut-and-cap, A-SAPT, and F-SAPT analyses. Note that the proflavine cation is treated as a single functional group, so the order-1 and order-2 analyses are equivalent for this system. This allows us to use visualization techniques adapted from A-SAPT to quickly depict the relevant interactions in the system: Figure 39 shows the qualitative and quantitative results from these analyses. In each panel, the functional groups are colored according to the sign and magnitude of their electrostatic interaction with the other monomer, with red indicating attraction and blue indicating repulsion. The small numbers in the figure indicate the quantitative electrostatic contributions, in kcal mol^{-1}. The color scales saturate at $\pm$30 kcal mol^{-1} to provide adequate resolution for the smaller contributions. This means that the proflavine contributions will be saturated in all three analyses, as will some of the DNA contributions in the A-SAPT analysis.

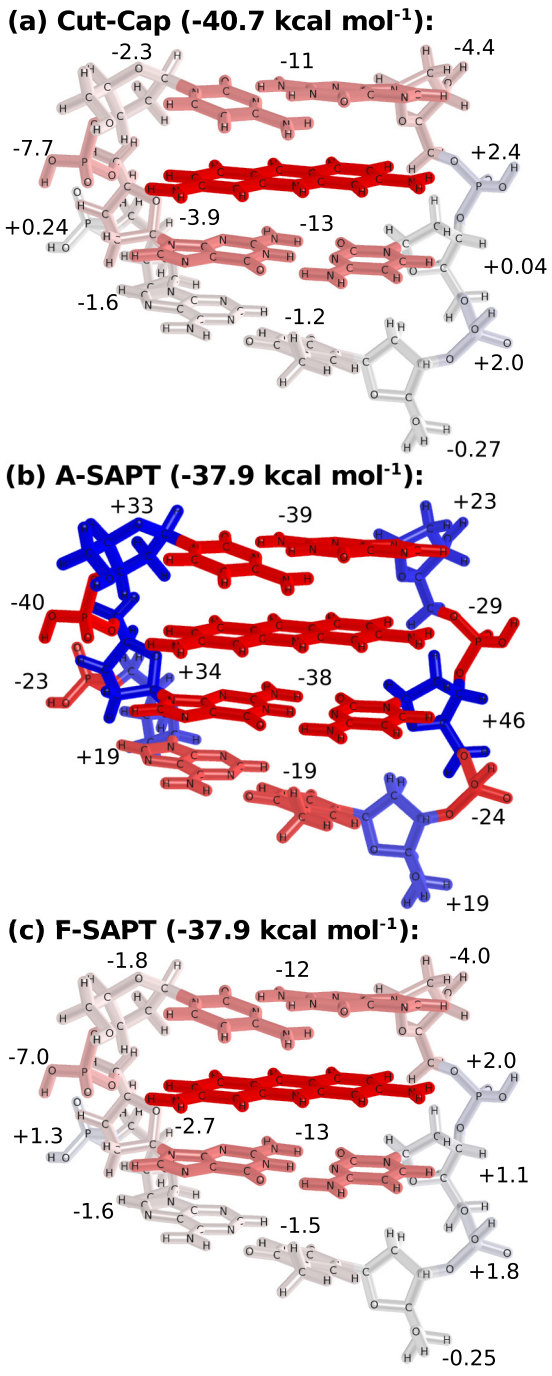

Figure 39: Proflavine^{+}-DNA SAPT0/jun-cc-pVDZ electrostatic energy partitioned to functional groups via (a) Cut-Cap, (b) A-SAPT, and (c) reduced F-SAPT schemes. Red indicates attraction, blue indicates repulsion, white indicates no net interaction. The color scales saturate at ± 30 kcal mol^{-1}. Small numbers indicate functional group electrostatic contributions in kcal mol^{-1}.

Just as with phenol dimer, the quantitative A-SAPT electrostatic partition does not provide much insight. The sugars have large repulsive electrostatic interactions, while the base pairs and phosphates have large attractive electrostatic interactions. This is easily explained when one considers that the link bonds involving sugars are either C–N or C–O, with the C on the sugar: the electron pairs in these polar covalent bonds will be assigned more preferentially to the N or O atoms of the base pairs or phosphates, respectively. This means that the sugars will attain net positive charges, while the base pairs and phosphates will attain net negative charges. The charge-charge interactions between these partially charged functional groups and the +1 charged proflavine cation strongly dominate the quantitative A-SAPT partition. Within the base pairs and sugars, it is apparent that the nearest neighbor base pairs and PL12 phosphate have the strongest electrostatic interactions, and that the next nearest neighbor base pair is less involved. However, as was the case with order-1 atomic visualizations in A-SAPT, this is, at best, a qualitative conclusion: there are oscillating contributions of $> \pm 10$ kcal mol^{-1} from all functional groups, and the strongest overall contributor is actually SR2, with a contribution of +46 kcal mol^{-1}.

The invocation of reduced F-SAPT markedly cleans up the electrostatic partition, yielding results which strongly resemble cut-and-cap fragmentation, while reproducing the full SAPT0 electrostatic energy. Quoting the F-SAPT partition, the rough conclusions from our DF-SAPT0 cut-and-cap analysis are immediately ratified, along with additional detail. The first finding above is apparent in the substantial -12 and -13 kcal mol^{-1} electrostatic contributions of BP1 and BP2, respectively, indicating an interesting cation-π and π-stacking electrostatic interaction between the proflavine and these two neighboring base pairs. The second finding above is apparent from the small contributions of PL23, PR23, SL3, SR3, and BP3, which individually contribute no more than ± 1.8 kcal mol^{-1} to the electrostatic energy, and collectively contribute a negligible -0.25 kcal mol^{-1} to the electrostatic energy. The third finding above is also

visible, with some enhancement provided by F-SAPT. With an electrostatic contribution of -7.0 kcal mol^{-1}, PL12 is clearly the strongest contributor from the backbone, and is the strongest overall contributor after the nearest neighbor base pairs. This supports our intuitive conclusion of long hydrogen bonding between proflavine and PL12 from the 2011 paper. However, it is interesting to note that there are other non-negligible contributors from the two backbones, e.g., SR1 (-4.0 kcal mol^{-1}) and SL2 (-2.7 kcal mol^{-1}). Both of these sugars are oriented so that the lone pairs of the ring oxygen are facing the proflavine cation, providing favorable electrostatic orientations. Additionally, SL1 is distorted so that the terminal hydroxyl group is oriented favorably with respect to the proflavine cation; while the ring oxygen is not favorably aligned for this sugar, the hydroxyl group is able to provide for a net favorable contribution of -1.8 kcal mol^{-1}. The contributions from SL1, PL12, and SL2 all add to the attractive contribution of the left-side backbone, indicating that while long hydrogen bonds with PL12 are the dominant contributors, the nearby sugars of the left-side backbone also contribute to the overall stability of the complex. Moreover, the favorable electrostatic contribution from SR1 is largely canceled by unfavorable contributions from PR12, SR2, etc. This implies that our original conclusion of negligible interactions between proflavine and the right-side backbone only holds in the net: non-trivial interactions do occur with various units of the right-side backbone, but these nearly cancel overall.

Note that the total electrostatic energy at SAPT0/jun-cc-pVDZ is -37.9 kcal mol^{-1}, while the pairwise cut-and-cap electrostatic energy is -40.7 kcal mol^{-1}, i.e., a many-body error of -2.8 kcal mol^{-1}. In particular, the cut-and-cap contributions are generally slightly more attractive than their F-SAPT counterparts, which is possibly due to the greater polarization of capping C–H, O–H, and N–H σ bonds in the cut-and-cap analysis. F-SAPT is able to avoid this error while producing a chemically useful partition.

9.4.4 Cucurbituril Inclusion Example

To demonstrate the insight provided by a complete F-SAPT analysis, we consider a cucurbit[7]uril$\cdots$ bicyclo[2.2.2]octane-1,4-dimethanol inclusion complex (geometry obtained from Ref. 50, originally proposed in Ref. 199), depicted in Figure 40. Inclusion complexes involving cucurbituril hosts and polar or cationic guest molecules are of great interest in supramolecular chemistry, due to their enormous binding constants and amazing specificities. Myriad inclusion complexes involving neutral and cationic ferrocene derivatives,[137,181,244,199] cationic adamantyl derivatives,[181] and neutral and cationic bicyclo[2.2.2]octane derivatives[199] as guests have been reported in both experimental and theoretical studies. The proposed origins for such strong binding affinities are the complimentary shapes of the host and guest, which lead to strong noncovalent enthalpic contributions, and to a remarkably small entropic penalty in the inclusion complex, arising from the rigidity of the constituent monomers. Cationic guests are found to have somewhat higher binding affinities, presumably due to favorable electrostatic interactions between the cationic side substituents of the guest and the rings of carbonyl groups in the cucurbituril (though this favorability is somewhat diminished by the corresponding desolvation penalty). The complex considered in the present study does not have a cationic guest, but rather features a pair of polar, neutral methanol groups attached to a bicyclo[2.2.2]octane core. This guest was proposed in the mining minima algorithm (M2) theoretical study of Gilson and co-workers,[199] and was further investigated by the noncovalent interactions index (NCI) of Yang and co-workers.[50] The NCI visualization heuristic of the latter was able to detect strong van der Waals interactions throughout the inclusion cavity (presumably indicating dispersion between the host and the core of the guest), and additionally indicated that "the hydroxyl substituents on the guest establish strong hydrogen bonds with two carbonyls of the [cucurbituril] host."

The functional-group partition selected for this complex is depicted in Figure 40.

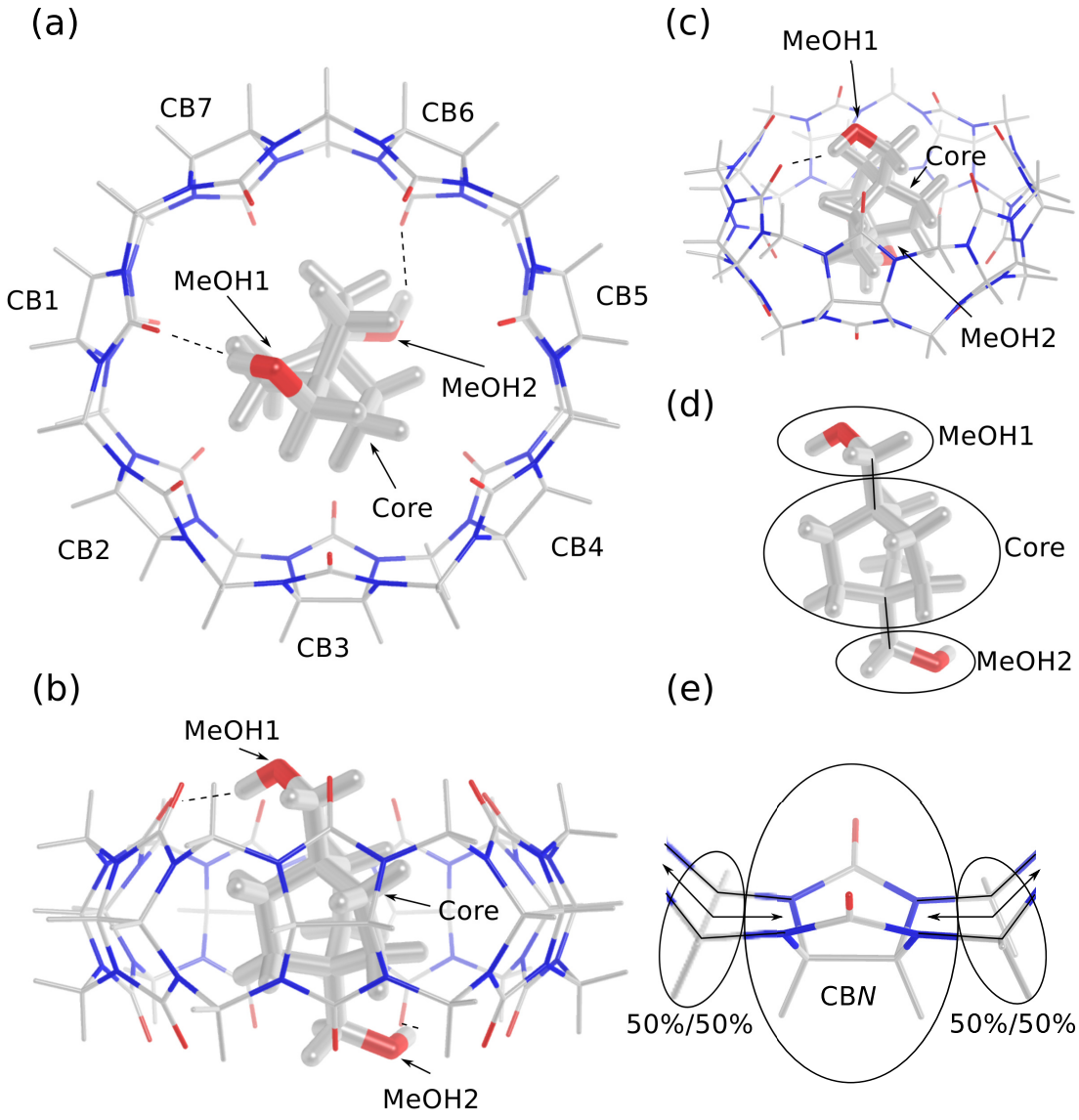

Figure 40: Cucurbit[7]uril$\cdots$bicyclo[2.2.2]octane-1,4-dimethanol inclusion complex geometry and functional group partitioning scheme. (a) Top view. (b) Side view. (c) Birds-eye view. (d) bicyclo[2.2.2]octane-1,4-dimethanol functional group partiton. (e) Cucurbit[7]uril functional group partition. Ovals denote functional group selection. Link bonds are indicated by solid black lines. Intermolecular hydrogen bonds are indicated by dashed black lines.

For the guest, we divide the monomer into a bicyclo[2.2.2]octane core (Core) and two methanol groups (MeOH1 and MeOH2). We alternatively could have partitioned to the level of hydroxyl groups and obtained similar results; the main deviations between these choices of partition are small exchange and dispersion contributions from the

CH_2 links in the methanols. For the host, we use a slightly more exotic partition to simplify the analysis. We first perform a reduced F-SAPT analysis with 14 host functional groups: the 7 uniquely identifiable cucurbituril units and the 7 sets of linking CH_2 pairs. Then, we assign the (minor) contributions from the CH_2 pairs evenly to their adjacent cucurbituril units (CBN). This "collapsed F-SAPT" scheme allows us to effectively slice the linking CH_2 pairs down the middle. Figures 41 and 42 depict the quantitative results from the collapsed F-SAPT partition, for all SAPT terms. Each row of these figures depicts an individual SAPT term, while each column within the row depicts the contributions to the core, MeOH1, and MeOH2 portions of the guest, from left to right. The colorings of the cucurbituril units correspond to their contributions from the current term and portion of the guest, while the colorings of the guest units correspond to their contributions from the current term and all cucurbituril units. The color scales saturate at ± 20 kcal mol^{-1}, implying that the total core contributions will sometimes be saturated.

The obvious conclusion from the collapsed F-SAPT analysis is the large and favorable interaction between the host and guest core, which contributes -28.7 kcal mol^{-1} of the net interaction strength of -28.2 kcal mol^{-1}. The major contributors to this interaction are electrostatics (-15.1 kcal mol^{-1}), exchange ($+24.5$ kcal mol^{-1}) and especially dispersion (-33.2 kcal mol^{-1}). The dispersion contribution is quite isotropic througout the cucurbituril, while the electrostatic and exchange interactions are much more anisotropic. The contributions of CB5 are the smallest overall, corresponding to CB5 being the most geometrically distant cucurbituril unit from the guest. The always-attractive and strongly distance-dependent electrostatic interaction between the host and nonpolar guest core is strongly indicative of a favorable charge penetration interaction (though a distributed multipole analysis would be required to quantitatively prove this claim). The induction contribution from the cucurbituril polarizing the core is surprisingly large, with a net contribution of -4.7 kcal mol^{-1}.

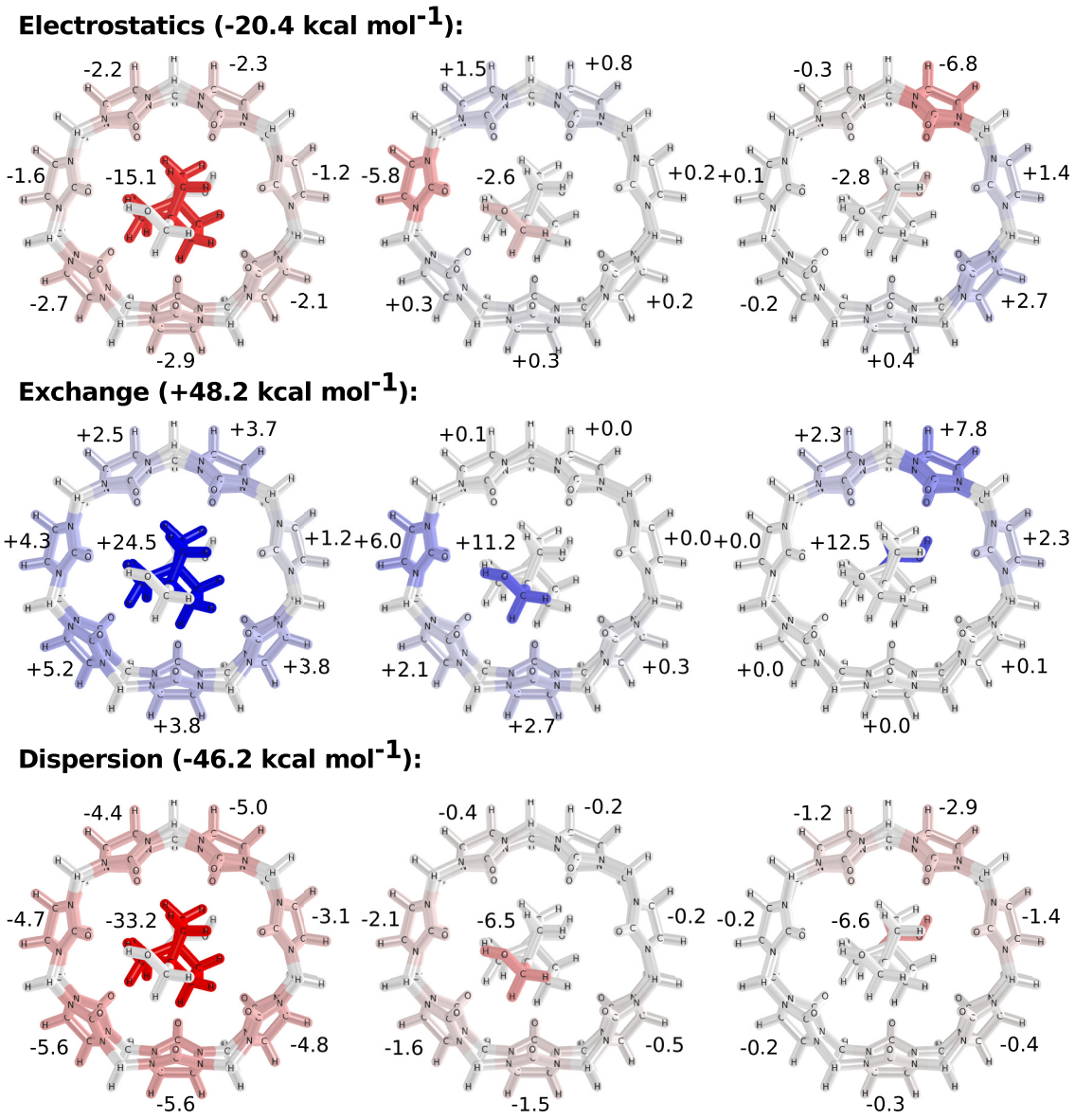

Figure 41: Cucurbit[7]uril$\cdots$bicyclo[2.2.2]octane-1,4-dimethanol inclusion complex SAPT0/jun-cc-pVDZ electrostatic, exchange, and dispersion terms partitioned by collapsed F-SAPT scheme discussed in the main text. Left: contributions to Core. Middle: contributions to MeOH1. Right: contributions to MeOH2. Red indicates attraction, blue indicates repulsion, white indicates no net interaction. The color scales saturate at ± 20 kcal mol^{-1}. Small numbers indicate functional group electrostatic contributions in kcal mol^{-1}.

Each polar cucurbituril contributes a small -0.6 to -0.7 kcal mol^{-1} fraction, but these add up to a substantial total. The induction term for the nonpolar guest core polarizing the cucurbituril is quantitatively negligible (-0.3 kcal mol^{-1}), as expected.

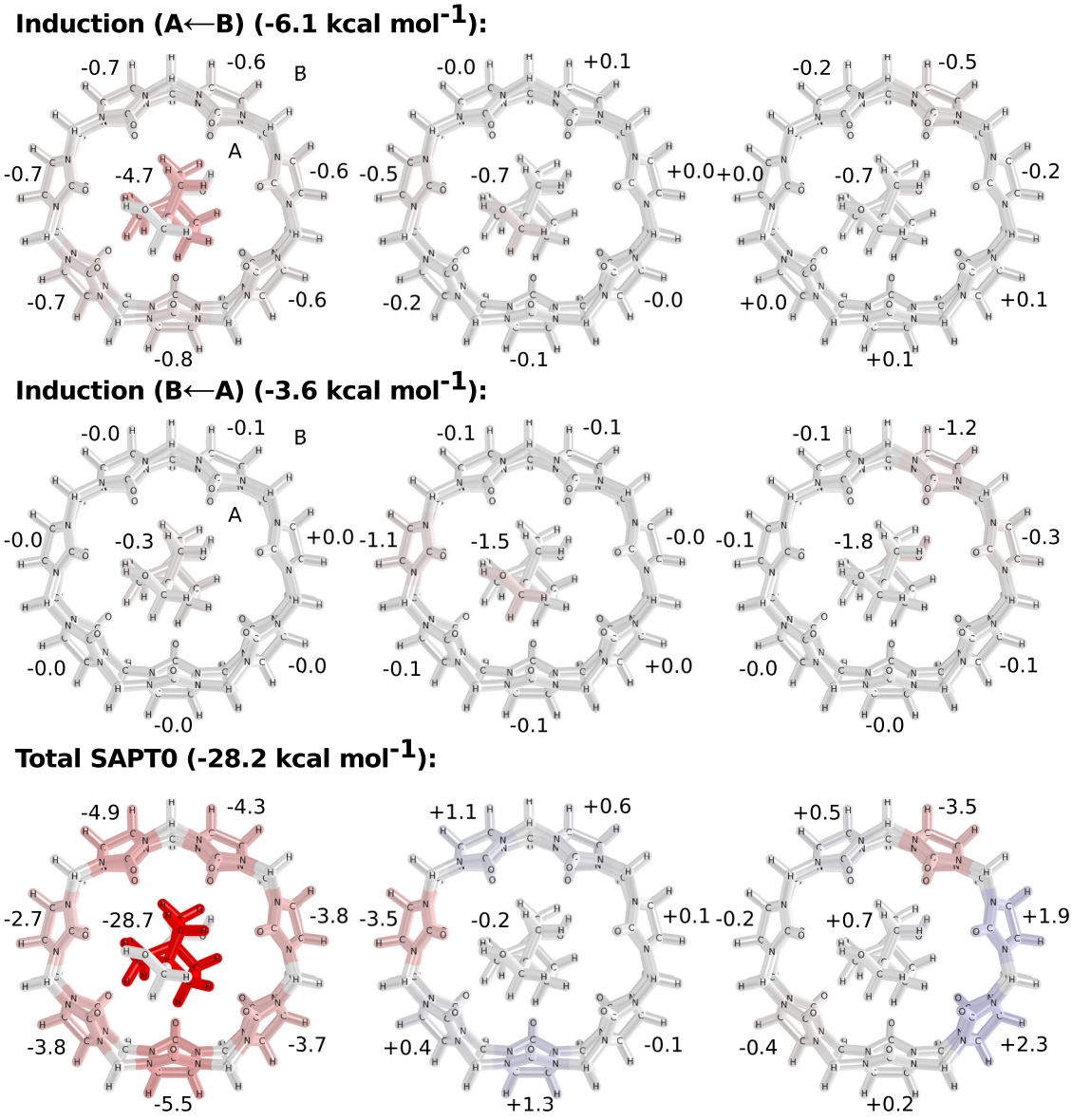

Figure 42: Cucurbit[7]uril···bicyclo[2.2.2]octane-1,4-dimethanol inclusion complex SAPT0/jun-cc-pVDZ induction and total terms partitioned by collapsed F-SAPT scheme discussed in the main text. Left: contributions to Core. Middle: contributions to MeOH1. Right: contributions to MeOH2. Red indicates attraction, blue indicates repulsion, white indicates no net interaction. The color scales saturate at ±20 kcal mol^{-1}. Small numbers indicate functional group electrostatic contributions in kcal mol^{-1}.

The surprising finding from F-SAPT is the quantitatively negligible net contribution from the methanol groups (−0.2 kcal mol^{-1} and +0.7 kcal mol^{-1} for MeOH1 and MeOH2, respectively). Evidence of hydrogen bonding with the nearest carbonyl

units is present in favorable net contributions in the electrostatics, exchange, induction (B←A), and dispersion terms for MeOH1$\cdots$CB1 and MeOH2$\cdots$CB6. However, this is essentially wholly canceled by unfavorable interactions in electrostatics and exchange from other cucurbituril units. The unfavorable electrostatic contributions are strongest on the next two cucurbituril units clockwise from the closest contact, and probably arise from clash between the hydroxyl and carbonyl lone pairs. The unfavorable exchange contributions appear counterclockwise from the contact in MeOH1, and on the two adjacent cucurbituril units to the contact in MeOH2, an thus are more dependent on the distance between the host and MeOH unit than was the case for electrostatics. Note that some of the unfavorable exchange contributions can be attributed to the linking CH_2 units in the methanols. However, repeating the collapsed F-SAPT analysis with partitioning at the level of hydroxyl groups and bicyclo[2.2.2]octane-1,4-CH_2 indicates that this is a minor effect. In the latter analysis, the two hydroxyls contribute total interaction strengths of -1.8 kcal^{-1} and -0.8 kcal^{-1} overall, e.g., much less than one hydrogen bond each. Thus it is apparent that the presence of neutral, polar methanol or hydroxyl units do not quantitatively stabilize the complex: a favorable dipole-dipole interaction does occur with the carbonyl of the nearest cucurbituril unit, but this is canceled by unfavorable electrostatic and exchange interations with other cucurbituril units. This may help to explain the enhanced binding affinities generally observed with cationic guests. A guest with cationic side substituents can adopt a geometry in which the cationis are in the middle of the ring of carbonyls on both sides of the cucurbituril, interacting favorably with all carbonyls. This is not possible for neutral side substituents possessing a dipole oriented in the plane of the ring of carbonyls, for which a single favorable dipole-dipole interaction will be largely canceled by unfavorable dipole-dipole interactions with the rest of the ring. An additional, more subtle effect may also be relevant: the symmetry-breaking hydrogen-bonded association of a neutral, polar side substituent

to any one cucurbituril unit presents a barrier to free rotation of the guest within the host. This barrier to rotation may present an entropic penalty relative to a guest with a charged side substituent which does not symmetry break.

It is worth noting that the collapsed cut-and-cap, A-SAPT, and F-SAPT analyses for this system (presented in the supplementary material) all produce quite similar partitions for this system. For A-SAPT, this is serendipitous: the raw A-SAPT partition summed to the level of CB units and linking CH_2 pairs shows significant oscillations in the electrostatics, but this is attenuated by the even assignment of the links to the neighboring CB units in the collapsed analysis. Moreover, the cut-and-cap analysis deviates from the total SAPT0 limit by 2.3 kcal mol^{-1}, with larger deviations in individual terms, e.g., 5.6 kcal mol^{-1} in exchange.

9.5 Conclusions

In this work, we have detailed an "F-SAPT" modification to the A-SAPT partition which yields an alternative decomposition in terms of pairs of chemical functional groups. A-SAPT provides a partition in the force-field-type picture of atom-pairwise contributions to the interaction energy, and is therefore much more suited to the development of *ab initio* force fields than to chemical understanding of complicated noncovalent interactions. In particular, summing A-SAPT atomic contributions to produce contributions in terms of chemical functional groups can provide non-intuitive results, especially for the electrostatics term, due to the presence of intrinsic charges in the A-SAPT functional groups. By contrast, F-SAPT is formulated entirely in terms of the chemically-intuitive picture of fragments of neutral or integral charge via assignment of whole occupied electrons or protons to fragments, thus ameliorating the charge-charge contaminations present in the A-SAPT electrostatics partition. The F-SAPT partition shows strong resemblence to the chemically intuitive results produced by cut-and-cap fragmentation, but additionally respects the many-body limit of the

SAPT approach. The F-SAPT partition is shown to be useful and robust over such complicated intermolecular interaction motifs as phenol dimer, a proflavine$^+$-DNA intercalation complex, and a cucurbituril host-guest inclusion complex. Additionally, the computational overhead of F-SAPT is very small, so F-SAPT should be tractable on any system for which SAPT is tractable. The largest F-SAPT0/jun-cc-pVDZ result shown in this work (proflavine$^+$-DNA) has 228 atoms and 2845 basis functions. Moreover, we have implemented F-SAPT0 as a minor extension to our A-SAPT0 code, so both analyses may be obtained together.

This work and the preceding one form a powerful toolset for the applications chemist. A-SAPT provides both automatic extraction of atom-pairwise potentials for use in force field development, and black box visualizations of the key contributions in noncovalent interactions. F-SAPT provides a more chemically robust assignment of the contributions of individual noncovalent contacts, for use in understanding and designing complex intermolecular interactions. Future improvements to the methodology include sensitivity testing over a broad range of interaction motifs and quasiparticle localization techniques, application to more-accurate SAPT levels, and development of a streamlined workflow for A-SAPT/F-SAPT computation and postprocessing. We should be careful to remember the A-SAPT/F-SAPT are neither unique, nor experimentally observable partitions of the interaction energy components, but so far they seem to be providing robust and deep insight into the origins of complicated noncovalent contacts. Even at the current level of completeness, we are already beginning to use A-SAPT/F-SAPT to probe the origins of intermolecular interactions in drug-ligand interactions, organic materials stacking, dispersion-assisted hydrogen bonding, and substituent effects in stacked benzene dimers.

CHAPTER X

ON THE ORIGINS OF THE SUBSTITUENT EFFECT IN BENZENE DIMER

The following is adapted from Ref. 221.

10.1 Abstract

The influence of substituents on π-stacking interactions has previously been explained by two competing hypotheses: a nonlocal effect in which the tuning of the π density by the substituent alters the interaction (the Hunter-Sanders picture), or a local effect in which the direct interaction of the added substituent and changed polarity of the phenyl-substituent σ bond alter the interaction (the Wheeler-Houk picture). In this work, we apply the recently developed functional-group partition of symmetry adapted perturbation theory (F-SAPT) to directly quantify these two effects *in situ*. The results show that both pictures do contribute to the change in interaction energy, but that the Wheeler-Houk picture is usually dominant.

10.2 Introduction

Noncovalent π-stacking interactions are ubiquitous in nature, and have been found to play critical roles in the stabilization of DNA,[43] the strength and specificity of drug-protein interactions,[261] and myriad other areas of supramolecular chemistry.[262] These interactions are strongly modulated by the presence of substituents around the π ring,[128,54] leading to a powerful avenue for the rational design of noncovalent contacts. However, the physical mechanism by which the substituents tune the interaction energy has been a source of considerable exploration and debate in the literature.[317] The long-held hypothesis of Hunter and Sanders (HS) posits that the

substituent principally serves to inductively add or withdraw density from the π system of the substituted benzene, thus changing the electrostatic interaction.[128,129] This effect is nonlocal and indirect, as the substituent affects the second benzene only through the π system of the substituted benzene. A second hypothesis proposed by Wheeler and Houk (WH) posits that the direct interaction between the substituent and second benzene, plus the change due to the altered polarity of the linking phenyl-substituent σ bond (which is of C-X topology in Ph-X vs. C-H in Ph-H) wholly explains the influence of the substituent on the interaction energy.[318] This effect is local (confined to the substituent and linking σ bond) and both direct and indirect (direct in the X$\cdots$Bz interaction, indirect in the changed $\sigma\cdots$Bz interaction[316]). A markedly interesting consequence of the WH hypothesis is that the π system is not involved in the substituent effect.

To this point, the evidence presented for these two hypotheses has been admirably thorough, but ultimately circumstantial in nature. The main evidence invoked for the HS picture is the correlation in several experiments[54,46,47] of the substituent effect with the Hammett σ_{m} constant (a loose quantification of the inductive adding- or withdrawing-power of a substituent vs. hydrogen). However, high-level electronic structure computations definitively show that both electron donating and electron withdrawing substituents enhance the interaction energy for gas-phase cofacial benzene dimers.[276,311] Some more recent experiments also support this finding.[319] Additionally, symmetry-adapted perturbation theory[139] (SAPT) shows that significant contributions to the substituent effect are provided by enhanced dispersion interactions in substituted benzene dimers.[277,311] Another interesting consideration is additivity: theoretically[249,316] and, very recently, experimentally,[130] it has been shown that substituent effects are almost wholly additive in benzene dimers with multiple substituents. This finding seems inconsistent with the HS model, given the diminishing flexibility or polarizability expected of the π system as additional substituents

are added. For the WH hypothesis, the primary evidence is the excellent and nearly one-to-one correlation between the substituent effect in Ph-X$\cdots$Bz and the direct interaction energy in an analogous H-X$\cdots$Bz model system which removes the π system of the substituted monomer completely.[318,316] Note however that this evidence, while compelling, is circumstantial: the H-X$\cdots$Bz model has different physics than the true Ph-X$\cdots$Bz system of interest. In particular, the invocation of the H-X$\cdots$Bz model *a priori* precludes a HS contribution. In the real system, it is possible that the HS picture may play some role and that the WH contribution may be different than that observed in the H-X$\cdots$Bz model system (due to the different electronic structure of the X functional group and linking σ bond in Ph-X vs. H-X). Finally, it is of note that the recent and thorough experimental study of Shimizu and co-workers[130] points strongly toward the direct interaction model of the WH picture, but states in the conclusions that "due to the limitation in the number of substituents in our system and the measurement error, we cannot definitively exclude the indirect [substituent effect] model." It is thus apparent that a theoretical tool with the capability to provide a robust and direct quantification of the substituent effect arising from these two pictures would provide a useful addition to this longstanding line of inquiry.

10.3 Theory

10.3.1 Difference F-SAPT Scheme

In this study, we apply the newly-developed functional-group SAPT (F-SAPT) analysis[220] to probe the origins of the substituent effect *in situ* (i.e., directly in the substituted benzene dimers, not in truncated model systems). F-SAPT begins from the formalism of SAPT, which provides an *ab initio* quantification of the electrostatics, exchange, induction/polarization, and dispersion contributions to the intermolecular interaction energy. F-SAPT asserts the picture that the constituent functional groups of weakly interacting monomers are composed of collections of protons and localized

occupied electrons, and then accumulates an effective two-body partition of the SAPT terms according to these groupings of particles. The result is a robust classification of the many-body SAPT interaction energy (and physical components) into effective interactions between pairs of functional groups. With F-SAPT, we can compute the SAPT interaction energy in Ph-X$\cdots$Bz and obtain contributions from the Ph functional group, the X functional group, and the linking σ bond (Link). Likewise, we can compute the SAPT interaction energy in Ph-H$\cdots$Bz and obtain contributions from the Ph functional group and the Link fragment (the H proton is part of the Link fragment in this system, so the X functional group contains no particles in this system and its interaction energy contribution is zero). The difference in interaction energies between these two systems, $\Delta E_{\rm int} \equiv E_{\rm int}({\rm Ph-X}) - E_{\rm int}({\rm Ph-H})$, is the substituent effect we seek, and now can be partitioned into contributions from the Ph, X, and Link fragments, e.g., $\Delta E_{\rm int}^{\rm Ph} \equiv E_{\rm int}^{\rm Ph}({\rm Ph-X}) - E_{\rm int}^{\rm Ph}({\rm Ph-H})$. The quantity $\Delta E_{\rm int}^{\rm Ph}$ will contain the difference in interaction energy due to the polarization of the Ph fragment by the substituent, and is the indirect HS contribution to the substituent effect. The quantity $\Delta E_{\rm int}^{\rm X}$ will contain the direct interaction due to the substituent, and is the direct WH contribution to the substituent effect. The quantity $\Delta E_{\rm int}^{\rm Link}$ will contain the difference in interaction energy due to the altered polarity of the linking σ bond, and is the indirect WH contribution to the substituent effect. We refer to this process of partitioning a difference in interaction energies as a "difference F-SAPT." See Figure 43 for a schematic of this computational experiment. To simplify the analysis, in this study we do not permit any contribution to the substituent effect due to ring relaxation or intermolecular displacement: these effects are higher order and not directly encapsulated within either the WH or HS picture.

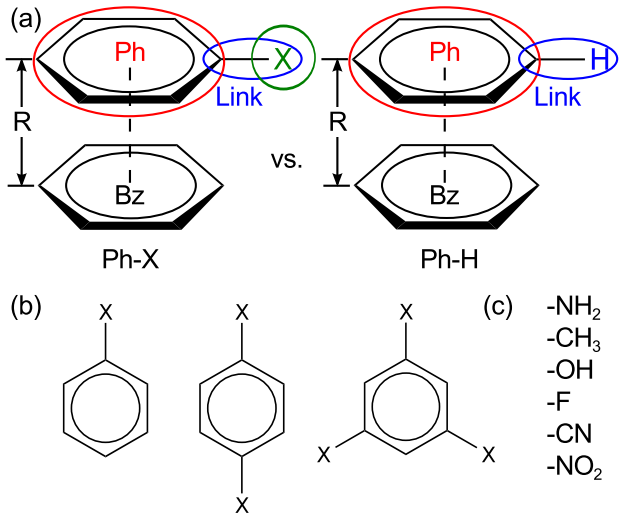

Figure 43: (a) Schematic of difference F-SAPT partition of Ph-X and Ph-H to provide Ph, Link, and X contributions. (b) Substituent placements used in this work. (c) Substituents used in this work, ordered by σ_{m} from most donating (top) to most withdrawing (bottom).

10.3.2 Computational Details

For the computational methodology, we use F-SAPT0/jun-cc-pVDZ with density fitting accelerations in the Psi4 program,[298] as described in our previous papers.[222,220] The underlying SAPT0/jun-cc-pVDZ methodology has been found to be accurate relative to coupled-cluster complete-basis set [CCSD(T)/CBS] benchmarks for interaction energies.[124,214] To provide the partition of the electrons into functional groups, we use the robust intrinsic bond orbital (IBO) localization methodology of Knizia.[149] To examine the origins of substituent effects across a broad range of withdrawing/donating substituents, we use substituted benzene dimers with -NH_2, -CH_3, -OH, -F, -CN, and -NO_2 substituents, and study singly-, doubly-, and triply-substitued dimers, as seen in Figure 43. The Ph-X monomers are optimized at the DF-MP2/aug-cc-pVDZ level of theory, and the corresponding Ph-H monomers are obtained by removing X and placing a capping proton a distance of 1.094 Å along the linking C-X bond. The Ph-X$\cdots$Bz and Ph-H$\cdots$Bz dimers are obtained by aligning

the rings of the monomers and then separating by a distance R. For the results in the primary manuscript, a uniform R of 3.90 Å from the Bz$\cdots$Bz DF-MP2/jun-cc-pVDZ equilibrium is used; additional results from the individual DF-MP2/jun-cc-pVDZ equilibrium separations of each substituted benzene dimer show the same quantitative trends. More details on the computational experiment, including explicit data, results for both choices of R, and sensitivity testing with the aug-cc-pVTZ basis set are available in the Supporting Information. Note that the aug-cc-pVTZ analysis is performed only to validate that there is not significant BSSE contamination in the difference F-SAPT0/jun-cc-pVDZ analysis (e.g., due to the different atoms present in Ph-X vs. Ph-H): F-SAPT0/jun-cc-pVDZ is expected to be much more accurate than F-SAPT0/aug-cc-pVTZ due to systematic error cancelation in the dispersion term,[124,214] and will be used for all chemical analysis in this work. Another point of interest is the possibility of multiple local minima in the IBO localization schemes that all partitions rely on: we have investigated this in the Supporting Information and found that the partitions are all quite robust to different localizations of the π orbitals of Ph-X.

10.4 Results and Discussion

The difference F-SAPT results for singly-substituted monomers at $R = 3.90$ Å are shown in Figure 44. Here, we will comment on only the singly-substituted cases, and discuss the transferability to multiply-substituted cases below. The first major finding from difference F-SAPT is that *both* the HS and WH pictures can contribute to the substituent effect, depending on the term and substituent. The second major finding is that the WH picture usually dominates the total. For electrostatics, the total contribution tracks well in sign and magnitude with σ_{m}, but is dominated by WH contributions, suggesting that the local electrostatic environments of the X+Link groups govern this term. HS electrostatic contributions are present, but, contrary to

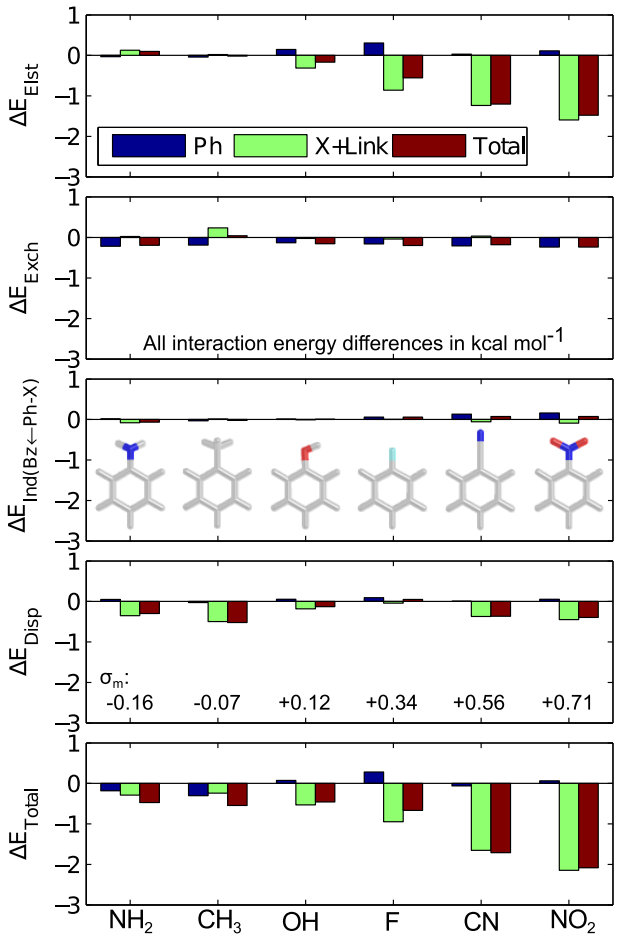

Figure 44: Difference F-SAPT analysis at F-SAPT0/jun-cc-pVDZ ($R = 3.90$ Å, from benzene dimer equilibrium) for singly-substituted benzene dimers. $\Delta E_{\mathrm{ind(Ph-X\leftarrow Bz)}}$ is negligible and therefore not shown.

expectations, are attractive for electron-donating substituents, repulsive for electron withdrawing substituents, and the corresponding magnitudes do not track well with σ_{m}. This suggests that HS-type electrostatic contributions are not simple differential quadrupole-quadruple interactions, but are more likely related to differential charge penetration (as evidenced by the opposite sign vs. that predicted by the quadrupole-quadrupole picture) and are quite susceptible to the exact polarization of the Ph unit by the substituent. For exchange, the total contribution is relatively constant across all substituents, and is dominated by attractive HS contributions. Difference density analysis of the Ph unit (see Supporting Information) indicates that nonzero

contraction of the π clouds occurs in some form for all substituents, leading to the attractive HS contributions seen here. However, the exact magnitude and topology of the Ph polarization varies markedly with the substituent considered, implying that the nearly constant magnitudes of the HS contributions to exchange are somewhat of a coincidence. WH contributions to exchange are negligible, with the exception of a repulsive contribution from -CH_3 due to direct steric clash. For induction, only the Ph-X monomer polarizing the Bz monomer produces a non-negligible effect (for the other direction, the underlying $E_{\mathrm{ind(Bz\leftarrow Bz)}}$ induction term is only on the order of 0.1 kcal mol^{-1} in magnitude to begin with, and does not change much upon substitution). The strength and sign of this induction contribution varies with σ_{m}, and is characterized by opposing HS and WH contributions For dispersion, the total is generally attractive and uncorrelated with σ_{m}, and is almost wholly dominated by the WH picture (i.e., direct dispersion between the substituent and the opposing benzene). Overall, electrostatics and then dispersion are the dominant terms, followed by much smaller contributions from exchange and induction. Since the WH contributions dominate the electrostatics and dispersion terms, the total substituent effect is usually governed by the WH picture. In fact, the only substituent for which the HS contribution is greater is -CH_3, for which the larger individual exchange and dispersion contributions from the WH picture nearly cancel.

Another consideration is transferability of the results to multiply substituted cases. The full difference F-SAPT0/jun-cc-pVDZ partitions for singly-, doubly-, and triply-substituted cases are presented in Figures S2-S4 of the Supporting Information, and demonstrate remarkable additivity of all nontrivial contributions, as shown succinctly in Figure S14. Particularly interesting is that *both* the WH and HS contributions are additive: this suggests that HS effects are possibly dominated by local polarizations of the phenyl density around added substitutents (e.g., the accumulation of phenyl charge around the ring carbon the substituent is attached to, see below), rather than

large scale changes of the phenyl density throughout the phenyl ring. In fact, the only term which shows noticable nonadditivity is induction, which is expected for this nonlinear term.

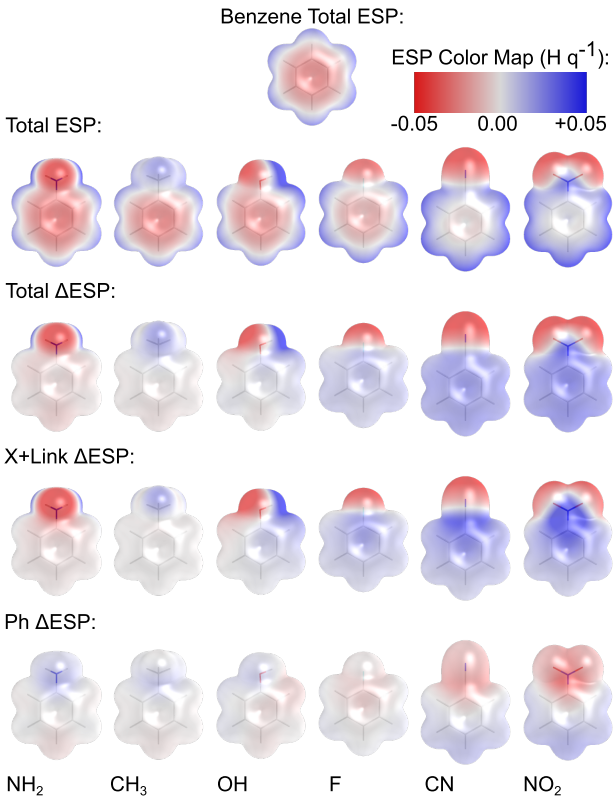

Figure 45: Difference ESP analysis at Hartree-Fock/jun-cc-pVDZ for singly-substituted benzene monomers. The ESP is mapped onto the isodensity surface of the substituted benzene monomer at 0.002 electron bohr^{-3}. For difference ESPs, red indicates areas with enhanced nucleophilicity relative to benzene, blue indicates areas with reduced nucleophilicity relative to benzene.

Further insights are provided by plots of differences in the electrostatic potential (ESP) between substituted and unsubstituted benzene monomers, partitioned in the same manner as in F-SAPT. These results are available in Figure 45 for Hartree-Fock/jun-cc-pVDZ methodology (corresponding to the treatment of electrostatics in SAPT0/jun-cc-pVDZ). As with the F-SAPT results, both WH and HS contributions occur, but the WH contributions usually dominate the change in the molecular ESP. The Ph contributions to the difference ESP track quite well in sign and strength

and magnitude with σ_m, but are relatively weak overall. The X+Link contributions are much stronger, and also quite localized, producing strong electric fields in the neighborhood of the substituent. Additional plots of partitioned differences in the density between substituted and unsubstitutend benzene monomers are available in the Supporting Information. These show that the extra electrons of the substituent and altered polarity of the linking σ bond are by far the dominant effects in the difference density, supporting the local interaction model. With regard to the difference density of the Ph unit (much smaller than X+Link), a more complicated picture emerges than that predicted by σ_m alone. The principal effect for most substituents is a concentrated accumulation of density around the ring carbon involved the linking σ bond (especially in the σ network around this carbon), presumably to stabilize the withdrawal along the altered linking σ bond. This accumulation of density to cover the polarization of the linking σ bond is necessarily mirrored by a depletion of density elsewhere in the Ph system, and usually comes from diffuse contributions from the π system, though there is a complicated and nonuniform nodal structure of the π difference density around the ring. The possible exception to these observations is the -CH_3 substituent, which exhibits only very mild link bond polarization and π withdrawal. With this exception identified, a very interesting picture emerges: essentially all substituents exhibit some contraction of the π system to cover the polarization of the linking σ bond. This supports the observation of always attractive HS contributions to the exchange terms above. The difference ESP for the Ph unit seems to track with where the depletion comes from in the ring (e.g., from the π system above the linking carbon in NH_2, or from other carbons in the Ph unit for electron withdrawing substituents), partially explaining both the trends in the Ph difference ESP and the complicated HS contributions to the electrostatics. The main conclusions from these analyses are (1) the local changes caused by the addition of X and polarization of the linking σ bond dominate both the ESP and density differences

(2) the minor Ph contributions to the difference ESP track well with the magnitude and sign of σ_{m}, but the minor Ph contributions to the difference density do not (and are quite complicated overall), indicating that σ_{m} is not a good measure of π electron donation or withdrawal caused by the substituent.

Overall, the difference F-SAPT, ESP, and density analyses indicate that the HS and WH pictures both contribute to the substituent effect in sandwich benzene dimers. The HS picture is usually dominant in the exchange term and also contributes to electrostatics and induction, but is much more complicated than the simple picture of π donation or withdrawal usually advocated in discussing this picture. The WH picture dominates the larger electrostatics and dispersion terms, and therefore is usually the dominant contributor to the total substituent effect. The only exception is if direct steric interactions from the WH term mask attractive WH contributions from electrostatics and/or dispersion, as occured for -CH_3. Moreover, these conclusions transfer quite readily to doubly- and triply-substituted cases, indicating additivity of *both* WH and HS effects as the number of substituents is increased. In any case, this *in situ* F-SAPT study of substituent effects clearly indicates the broad qualitative correctness of the WH model, albeit with the caveat that minor HS effects do occur.

CHAPTER XI

INTRAMOLECULAR SYMMETRY-ADAPTED PERTUBATION THEORY

The following is adapted from Ref. 215.

11.1 Abstract

We develop a simple methodology for the computation of symmetry-adapted perturbation theory (SAPT) interaction energy contributions for intramolecular noncovalent interactions. In this approach, the local occupied orbitals of the total Hartree-Fock (HF) wavefunction are used to partition the fully interacting system into three chemically identifiable units: the noncovalent fragments A and B, and a covalent linker C. Once these units are identified, the noninteracting HF wavefunctions of the fragments A and B are separately optimized while embedded in the HF wavefunction of C, providing the dressed zeroth order wavefunctions for A and B in the presence of C. Standard two-body SAPT is then applied between the relaxed wavefunctions for A and B, particularly SAPT0 in this work. This Intramolecular SAPT (ISAPT) procedure is found to be remarkably straightforward and efficient, as evidenced by example applications ranging from diols to hexaphenyl-ethane derivatives.

11.2 Introduction

The quantification and characterization of noncovalent interactions is an important task for theoretical chemistry. Such interactions may occur in an intermolecular sense, e.g., in the hydrogen bonding in two water molecules, wherein no covalent links exist between the two molecules of interest, or in an intramolecular sense, e.g., in the formation of hydrogen bonds between covalently linked hydroxyl groups. In

the former case, the intermolecular interaction energy and noninteracting zeroth-order wavefunctions of the monomers are easily-defined quantities, and thus much success has been obtained in developing methodology for intermolecular interactions. Example techniques for this paradigm includes supermolecular interaction energies, energy decomposition analysis (EDA),[147] absolutely-localized molecular orbital techniques (ALMO),[13] and symmetry-adapted perturbation theory (SAPT).[138,139] The latter is particularly well developed, and has the laudable property that it is formally completable to the full *ab initio* limit. We will therefore focus wholly on SAPT in this work.

By contrast, methodology for the characterization of intramolecular interactions is virtually nonexistent. The main reason for this deficiency is that it has proven quite difficult to isolate sensible zeroth-order wavefunctions for the intramolecular fragments before the mutual interaction is activated. For a canonical example, when one looks at 2,4-pentadiol (Figure 46), it is fairly obvious that there should be an intramolecular hydrogen bond between the two hydroxyl groups (fragments A and B), and that this should be largely but not exactly like the intermolecular hydrogen bond between two water molecules. However, the presence of the carbohydrate backbone (the linker C) markedly complicates the task of choosing a noninteracting wavefunction for A and B. On the one hand, it is difficult to determine which particles (electrons and protons) should be assigned to A, B, and C; an improper choice can yield zeroth-order wavefunctions for A and B with intrinsic charges, that may overwhelm the subsequent A$\cdots$B analysis. On the other hand, even if a chemically reasonable set of particles is used in the partition, the robust production of zeroth-order wavefunctions for A and B has been elusive. While the A$\cdots$B interaction is weak and thus treatable by perturbation theory (e.g., SAPT), the A$\cdots$C and B$\cdots$C interactions are extremely strong, i.e., covalent bonds! Thus, to elucidate the noncovalent part of the interaction, the A$\cdots$C and B$\cdots$C interactions and orthogonality

constraints must be maintained throughout the production of the zeroth-order wavefunctions. One approach to address these considerations has been recently reported,[92] using the chemical Hamiltonian approach (CHA).[192] While initial tests were promising, the CHA approach is based on assignment of the basis functions to atoms, and thus is not well-defined at the complete basis set limit. Moreover, the CHA approach is more complicated than that proposed here.

In this work, we demonstrate that a simple, chemically reasonable, and computationally robust procedure exists for the partitioning and determination of the zeroth-order wavefunctions of A and B, and that standard SAPT can be immediately adapted to this choice of reference.

11.3 Theory

11.3.1 I-SAPT Scheme

To frame the discussion, consider 2,4-pentadiol, depicted in Figure 46(a). We seek to develop a SAPT procedure to elucidate the strength and nature of the A$\cdots$B interaction between the two hydroxyl groups, within the linking pentane C.

The first task is to assign particles (electrons and protons) to the fragments A and B, and the linker C. If the complete HF wavefunction is localized by a robust orbital localization methodology, the chemical outline of the system becomes immediately obvious. For each hydroxyl, we identify the O 1s core orbital, the O-H σ orbital, and the two O lone pairs. For the pentane linker, we identify the five C 1s core orbitals, the four C-C σ bonds and the ten C-H σ bonds. Further, two linking C-O σ bonds are present (one between A and C, and one between B and C). For the moment, let us assign these linking σ bonds to the linker C; other choices will be discussed below. For the nuclei, we assign protons as in our functional-group SAPT (F-SAPT) partition,[220] so as to avoid charge oscillation between the fragments. That is, we initially assign all protons on the atoms of each fragment A, B, or C to that

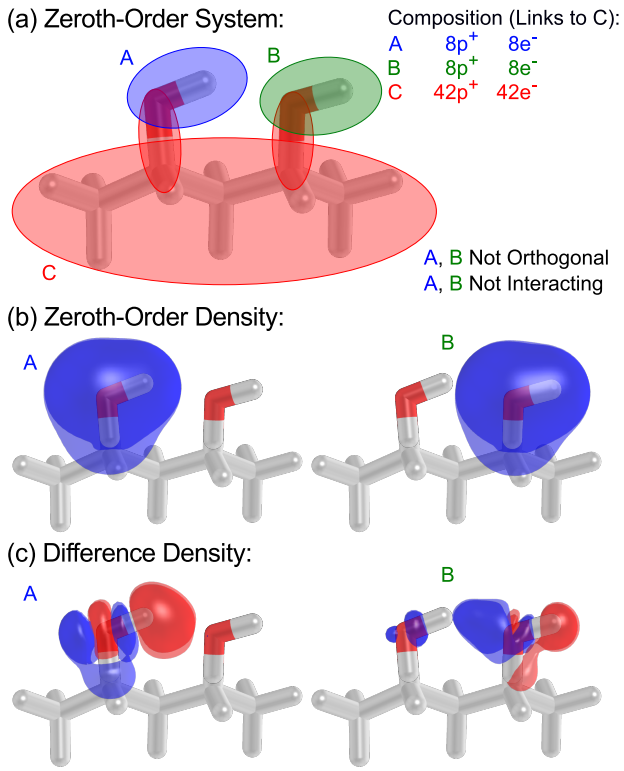

Figure 46: (a) Schematic of the zeroth-order system for 2,4-pentadiol for the choice of linking σ bonds assigned to fragment C. (b) Zeroth-order density field for 2,4-pentadiol (links to C) superimposed with the zeroth-order density field for the corresponding water dimer. Isosurfaces at 0.004 electron bohr^{-3}. (c) Difference density fields superimposed as in (b) showing where density depletes (red) or accumulates (blue) when the HF interaction between A and B is activated. Isosurfaces at $\pm$0.0008 electron bohr^{-3}.

fragment. Then, we remove one proton from each heavy atom involved in a linking σ bond, and instead group that proton with the fragment owning the linking σ bond, yielding an electrically neutral σ link.

At this point, the particle assignment to the fragments is complete (and, given a list of the atoms in A, B, and C, can be completely automated by examination of the sum of orbital atomic charges). However, all three fragments in the system are fully interacting at a HF level; in particular, the wavefunctions of fragments A and B are interacting and orthogonal. The next task is to provide appropriate zeroth-order wavefunctions with the A$\cdots$B interaction and orthogonality constraint

removed. We might be tempted to find separately noninteracting wavefunctions for A, B, *and* C, but this does not make chemical sense; the infinite-order HF interactions and orthogonality constraints A$\cdots$C and B$\cdots$C are critical to provide the correct topology of the covalent links between these fragments, and must be retained if one is to extract the noncovalent part of the interaction A$\cdots$B later. To obtain the zeroth-order wavefunctions for A and B, each with the interactions with C built in, we invoke HF-in-HF embedding as outlined by Manby and coworkers.[185] That is, we perform a standard HF computation for the protons and electrons assigned to fragment A, but with the Coulomb and exchange potentials of the particles of C provided as a spin-specific external potential $\hat{W}_C^\alpha \equiv \hat{V}_C + \hat{J}_C - \hat{K}_C^\alpha$, and with the occupied orbitals of C deleted from the subspace in which the Fock matrix for A is diagonalized. Here, $\hat{V}_C$ is the Coulomb potential of the protons of C, $\hat{J}_C$ is the Coulomb potential of the occupied electrons of C, and $\hat{K}_C^\alpha$ is the exchange potential of the occupied α electrons of C.

At this point, one obtains a set of relaxed occupied and virtual orbitals and corresponding eigenvalues for A and B, each embedded in the HF wavefunction for the linker C, but noninteracting and nonorthogonal between A and B. Next, one may simply substitute the orbitals, eigenvalues, and electrostatic potentials of the A and B systems into a standard two-body SAPT methodology, and obtain the SAPT analysis of the A$\cdots$B interaction, with C built in as a modification of the vacuum.

The overall ISAPT procedure is summarized as follows:

1. Compute the HF wavefunction of the complete chemical system.

2. Localize the occupied orbitals.

3. Assign local occupied orbitals and protons to A, B, and C by consideration of orbital atomic charges.

4. Relax the HF wavefunction for A to be noninteracting and nonorthogonal to B,

but fully interacting and orthogonal to C (via embedding the HF wavefunction for A in the frozen HF wavefunction for C).

5. Repeat the previous step for B.

6. Perform a standard two-body SAPT analysis between the relaxed HF wavefunctions for A and B.

11.3.2 Computational Details

All computations have been performed in a development version of Psi4,[298] using density-fitted SAPT0 (DF-SAPT0) in the jun-cc-pVDZ basis set,[138,118,214] and appropriate -JKFIT and -MP2FIT auxiliary basis sets. The full system's basis set is used throughout, analogous to a dimer-centered basis set in standard intermolecular SAPT. Localization to identify the occupied orbitals of C is accomplished by the robust intrinsic bond orbital (IBO) methodology of Knizia[149] with the cc-pVTZ cores as the minimal AO basis set. To maintain orthogonality against C in the HF-in-HF embedding procedure, we simply diagonalize the Fock matrix in a reduced subspace with the columns spanning the occupied orbitals of C removed. The $\delta^{(2)}_{\mathrm{HF,r}}$ correction to induction is applied throughout this work, using the appropriate HF interaction energy for the zeroth-order wavefunction for A and B embedded in C.

11.4 Results and Discussion

The chemical soundness of our approach is indicated by Figure 46(b)-(c), which compares the zeroth-order density fields and difference density fields upon activation of the HF interaction between the hydroxyls in 2,4-pentadiol and the corresponding water dimer. The qualitative shapes of the density fields and difference density fields are remarkably similar, with the only deviations occuring where the σ linkers would be (assigned to C in this case). In particular, the key changes caused by the interaction are faithfully reproduced, e.g., the interstitial polarization and charge transfer in B

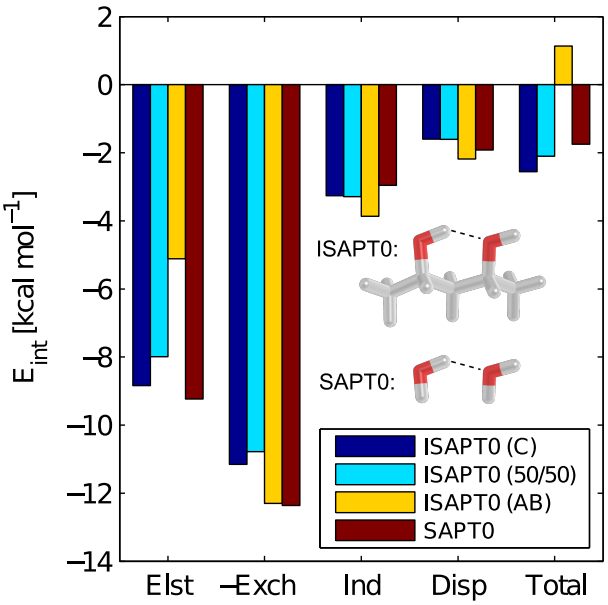

Figure 47: Comparison of ISAPT0/jun-cc-pVDZ for the hydroxyl groups of 2,4-pentadiol to standard intermolecular SAPT0/jun-cc-pVDZ for the corresponding water dimer. Note that $-E_{\text{exch}}$ is plotted to for convenience. ISAPT partitions for the linking σ bonds assigned to C, shared evenly between C and A or B, or assigned to A or B are depicted.

are identical between ISAPT and supermolecular SAPT.

Figure 47 compares the ISAPT0/jun-cc-pVDZ interaction energy components for 2,4-pentadiol with SAPT0/jun-cc-pVDZ for the corresponding water dimer (this latter approach is often referred to as a "cut-and-cap" analysis). To explore the choices of linking σ bond assignment, we have developed a code that can assign the link bonds wholly to C, wholly to A or B, or evenly between C and A or B (in this last case, the link bonds are frozen in second-order terms, and provide only electrostatic potentials to the A$\cdots$B interaction). Overall, the interaction energy components are in excellent agreement with cut-and-cap analysis. The agreement is particularly good for exchange, induction, and dispersion. For these terms, assignment of the linking σ bonds to A or B results in a larger magnitude of each interaction term, as more occupied orbitals are involved in steric overlap or induction/dispersion response. The disappointing case is electrostatics with assignment of the linking σ bonds to A or

B. Here, the protons taken from the C atoms are largely unshielded by the linking σ bond, giving rise to repulsive electrostatic contributions from the $\sigma\cdots\sigma$ multipoles. This contribution is chemically unexpected, but is correct given the choice of particle assignment to fragemnts. We note that such unexpected multipole-multipole electrostatic interactions can also occur in other systems if the linking σ bonds are given to C (see below for such examples as dipole-dipole interactions in methyl-methyl interactions). At present, we can offer no solution to this issue, but note that this anomaly is certainly smaller than the charge-charge or radical contamination inherent in many other choices of molecular partition, and is consistent with our definition of zeroth-order wavefunction.

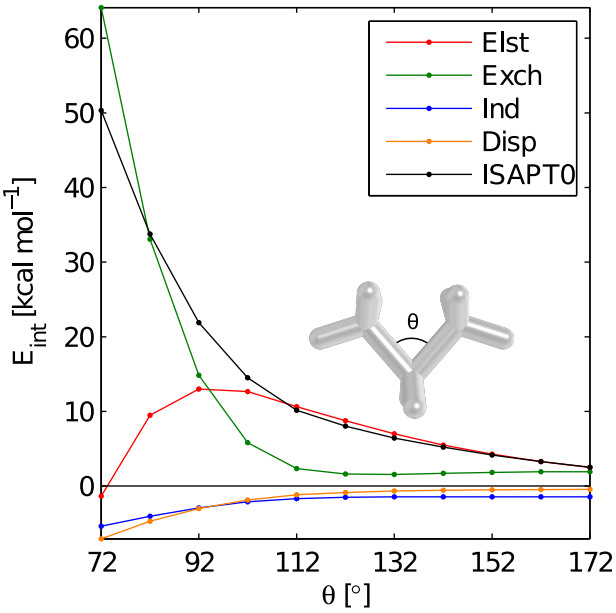

Figure 48: ISAPT0/jun-cc-pVDZ analysis for the interaction between the two terminal methyl groups in propane vs. the C-C-C bond angle θ. Links are assigned to C.

A more torturous test is the noncovalent interaction of the methyl groups in propane, depicted for various C-C-C bond angles in Figure 48. Note that for this system, cut-and-cap analysis is not even qualitatively correct; the capping protons

on the corresponding methyl dimer would be less than 1 Å apart and the resultant steric clash would overwhelm the analysis. Here, ISAPT is able to produce a smooth and largely sensible picture, by defining each of the methyl units A and B as the three C-H σ bonds plus the C 1s, embedded in the field of the CH_2 linking unit. At small angles, exchange dominates the interaction, though significant dispersion, induction, and charge penetration contributions are also apparent. The exchange term exhibits a minimum at around 122°, beyond which exchange increases slightly, consistent with a mild overlap developing between the two previously uninvolved C-H σ bonds (the outermost two C-H σ bonds in the figure). At large angles, the interaction is dominated by repulsive electrostatic interactions, which are consistent with the anti-aligned dipoles present in the zeroth-order wavefunctions for the methyl groups. These dipoles arise from the sum of the three C-H bond dipoles of each methyl group, which are not compensated by the link bond (which is assigned to C). Several alternative choices of link bond assignment are not appropriate in this case: if the link bonds are shared or wholly assigned to A or B, the overlap between the two linking σ bonds causes enormous charge penetration contributions to the electrostatics and induction terms.

A key example of the utility of ISAPT is hexaphenyl ethane (5H), and its t-Bu-substituted derivative (5T), depicted in Figure 49. These compounds have recently been of substantial interest in the experimental literature because the latter has been synthesized and found to have one of the longest C-C σ bonds known (on the order of 1.65 Å), but the former has not been experimentally observed. This is perhaps counterintuitive, as one would expect substantially more steric clash in the t-Bu-substituted variant. Shreiner,[263] and later Shreiner and Grimme[98] have hypothesized that 5T does experience increased steric clash between the two groups on opposite

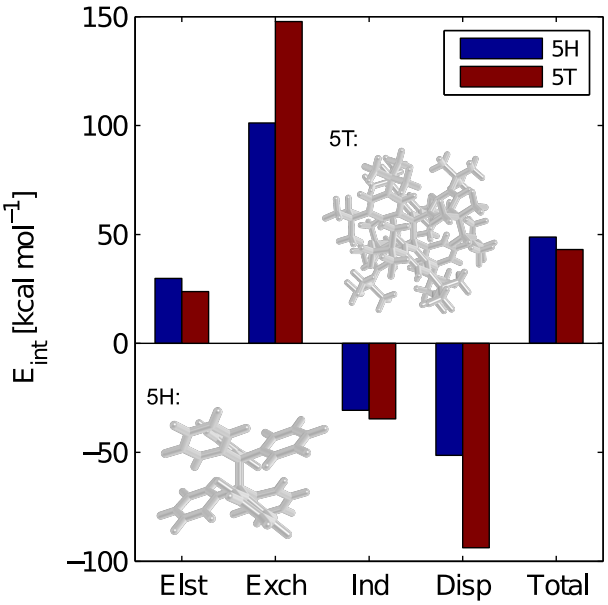

Figure 49: ISAPT0/jun-cc-pVDZ analysis for hexaphenyl-ethane (5H) and t-Bu-substituted derivative (5T).

sides of the long σ bond, but this is countered by an even greater dispersion stabilization afforded by extra t-Bu$\cdots$t-Bu and t-Bu$\cdots$Ph interactions. To this point, however, no direct quantification of the interaction energy components has been possible for this class of systems, principally because they are intractable with a cut-and-cap style analysis. ISAPT can easily handle these systems, by defining the link unit C as the long σ bond, and assigning the remainder to A and B. The results are quite interesting; the increase in exchange repulsion is 47 kcal mol^{-1}, while the increase in dispersion attraction is 41 kcal mol^{-1}. That is, this analysis does not support the hypothesis of a net increase in exchange plus dispersion in 5T. Additional CH-π contacts exist between the t-Bu and Ph groups, which are probably responsible for the more-attractive electrostatic and induction terms in 5T, but this yields only a 9 kcal mol^{-1} more favorable total noncovalent interaction in 5T vs. 5H. Another possible explanation for the enhanced stability of the 5T is destabilization of the radical

disocciation product vs. that of 5H, e.g., by steric clash between the t-Bu groups preventing favorable Ph$\cdots$Ph interactions. We are currently devising analyses to probe this hypothesis. Finally, we note that the overall electrostatics and total interaction energies for both of these systems are positive, which is a consequence of the dipoles present upon removal of the long σ bond between A and B. This effect should almost wholly cancel when one is considering differences between these two systems, as is done here.

It is apparent that ISAPT is a promising approach which allows for quantification of noncovalent interactions in systems for which cut-and-cap analysis is unsuitable. The method is quite easy to implement within an existing two-body SAPT code, and exhibits the same tractability limit. Future developments will involve the application of F-SAPT techniques to the methodology (e.g., to elucidate the fraction of t-Bu involvement vs. Ph$\cdots$Ph interactions in 5T), and efforts to mitigate the chemically unexpected multipole-multipole interactions near the linking σ bonds.

CHAPTER XII

CONCLUSIONS REGARDING PARTITIONED INTERMOLECULAR INTERACTIONS

Effective order-2 partitions of symmetry-adapted perturbation theory interaction components have been developed and utilized to elucidate the spatial origins of complicated noncovalent interactions in large systems. These partitions are realized via a two-stage procedure. In the first stage, an effective two-body interaction is extracted from each of the N-body SAPT perturbation series terms. In developing these effective two-body interactions, one occupied body is selected from each of the two monomers. The use of occupied bodies is preferred, as the occupied space provides the chemical outline of the noninteracting monomers before the interaction is activated. For the cases of SAPT0 or SAPT(DFT), the monomer wavefunctions are wholly specified by the nuclei and occupied electrons, making this an entirely sensible choice. In the second stage, a local set of occupied particles is used to span the occupied space. Several choices for these local occupied particles exist, leading to different flavors of partitioned SAPT.

In atomic SAPT (A-SAPT), the electronic density of the monomers is partitioned into atomic contributions via iterative stockholder analysis (ISA), and atoms are then constructed by joining each electronic atomic density with its corresponding nucleus. For electrostatic terms, atom-pairwise interaction energy contributions are directly constructed via a grid- and auxiliary-based approach which strongly resembles THC. For electronic response terms, contributions are first accumulated at the level of local occupied electrons, and then assigned to atoms via fuzzy orbital atomic charges. A-SAPT is found to produce automatic and useful decompositions, and may provide

quite useful in the direct determination of *ab initio* atom-pair potentials for use in parametrizing force fields. For the purposes of chemical intuition, A-SAPT is less suitable due to an artifact inherent in the definition of atoms in molecules. Under ISA or any other atomization procedure, the resultant set of atoms will necessarily carry partial charges due to electronegativity differences. When functional groups are identified by dividing groups of atoms joined by σ bonds, the functional groups inherit partial charges due primarily to the polarization of the linking σ bonds. This leads to charge-charge terms in the electrostatic partition, muddling the chemically expected dipole-dipole (etc) results. To provide robust order-2 contributions to what chemists typically think of as functional groups, an alternative technique is required.

In functional group SAPT (F-SAPT), order-2 contributions are initially accumulated to the level of protons and localized occupied electrons. For the consideration of localized occupied electrons, the intrinsic bond orbital (IBO) methodology is found to produce a robust set of localized orbitals which clearly outline the generalized Lewis structures of the monomers. From this generalized Lewis structure, net neutral contributions to the functional groups and linking σ bonds can be clearly and automatically determined, given a user-specified list of atoms to provide the functional-group definitions. The linking σ bond contributions can usually be split between the two bridged functional groups, simplifying the subsequent analysis. In the end, F-SAPT strongly resembles the popular cut-and-cap heuristic already in use for partitioning noncovalent interactions, but additionally maintains the many-body limit of the interaction energy. F-SAPT is shown to clearly and succinctly identify the relevant contacts in a number of large and complicated noncovalent interactions.

F-SAPT provides deep insight into a number of difficult and important chemical questions. One example is our recent study of the origins of the substituent effect in substituted sandwich benzene dimers. For years, there has been a substantial debate between two competing hypotheses on this topic. The Hunter-Sanders hypothesis

posits that a substituent principally serves to inductively modulate the π density of the phenyl group of a substituted benzene, leading to a second-order change in the intermolecular interaction between the two phenyl rings. The Wheeler-Houk hypothesis instead asserts that the substituent effect is local, and arises due to the changed electronic environment in the immediate neighborhood of the added substituent. F-SAPT allows us to directly partition the substituent effect into contributions from these two hypotheses, providing the first study of these hypothesis *in media res*. It is found that the Wheeler-Houk effect is dominant, but that nontrivial Hunter-Sanders contributions do exist.

As a spinoff of these efforts, a simple method to quantify intramolecular noncovalent interactions (I-SAPT) has been developed, based on a mixture of F-SAPT-style of division of a molecule into functional groups, followed by Hartree-Fock-in-Hartree-Fock wavefunction embedding to provide sensible zeroth-order wavefunctions for the noninteracting fragment within the field of the linking moeity. This approach allows one to use existing two-body SAPT methodology to quantify directly intramolecular noncovalent interactions. This approach may shed additional light on problems not tractable with standard cut-and-cap SAPT analysis, such as the problem of unexpected stability in t-Butyl substituted hexaphenyl ethane vs. its unsubstituted counterpart.

Many future developments and applications are possible in this general area. Within F-SAPT and A-SAPT, the methodology should transfer directly to more-accurate SAPT(DFT) methodology, and should also be developed for higher-order wavefunction-based SAPT. Within I-SAPT, AI-SAPT and FI-SAPT extensions should be easily possible, and would further enhance analysis of complicated intramolecular interactions. A remaining technical problem is the presence of vexing dipole-dipole terms which muddle the electrostatic interactions in certain cases in I-SAPT - this should be investigated, and, if possible, eliminated. With respect to applications,

many areas might benefit from the application of these methods: A-SAPT for force field parametrization, F-SAPT for understanding of complicated noncovalent interactions in drug design, crytal packing, etc, and I-SAPT for determination of the competition between deformation energy and intramolecular noncovalent interactions in flexible systems. Further afield, this general idea of robust assignment of chemical observables to local spatial origins may be possible, e.g., to quantify the functional group contributions to deformation energy or solvation free energy.

Bibliography

[1] Ababou, A., van der Vaart, A., Gogonea, V., and Jr., K. M. M., "Interaction energy decomposition in protein-protein association: A quantum mechanical study of barnase–barstar complex," *Biophys. Chem.*, vol. 125, no. 1, pp. 221 – 236, 2007.

[2] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, 1964.

[3] Alikhani, M., Fuster, F., and Silvi, B., "What can tell the topological analysis of ELF on hydrogen bonding?," *Struct. Chem.*, vol. 16, no. 3, pp. 203–210, 2005.

[4] Almlöf, J., "Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach," *Chem. Phys. Lett.*, vol. 181, pp. 319–320, 1991.

[5] Almlöf, J., Faegri, K., and Korsell, K., "Principles for a direct SCF approach to LCAO–MO ab-initio calculations," *J. Comp. Chem.*, vol. 3, no. 3, pp. 385–399, 1982.

[6] Aquilante, F., Gagliardi, L., Pedersen, T. B., and Lindh, R., "Atomic Cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency," *J. Chem. Phys.*, vol. 130, p. 154107, 2009.

[7] Aquilante, F., Lindh, R., and Pedersen, T. B., "Unbiased auxiliary basis sets for accurate two-electron integral approximations," *J. Chem. Phys.*, vol. 127, p. 114107, 2007.

[8] Aquilante, F., Pedersen, T. B., and Lindh, R., "Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals," *J. Chem. Phys.*, vol. 126, p. 194106, 2007.

[9] Aquilante, F., Bondo Pedersen, T., Sánchez de Merás, A., and Koch, H., "Fast noniterative orbital localization for large molecules," *J. Chem. Phys.*, vol. 125, no. 17, p. 174101, 2006.

[10] Aquilante, F., Lindh, R., and Pedersen, T. B., "Analytic derivatives for the Cholesky representation of the two-electron integrals," *J. Chem. Phys.*, vol. 129, no. 3, p. 034106, 2008.

[11] Arnstein, S. A. and Sherrill, C. D., "Substituent effects in parallel-displaced π-π interactions," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 2646–2655, 2008.

[12] Avila, G. and Tucker Carrington, J., "Nonproduct quadrature grids for solving the vibrational Schrödinger equation," *J. Chem. Phys.*, vol. 131, no. 17, p. 174103, 2009.

[13] Azar, R. J. and Head-Gordon, M., "An energy decomposition analysis for intermolecular interactions from an absolutely localized molecular orbital reference at the coupled-cluster singles and doubles level," *J. Chem. Phys.*, vol. 136, no. 2, p. 024103, 2012.

[14] Bader, R. F. W., *Atoms in Molecules: A Quantum Theory*, vol. 22 of *International Series of Monographs on Chemistry*. New York: Oxford, 1990.

[15] Bader, R. F. W., "A quantum theory of molecular structure and its applications," *Chem. Rev.*, vol. 91, no. 5, pp. 893–928, 1991.

[16] Bagus, P. S., Hermann, K., and Bauschlicher, C. W., "A new analysis of charge transfer and polarization for ligand-metal bonding: Model studies of Al_4CO and Al_4NH_3," *J. Chem. Phys.*, vol. 80, pp. 4378–4386, 1984.

[17] Bates, D. M., Smith, J. R., and Tschumper, G. S., "Efficient and accurate methods for the geometry optimization of water clusters: Application of analytic gradients for the two-body:many-body QM:QM fragmentation method to $(H2O)_n$, $n = 310$," *J. Chem. Theory Comput.*, vol. 7, no. 9, pp. 2753–2760, 2011.

[18] Bačić, Z. and Light, J. C., "Theoretical methods for rovibrational states of floppy molecules," *Annu. Rev. Phys. Chem.*, vol. 40, pp. 469–498, 1989.

[19] Baye, D. and Heenen, P. H., "Generalised meshes for quantum mechanical problems," *Journal of Physics A: Mathematical and General*, vol. 19, no. 11, p. 2041, 1986.

[20] Becke, A. D., "A multicenter numerical integration scheme for polyatomic molecules," *J. Chem. Phys.*, vol. 88, no. 4, pp. 2547–2553, 1988.

[21] Becke, A. D. and Edgecombe, K. E., "A simple measure of electron localization in atomic and molecular systems," *J. Chem. Phys.*, vol. 92, no. 9, pp. 5397–5403, 1990.

[22] Becke, A. D., "Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals," *J. Chem. Phys.*, vol. 107, no. 20, pp. 8554–8560, 1997.

[23] Becke, A. D. and Johnson, E. R., "Exchange-hole dipole moment and the dispersion interaction," *J. Chem. Phys.*, vol. 122, no. 15, p. 154104, 2005.

[24] Beebe, N. H. F. and Linderberg, J., "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," *Int. J. Quantum Chem.*, vol. 12, pp. 683–705, 1977.

[25] Bell, F., Lambrecht, D., and Head-Gordon, M., "Higher order singular value decomposition in quantum chemistry," *Mol. Phys.*, vol. 108, no. 19-20, pp. 2759–2773, 2010.

[26] Bender, M., Duguet, T., and Lacroix, D., "Particle-number restoration within the energy density functional formalism," *Phys. Rev. C*, vol. 79, p. 044319, Apr 2009.

[27] Benedikt, U., Auer, A. A., Espig, M., and Hackbusch, W., "Tensor decomposition in post-Hartree–Fock methods. I. Two-electron integrals and MP2," *J. Chem. Phys.*, vol. 134, no. 5, p. 054118, 2011.

[28] Benedikt, U., Böhm, K.-H., and Auer, A. A., "Tensor decomposition in post-Hartree–Fock methods. II. CCD implementation," *J. Chem. Phys.*, vol. 139, no. 22, p. 224101, 2013.

[29] Beran, G. J. O. and Nanda, K., "Predicting organic crystal lattice energies with chemical accuracy," *J. Chem. Phys. Lett.*, vol. 1, no. 24, pp. 3480–3487, 2010.

[30] Blaizot, J., "The Skyrme energy functional and RPA calculations," *Physics Letters B*, vol. 60, no. 5, pp. 435 – 438, 1976.

[31] Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A., "Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences," *Int. J. Quantum Chem.*, vol. 113, no. 18, pp. 2110–2142, 2013.

[32] Bondarev, D. A., Skawinski, W. J., and Venanzi, C. A., "Nature of intercalator amiloride-nucleobase stacking. An empirical potential and ab initio electron correlation study," *J. Phys. Chem. B*, vol. 104, pp. 815–822, 2000.

[33] Born, M. and Oppenheimer, R., "Zur quantentheorie der molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.

[34] Boys, S. F., "Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another," *Rev. Mod. Phys.*, vol. 32, pp. 296–299, Apr 1960.

[35] Boys, S. F. and Bernardi, F., "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.*, vol. 19, no. 4, pp. 553–566, 1970.

[36] Boys, S. and Rajagopal, P., "Quantum calculations: Which are accumulative in accuracy, unrestricted in expansion functions, and economical in computation," vol. 2 of *Advances in Quantum Chemistry*, pp. 1 – 24, Academic Press, 1966.

[37] Braess, D. and Hackbusch, W., "Approximation of $1/x$ by exponential sums in $[1, \infty)$," *IMA J. Numer. Anal.*, vol. 25, no. 4, pp. 685–697, October 2005.

[38] Brody, T. and Moshinsky, M., *Tables of transformation brackets for nuclear shell-model calculations*. Gordon and Breach Science Publishers, 1967.

[39] Bultinck, P., Van Alsenoy, C., Ayers, P. W., and Carbó-Dorca, R., "Critical analysis and extension of the Hirshfeld atoms in molecules," *J. Chem. Phys.*, vol. 126, no. 14, p. 144111, 2007.

[40] Bunse-Gerstner, A., Byers, R., and Mehrmann, V., "Numerical methods for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 14, no. 4, pp. 927–949, 1993.

[41] Cardoso, J.-F. and Souloumiac, A., "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, pp. 161–164, Jan. 1996.

[42] Carroll, J. D. and Chang, J. J. *Psychometrika*, vol. 31, p. 279, 1970.

[43] Černý, J., Kabeláč, M., and Hobza, P., "Double-helical → ladder structural transition in the B-DNA is induced by a loss of dispersion energy," *J. Am. Chem. Soc.*, vol. 130, no. 47, pp. 16055–16059, 2008.

[44] Chien, S.-H. and Gill, P. M. W., "SG-0: A small standard grid for DFT quadrature on large systems," *J. Comput. Chem.*, vol. 27, no. 6, pp. 730–739, 2006.

[45] Cížek, J., "On the correlation problem in atomic and molecular systems. calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods," *J. Chem. Phys.*, vol. 45, no. 11, pp. 4256–4266, 1966.

[46] Cockroft, S. L., Hunter, C. A., Lawson, K. R., Perkins, J., and Urch, C. J., "Electrostatic control of aromatic stacking interactions," *J. Am. Chem. Soc.*, vol. 127, pp. 8594–8595, 2005.

[47] Cockroft, S. L., Perkins, J., Zonta, C., Adams, H., Spey, S. E., Low, C. M. R., Vinter, J. G., Lawson, K. R., Urch, C. J., and Hunter, C. A., "Substituent effects on aromatic stacking interactions," *Org. Biomol. Chem.*, vol. 5, pp. 1062–1080, 2007.

[48] Colbert, D. T. and Miller, W. H., "A novel discrete variable representation for quantum mechanical reactive scattering via the S-matrix Kohn method," *J. Chem. Phys.*, vol. 96, no. 3, pp. 1982–1991, 1992.

[49] Condon, E. U., "The theory of complex spectra," *Phys. Rev.*, vol. 36, pp. 1121–1133, Oct 1930.

[50] Contreras-García, J., Johnson, E. R., Keinan, S., Chaudret, R., Piquemal, J.-P., Beratan, D. N., and Yang, W., "NCIPLOT: A program for plotting noncovalent interaction regions," *J. Chem. Theory Comput.*, vol. 7, no. 3, pp. 625–632, 2011.

[51] Contreras-Garca, J. and Recio, J., "Electron delocalization and bond formation under the ELF framework," *Theor. Chem. Acc.*, vol. 128, no. 4-6, pp. 411–418, 2011.

[52] Cooper, V. R., Thonhauser, T., Puzder, A., Schroder, E., Lundqvist, B. I., and Langreth, D. C., "Stacking interactions and the twist of DNA," *J. Am. Chem. Soc.*, vol. 130, pp. 1304–1308, 2008.

[53] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A., "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[54] Cozzi, F., Cinquini, M., Annuziata, R., Dwyer, T., and Siegel, J. S., "Polar/π interactions between stacked aryls in 1,8-diarylnaphthalenes," *J. Am. Chem. Soc.*, vol. 114, no. 14, pp. 5729–5733, 1992.

[55] Dawes, R. and Carrington, T., "A multidimensional discrete variable representation basis obtained by simultaneous diagonalization," *J. Chem. Phys.*, vol. 121, no. 2, pp. 726–736, 2004.

[56] Dawes, R. and Carrington, T., "How to choose one-dimensional basis functions so that a very efficient multidimensional basis may be extracted from a direct product of the one-dimensional functions: Energy levels of coupled systems with as many as 16 coordinates," *J. Chem. Phys.*, vol. 122, no. 13, p. 134101, 2005.

[57] Dawes, R. and Carrington, T., "Using simultaneous diagonalization and trace minimization to make an efficient and simple multidimensional basis for solving the vibrational schrödinger equation," *J. Chem. Phys.*, vol. 124, no. 5, p. 054102, 2006.

[58] de Silva, P. and Corminboeuf, C., "Simultaneous visualization of covalent and noncovalent interactions using regions of density overlap," *J. Chem. Theory Comput.*, vol. 10, no. 9, pp. 3745–3756, 2014.

[59] Dechargé, J. and Gogny, D., "Hartree-Fock-Bogolyubov calculations with the $d1$ effective interaction on spherical nuclei," *Phys. Rev. C*, vol. 21, pp. 1568–1593, Apr 1980.

[60] DePrince, A. E. and Sherrill, C. D., "Accuracy and efficiency of coupled-cluster theory using density fitting/Cholesky decomposition, frozen natural orbitals, and a t1-transformed hamiltonian," *J. Chem. Theory Comput.*, vol. 9, no. 6, pp. 2687–2696, 2013.

[61] Dickinson, A. S. and Certain, P. R., "Calculation of matrix elements for one-dimensional quantum-mechanical problems," *J. Chem. Phys.*, vol. 49, no. 9, pp. 4209–4211, 1968.

[62] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C., and Lundqvist, B. I., "Van der Waals density functional for general geometries," *Phys. Rev. Lett.*, vol. 92, p. 246401, Jun 2004.

[63] Dobaczewski, J. and Dudek, J., "Solution of the Skyrme-Hartree-Fock equations in the Cartesian deformed harmonic oscillator basis I. The method," *Computer Physics Communications*, vol. 102, no. 1, pp. 166 – 182, 1997.

[64] Dobaczewski, J., Satua, W., Carlsson, B., Engel, J., Olbratowski, P., Powaowski, P., Sadziak, M., Sarich, J., Schunck, N., Staszczak, A., Stoitsov, M., Zalewski, M., and Zduczuk, H., "Solution of the Skyrme-Hartree-Fock-Bogolyubov equations in the Cartesian deformed harmonic-oscillator basis.: (vi) hfodd (v2.40h): A new version of the program," *Comp. Phys. Comm.*, vol. 180, no. 11, pp. 2361 – 2391, 2009.

[65] Dobaczewski, J., Stoitsov, M. V., Nazarewicz, W., and Reinhard, P.-G., "Particle-number projection and the density functional theory," *Phys. Rev. C*, vol. 76, p. 054315, Nov 2007.

[66] Duguet, T., Bender, M., Bennaceur, K., Lacroix, D., and Lesinski, T., "Particle-number restoration within the energy density functional formalism: Nonviability of terms depending on noninteger powers of the density matrices," *Phys. Rev. C*, vol. 79, p. 044320, Apr 2009.

[67] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "Applicability of LCAO-X-alpha methods to molecules containing transition-metal atoms - nickel atom and nickel hydride," *Int. J. Quantum Chem. Symp.*, vol. 12, p. 81, 1977.

[68] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "On some approximations in applications of X-alpha theory," *J. Chem. Phys.*, vol. 71, no. 8, pp. 3396–3402, 1979.

[69] Dunlap, B., "Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry," *J. Mol. Struct. THEOCHEM*, vol. 529, no. 13, pp. 37 – 40, 2000.

[70] Dunning, T. H., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, no. 2, pp. 1007–1023, 1989.

[71] Dunning, T. H., "A road map for the calculation of molecular binding energies," *J. Phys. Chem. A*, vol. 104, pp. 9062–9080, 2000.

[72] Edmiston, C. and Ruedenberg, K., "Localized atomic and molecular orbitals," *Rev. Mod. Phys.*, vol. 35, pp. 457–464, Jul 1963.

[73] El-Sherbiny, A. and Poirier, R. A., "An evaluation of the radial part of numerical integration commonly used in DFT," *J. Comp. Chem.*, vol. 25, no. 11, pp. 1378–1384, 2004.

[74] Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Kennedy, M. R., Sherrill, C. D., and Merz, K. M., "Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes," *J. Chem. Theory Comput.*, vol. 7, pp. 790–797, 2011.

[75] Fedorov, D. G. and Kitaura, K., "Pair interaction energy decomposition analysis," *J. Comput. Chem.*, vol. 28, no. 1, pp. 222–237, 2007.

[76] Feyereisen, M., Fitzgerald, G., and Komornicki, A., "Use of approximate integrals in ab initio theory. an application in MP2 calculations," *Chem. Phys. Lett.*, vol. 208, pp. 359–363, 1993.

[77] Fogal, T. and Krüger, J., "Tuvok, an architecture for large scale volume rendering," in *Proceedings of the 15th International Workshop on Vision, Modeling, and Visualization*, November 2010.

[78] Foster, J. P. and Weinhold, F., "Natural hybrid orbitals," *J. Am. Chem. Soc.*, vol. 102, no. 24, pp. 7211–7218, 1980.

[79] Frenking, G. and Krapp, A., "Unicorns in the world of chemical bonding models," *J. Comput. Chem.*, vol. 28, no. 1, pp. 15–24, 2007.

[80] Friesner, R. A., "Solution of self-consistent electronic structure theory equations by a pseudospectral method," *Chem. Phys. Lett.*, vol. 116, p. 39, 1985.

[81] Friesner, R. A., "Solution of the Hartree-Fock equations by a pseudospectral method: Application to diatomic molecules," *J. Chem. Phys.*, vol. 85, p. 1462, 1986.

[82] Friesner, R. A., "Solution of the Hartree-Fock equations for polyatomic molecules by a pseudospectral method," *J. Chem. Phys.*, vol. 86, p. 3522, 1987.

[83] Friesner, R. A., "An automatic grid generation scheme for pseudospectral self-consistent calculations on polyatomic molecules," *J. Phys. Chem.*, vol. 92, p. 3091, 1988.

[84] Friesner, R. A., "New methods for electronic structure calculations on large molecules," *Annu. Rev. Phys. Chem.*, vol. 42, no. 1, pp. 341–367, 1991.

[85] Friesner, R. A., Murphy, R. B., Beachy, M. D., Ringnalda, M. N., Pollard, W. T., Dunietz, B. D., and Cao, Y., "Correlated ab initio electronic structure calculations for large molecules," *J. Phys. Chem. A*, vol. 103, pp. 1913–1928, 1999.

[86] Fukui, K., Yonezawa, T., and Shingu, H., "A molecular orbital theory of reactivity in aromatic hydrocarbons," *J. Chem. Phys.*, vol. 20, no. 4, pp. 722–725, 1952.

[87] Gill, P. M. W. and Chien, S.-H., "Radial quadrature for multiexponential integrands," *J. Comp. Chem.*, vol. 24, no. 6, pp. 732–740, 2003.

[88] Gill, P. M., Johnson, B. G., and Pople, J. A., "A standard grid for density functional calculations," *Chem. Phys. Lett.*, vol. 209, no. 56, pp. 506 – 512, 1993.

[89] Glendening, E. D. and Streitwieser, A., "Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding, strong ionic, and moderate donoracceptor interactions," *J. Chem. Phys.*, vol. 100, no. 4, pp. 2900–2909, 1994.

[90] Gogny, D., "Simple separable expansions for calculating matrix elements of two-body local interactions with harmonic oscillator functions," *Nucl. Phys. A*, vol. 237, no. 3, pp. 399 – 418, 1975.

[91] Golub, G. H. and Welsch, J. H. *Math. Comp.*, vol. 23, pp. 221–230, 1969.

[92] Gonthier, J. F. and Corminboeuf, C., "Exploration of zeroth-order wavefunctions and energies as a first step toward intramolecular symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 140, no. 15, p. 154107, 2014.

[93] Greeley, B. H., Russo, T. V., Mainz, D. T., Friesner, R. A., Langlois, J. M., Goddard, W. A., Donnelly, R. E., and Ringnalda, M. N., "New pseudospectral algorithms for electronic structure calculations: Length scale separation and analytical two-electron corrections," *J. Chem. Phys.*, vol. 101, pp. 4028–4041, 1994.

[94] Griffiths, D. J., *Introduction to Quantum Mechanics; 2nd ed.* Upper Saddle River, NJ: Pearson, 2005.

[95] Grimme, S., "Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies," *J. Chem. Phys.*, vol. 118, no. 20, pp. 9095–9102, 2003.

[96] Grimme, S., "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *J. Comput. Chem.*, vol. 27, no. 15, pp. 1787–1799, 2006.

[97] Grimme, S., Muck-Lichtenfeld, C., and Antony, J., "Analysis of non-covalent interactions in (bio)organic molecules using orbital-partitioned localized MP2," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 3327–3334, 2008.

[98] Grimme, S. and Schreiner, P. R., "Steric crowding can stabilize a labile molecule: Solving the hexaphenylethane riddle," *Angew. Chem., Int. Ed. Engl.*, vol. 50, no. 52, pp. 12639–12642, 2011.

[99] Gygi, F., Fattebert, J.-L., and Schwegler, E., "Computation of maximally localized Wannier functions usingasimultaneous diagonalization algorithm," *Comp. Phys. Comm.*, vol. 155, no. 1, pp. 1 – 6, 2003.

[100] Hampel, C. and Werner, H., "Local treatment of electron correlation in coupled cluster theory," *J. Chem. Phys.*, vol. 104, no. 16, pp. 6286–6297, 1996.

[101] Hapka, M., Żuchowski, P. S., Szcześniak, M. M., and Chałasiński, G., "Symmetry-adapted perturbation theory based on unrestricted Kohn-Sham orbitals for high-spin open-shell van der Waals complexes," *J. Chem. Phys.*, vol. 137, no. 16, p. 164104, 2012.

[102] Hariharan, P. and Pople, J., "The influence of polarization functions on molecular orbital hydrogenation energies," *Theo. Chem. Acc.*, vol. 28, no. 3, pp. 213–222, 1973.

[103] Harris, D. O., Engerholm, G. G., and Gwinn, W. D., "Calculation of matrix elements for one-dimensional quantum-mechanical problems and the application to anharmonic oscillators," *J. Chem. Phys.*, vol. 43, no. 5, pp. 1515–1517, 1965.

[104] Häser, M. and Almlöf, J., "Laplace transform techniques in Møller-Plesset perturbation theory," *J. Chem. Phys.*, vol. 96, pp. 489–494, 1992.

[105] Häser, M. and Ahlrichs, R., "Improvements on the direct SCF method," *J. Comput. Chem.*, vol. 10, no. 1, pp. 104–111, 1989.

[106] Heather, R. W. and Light, J. C., "Discrete variable theory of triatomic photodissociation," *J. Chem. Phys.*, vol. 79, no. 1, pp. 147–159, 1983.

[107] Heisenberg, W., "Über die gesamtstärke der von einem zustande ausgehenden absorptionslinien," *Z. Phys.*, vol. 33, pp. 879–893, 1925.

[108] Heisenberg, W., *Physics and Beyond: Encounters and Conversations*. New York: Harper and Row, 1972.

[109] Helgaker, T., Jørgensen, P., and Olsen, J., *Molecular Electronic Structure Theory*. New York: Wiley, 2000.

[110] Hesselmann, A. and Jansen, G., "The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchangecorrelation potential," *Phys. Chem. Chem. Phys.*, vol. 5, p. 5010, 2003.

[111] Hesselmann, A., Jansen, G., and Schütz, M., "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," *J. Chem. Phys.*, vol. 122, p. 014103, 2005.

[112] Hesselmann, A., "Improved supermolecular second order MøllerPlesset intermolecular interaction energies using time-dependent density functional response theory," *J. Chem. Phys.*, vol. 128, no. 14, p. 144112, 2008.

[113] Hill, J. G. and Platts, J. A., "Local electron correlation descriptions of the intermolecular stacking interactions between aromatic intercalators and nucleic acids," *Chem. Phys. Lett.*, vol. 479, pp. 279–283, 2009.

[114] Hirshfeld, F. L., "Bonded-atom fragments for describing molecular charge densities," *Theor. Chim. Acta*, vol. 44, pp. 129–138, 1977.

[115] Hobza, P. and Šponer, J., "Toward true DNA base-stacking energies: MP2, CCSD(T), and complete basis set calculations," *J. Am. Chem. Soc.*, vol. 124, pp. 11802–11808, 2002.

[116] Hohenstein, E. G., Duan, J., and Sherrill, C. D., "Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers," *J. Am. Chem. Soc.*, vol. 133, pp. 13244–13247, 2011.

[117] Hohenstein, E. G., Jaeger, H. M., Carrell, E. J., Tschumper, G. S., and Sherrill, C. D., "Accurate interaction energies for problematic dispersion-bound complexes: Homogeneous dimers of NCCN, P_2, and PCCP," *J. Chem. Theory Comput.*, vol. 7, pp. 2842–2851, 2011.

[118] Hohenstein, E. G., Parrish, R. M., Sherrill, C. D., Turney, J. M., and Schaefer, H. F., "Large-scale symmetry-adapted perturbation theory computations via density fitting and Laplace transformation techniques: Investigating the fundamental forces of DNA-intercalator interactions," *J. Chem. Phys.*, vol. 135, no. 17, p. 174107, 2011.

[119] Hohenstein, E. G., Kokkila, S. I. L., Parrish, R. M., and Martínez, T. J., "Quartic scaling second-order approximate coupled cluster singles and doubles via tensor hypercontraction: THC-CC2," *J. Chem. Phys.*, vol. 138, no. 12, p. 124111, 2013.

[120] Hohenstein, E. G., Kokkila, S. I. L., Parrish, R. M., and Martńez, T. J., "Tensor hypercontraction equation-of-motion second-order approximate

coupled cluster: Electronic excitation energies in $O(N^4)$ time," *J. Phys. Chem. B*, vol. 117, no. 42, pp. 12972–12978, 2013.

[121] Hohenstein, E. G., Parrish, R. M., and Martínez, T. J., "Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory," *J. Chem. Phys.*, vol. 137, no. 4, p. 044103, 2012.

[122] Hohenstein, E. G., Parrish, R. M., Sherrill, C. D., and Martínez, T. J., "Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions," *J. Chem. Phys.*, vol. 137, no. 22, p. 221101, 2012.

[123] Hohenstein, E. G. and Sherrill, C. D., "Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes," *J. Chem. Phys.*, vol. 132, no. 18, p. 184111, 2010.

[124] Hohenstein, E. G. and Sherrill, C. D., "Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 133, no. 1, p. 014101, 2010.

[125] Hongo, K., Cuong, N. T., and Maezono, R., "The importance of electron correlation on stacking interaction of adenine-thymine base-pair step in B-DNA: A quantum monte carlo study," *J. Chem. Theory Comput.*, vol. 9, no. 2, pp. 1081–1086, 2013.

[126] Høyvik, I.-M., Jansik, B., and Jørgensen, P., "Orbital localization using fourth central moment minimization," *J. Chem. Phys.*, vol. 137, no. 22, p. 224114, 2012.

[127] Huang, Y., Shao, Y., and Beran, G. J. O., "Accelerating MP2C dispersion corrections for dimers and molecular crystals," *J. Chem. Phys.*, vol. 138, no. 22, p. 224112, 2013.

[128] Hunter, C. A. and Sanders, J. K. M., "The nature of π-π Interactions," *J. Am. Chem. Soc.*, vol. 112, no. 14, pp. 5525–5534, 1990.

[129] Hunter, C. A., Lawson, K. R., Perkins, J., and Urch, C. J., "Aromatic interactions," *J. Chem. Soc., Perkin Trans. 2*, pp. 651–669, 2001.

[130] Hwang, J., Li, P., Carroll, W. R., Smith, M. D., Pellechia, P. J., and Shimizu, K. D., "Additivity of substituent effects in aromatic stacking interactions," *J. Am. Chem. Soc.*, vol. 136, no. 40, pp. 14060–14067, 2014.

[131] Hylleraas, E., "Neue berechnung der energie des heliums im grundzustande, sowie des tiefsten terms von ortho-helium," *Z. Phys. A*, vol. 54, no. 5-6, pp. 347–366, 1929.

[132] Isaacs, E. D., Shukla, A., Platzman, P. M., Hamann, D. R., Barbiellini, B., and Tulk, C. A., "Covalency of the hydrogen bond in ice: A direct X-ray measurement," *Phys. Rev. Lett.*, vol. 82, pp. 600–603, Jan 1999.

[133] Ishikawa, T., Mochizuki, Y., Amari, S., Nakano, T., Tokiwa, H., Tanaka, S., and Tanaka, K., "Fragment interaction analysis based on local MP2," *Theor. Chem. Acc.*, vol. 118, no. 5-6, pp. 937–945, 2007.

[134] Jacobi, C. G. J. *Journal für die Reine und Angewandte Mathematikr*, vol. 1846, pp. 51–94, 1846.

[135] Jacobson, L. D. and Herbert, J. M., "An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion," *J. Chem. Phys.*, vol. 134, no. 9, p. 094118, 2011.

[136] Janowski, T., "On the near equivalence of intrinsic atomic orbitals and quasi-atomic orbitals," *J. Chem. Theory Comput.*, vol. 10, no. 8, pp. 3085–3091, 2014.

[137] Jeon, W. S., Moon, K., Park, S. H., Chun, H., Ko, Y. H., Lee, J. Y., Lee, E. S., Samal, S., Selvapalam, N., Rekharsky, M. V., Sindelar, V., Sobransingh, D., Inoue, Y., Kaifer, A. E., and Kim, K., "Complexation of ferrocene derivatives by the cucurbit[7]uril host: a comparative study of the cucurbituril and cyclodextrin host families," *J. Am. Chem. Soc.*, vol. 127, no. 37, pp. 12984–12989, 2005.

[138] Jeziorski, B., Moszynski, R., Ratkiewicz, A., Rybak, S., Szalewicz, K., and Williams, H. L. in *Methods and Techniques in Computational Chemistry: METECC94* (Clementi, E., ed.), vol. B (Medium-Size Systems), p. 79, STEF: Cagliari, 1993.

[139] Jeziorski, B., Moszynski, R., and Szalewicz, K., "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem. Rev.*, vol. 94, pp. 1887–1930, 1994.

[140] Johnson, B. G., Gill, P. M., and Pople, J. A., "A rotationally invariant procedure for density functional calculations," *Chem. Phys. Lett.*, vol. 220, no. 6, pp. 377 – 384, 1994.

[141] Johnson, E. R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A. J., and Yang, W., "Revealing noncovalent interactions," *J. Am. Chem. Soc.*, vol. 132, no. 18, pp. 6498–6506, 2010.

[142] Jung, Y., Sodt, A., Gill, P. M. W., and Head-Gordon, M., "Auxiliary basis expansions for large-scale electronic structure calculations," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 19, pp. 6692–6697, 2005.

[143] Jurečka, P., Šponer, J., Černý, J., and Hobza, P., "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 1985–1993, 2006.

[144] Kendall, R. A. and Fruchtl, H. A., "The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development," *Theor. Chem. Acc.*, vol. 97, pp. 158–163, 1997.

[145] Khaliullin, R. Z., Bell, A. T., and Head-Gordon, M., "Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals," *J. Chem. Phys.*, vol. 128, no. 18, p. 184112, 2008.

[146] Khaliullin, R. Z., Head-Gordon, M., and Bell, A. T., "An efficient self-consistent field method for large systems of weakly interacting components," *J. Chem. Phys.*, vol. 124, no. 20, p. 204105, 2006.

[147] Kitaura, K. and Morokuma, K., "New energy decomposition scheme for molecular-interactions within Hartree-Fock approximation," *Int. J. Quantum Chem.*, vol. 10, pp. 325–340, 1976.

[148] Kitaura, K., Ikeo, E., Asada, T., Nakano, T., and Uebayasi, M., "Fragment molecular orbital method: an approximate computational method for large molecules," *Chem. Phys. Lett.*, vol. 313, no. 34, pp. 701 – 706, 1999.

[149] Knizia, G., "Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts," *J. Chem. Theory Comput.*, vol. 9, no. 11, pp. 4834–4843, 2013.

[150] Ko, C., Malick, D. K., Braden, D. A., Friesner, R. A., and Martínez, T. J., "Pseudospectral time-dependent density functional theory," *J. Chem. Phys.*, vol. 128, no. 10, p. 104103, 2008.

[151] Koch, H., de Meras, A. S., and Pedersen, T. B., "Reduced scaling in electronic structure calculations using Cholesky decompositions," *J. Chem. Phys.*, vol. 118, pp. 9481–9484, 2003.

[152] Koch, H., Christiansen, O., Kobayashi, R., Jørgensen, P., and Helgaker, T., "A direct atomic orbital driven implementation of the coupled cluster singles and doubles (CCSD) model," *Chem. Phys. Lett.*, vol. 228, no. 13, pp. 233 – 238, 1994.

[153] Koch, W. and Holthausen, M. C., *A Chemist's Guide to Density Functional Theory*. New York: Wiley-VCH, 2001.

[154] Kokkila, S., Hohenstein, E. G., Parrish, R. M., Wang, L.-P., and Martínez, T., "Tensor hypercontraction second-order Møller-Plesset perturbation theory: Grid optimization and reaction energies," *submitted*.

[155] Kolář, M., Kubař, T., and Hobza, P., "Sequence-dependent configurational entropy change of DNA upon intercalation," *J. Phys. Chem. B*, vol. 114, pp. 13446–13454, 2010.

[156] Komornicki, A. and Fitzgerald, G., "Molecular gradients and hessians implemented in density functional theory," *J. Chem. Phys.*, vol. 98, no. 2, pp. 1398–1421, 1993.

[157] Korona, T., "First-order exchange energy of intermolecular interactions from coupled cluster density matrices and their cumulants," *J. Chem. Phys.*, vol. 128, no. 22, p. 224104, 2008.

[158] Korona, T., "Second-order exchange-induction energy of intermolecular interactions from coupled cluster density matrices and their cumulants," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 6509–6519, 2008.

[159] Korona, T., "Exchange-dispersion energy: A formulation in terms of monomer properties and coupled cluster treatment of intramonomer correlation," *J. Chem. Theory Comput.*, vol. 5, no. 10, pp. 2663–2678, 2009.

[160] Korona, T., "A coupled cluster treatment of intramonomer electron correlation within symmetry-adapted perturbation theory: benchmark calculations and a comparison with a density-functional theory description," *Mol. Phys.*, vol. 111, no. 24, pp. 3705–3715, 2013.

[161] Korona, T. and Jeziorski, B., "Dispersion energy from density-fitted density susceptibilities of singles and doubles coupled cluster theory," *J. Chem. Phys.*, vol. 128, no. 14, p. 144107, 2008.

[162] Kossmann, S. and Neese, F., "Efficient structure optimization with second-order many-body perturbation theory: The RIJCOSX-MP2 method," *J. Chem. Theory Comput.*, vol. 6, no. 8, pp. 2325–2338, 2010.

[163] Kristensen, K., Ettenhuber, P., Eriksen, J. J., Jensen, F., and Jørgensen, P., "The same number of optimized parameters scheme for determining intermolecular interaction energies," *J. Chem. Phys.*, vol. 142, no. 11, p. 114116, 2015.

[164] Kumar, A., Elstner, M., and Suhai, S., "SCC-DFTB-D study of intercalating carcinogens: Benzo(a)pyrene and its metabolites complexed with the G-C base pair," *Int. J. Quantum Chem.*, vol. 95, pp. 44–59, 2003.

[165] Kutzelnigg, W. and Klopper, W., "Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory," *J. Chem. Phys.*, vol. 94, no. 3, pp. 1985–2001, 1991.

[166] Lacroix, D., Duguet, T., and Bender, M., "Configuration mixing within the energy density functional formalism: Removing spurious contributions from nondiagonal energy kernels," *Phys. Rev. C*, vol. 79, p. 044318, Apr 2009.

[167] Lambrecht, D. S. and Ochsenfeld, C., "Multipole-based integral estimates for the rigorous description of distance dependence in two-electron integrals," *J. Chem. Phys.*, vol. 123, no. 18, p. 184101, 2005.

[168] Langner, K. M., Kedzierski, P., Sokalski, W. A., and Leszczynski, J., "Physical nature of ethidium and proflavine interactions with nucleic acid bases in the intercalation plane," *J. Phys. Chem. B*, vol. 110, pp. 9720–9727, 2006.

[169] Lao, K. U. and Herbert, J. M., "An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method," *J. Chem. Phys.*, vol. 139, no. 3, p. 034107, 2013.

[170] Lawson, C. L., Hanson, R. J., Kincaid, D. R., and Krogh, F. T., "Basic linear algebra subprograms for Fortran usage," *ACM Trans. Math. Softw.*, vol. 5, pp. 308–323, Sept. 1979.

[171] Lebedev, V. I. and Laikov, D. N., "A quadrature formula for the sphere of the 131st algebraic order of accuracy," *Doklady Mathematics*, vol. 59, pp. 477–481, 1999.

[172] Lehtola, S. and Jónsson, H., "Pipek–Mezey orbital localization using various partial charge estimates," *J. Chem. Theory Comput.*, vol. 10, no. 2, pp. 642–649, 2014.

[173] Lewis, G. N., "The atom and the molecule," *J. Am. Chem. Soc.*, vol. 38, no. 4, pp. 762–785, 1916.

[174] Li, S., Cooper, V. R., Thonhauser, T., Lundqvist, B. I., and Langreth, D. C., "Stacking interactions and DNA intercalation," *J. Phys. Chem. B*, vol. 113, pp. 11166–11172, 2009.

[175] Light, J. C., Hamilton, I. P., and Lill, J. V., "Generalized discrete variable approximation in quantum mechanics," *J. Chem. Phys.*, vol. 82, no. 3, pp. 1400–1409, 1985.

[176] Light, J. C. and Carrington, T., "Discrete-variable representations and their utilization," *Adv. Chem. Phys.*, vol. 114, pp. 263–310, 2000.

[177] Lill, J., Parker, G., and Light, J., "Discrete variable representations and sudden models in quantum scattering theory," *Chem. Phys. Lett.*, vol. 89, no. 6, pp. 483 – 489, 1982.

[178] Lillestolen, T. C. and Wheatley, R. J., "Redefining the atom: atomic charge densities produced by an iterative stockholder approach," *Chem. Commun.*, vol. 2008, pp. 5909–5911, 2008.

[179] Lillestolen, T. C. and Wheatley, R. J., "Atomic charge densities generated using an iterative stockholder procedure," *J. Chem. Phys.*, vol. 131, no. 14, p. 144101, 2009.

[180] Lindh, R., Malmqvist, P.-A., and Gagliardi, L., "Molecular integrals by numerical quadrature. I. Radial integration," *Theor. Chem. Acc.*, vol. 106, pp. 178–187, 2001.

[181] Liu, S., Ruspic, C., Mukhopadhyay, P., Chakrabarti, S., Zavalij, P. Y., and Isaacs, L., "The cucurbit[n]uril family: prime components for self-sorting systems," *J. Am. Chem. Soc.*, vol. 127, no. 45, pp. 15959–15967, 2005.

[182] Lotrich, V. F. and Szalewicz, K., "Symmetry-adapted perturbation theory of three-body nonadditivity of intermolecular interaction energy," *J. Chem. Phys.*, vol. 106, no. 23, pp. 9668–9687, 1997.

[183] Lu, W. C., Wang, C. Z., Schmidt, M. W., Bytautas, L., Ho, K. M., and Ruedenberg, K., "Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals," *J. Chem. Phys.*, vol. 120, no. 6, pp. 2629–2637, 2004.

[184] Maehigashi, T., Persil, O., Hud, N. V., and Williams, L. D., "NDB ID: DD0103," *unpublished*.

[185] Manby, F. R., Stella, M., Goodpaster, J. D., and Miller, T. F., "A simple, exact density-functional-theory embedding scheme," *J. Chem. Theory Comput.*, vol. 8, no. 8, pp. 2564–2568, 2012.

[186] Marshall, M. S., Steele, R. P., Thanthiriwatte, K. S., and Sherrill, C. D., "Potential energy curves for cation–π interactions: Off-axis configurations are also attractive," *J. Phys. Chem. A*, vol. 113, pp. 13628–13632, 2009.

[187] Martínez, T. J. and Carter, E. A., "Pseudospectral methods applied to the electron correlation problem," in *Modern Electronic Structure Theory* (Yarkony, D. R., ed.), vol. 2 of *Advanced Series in Physical Chemistry*, pp. 1132–1165, Singapore: World Scientific, 1995.

[188] Martínez, T. J., Mehta, A., and Carter, E. A., "Pseudospectral full configuration interaction," *J. Chem. Phys.*, vol. 97, pp. 1876–1880, 1992.

[189] Martínez, T. J. and Carter, E. A., "Pseudospectral double excitation configuration interaction," *J. Chem. Phys.*, vol. 98, no. 9, pp. 7081–7085, 1993.

[190] Martínez, T. J. and Carter, E. A., "Pseudospectral Møller–plesset perturbation theory through third order," *J. Chem. Phys.*, vol. 100, no. 5, pp. 3631–3638, 1994.

[191] Maurer, S. A., Lambrecht, D. S., Flaig, D., and Ochsenfeld, C., "Distance-dependent Schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. rigorous upper bounds," *J. Chem. Phys.*, vol. 136, no. 14, p. 144107, 2012.

[192] Mayer, I., "Towards a "chemical" Hamiltonian," *Int. J. Quantum Chem.*, vol. 23, no. 2, pp. 341–363, 1983.

[193] McMurchie, L. E. and Davidson, E. R., "One- and two-electron integrals over Cartesian Gaussian functions," *J. Comput. Phys.*, vol. 26, no. 2, pp. 218 – 231, 1978.

[194] Merlot, P., Kjærgaard, T., Helgaker, T., Lindh, R., Aquilante, F., Reine, S., and Pedersen, T. B., "Attractive electron-electron interactions within robust local fitting approximations," *J. Comput. Chem.*, vol. 34, no. 17, pp. 1486–1496, 2013.

[195] Misquitta, A. J., Podeszwa, R., Jeziorski, B., and Szalewicz, K., "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," *J. Chem. Phys.*, vol. 123, p. 214103, 2005.

[196] Misquitta, A. J. and Stone, A. J., "Distributed polarizabilities obtained using a constrained density-fitting algorithm," *J. Chem. Phys.*, vol. 124, no. 2, p. 024111, 2006.

[197] Misquitta, A. J. and Szalewicz, K., "Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers," *J. Chem. Phys.*, vol. 122, no. 21, p. 214109, 2005.

[198] Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A., "Predicting absolute binding free energies to a simple model site," *J. Mol. Biol.*, vol. 371, pp. 1118–1134, 2007.

[199] Moghaddam, S., Inoue, Y., and Gilson, M. K., "Host-guest complexes with protein-ligand–like affinities: Computational analysis and design," *J. Am. Chem. Soc.*, vol. 131, no. 11, pp. 4012–4021, 2009.

[200] Møller, C. and Plesset, M. S., "Note on an approximation treatment for many-electron systems," *Phys. Rev.*, vol. 46, pp. 618–622, 1934.

[201] Moszynski, R., Jeziorski, B., and Szalewicz, K., "Many-body theory of exchange effects in intermolecular interactions. second-quantization approach and comparison with full configuration interaction results," *J. Chem. Phys.*, vol. 100, no. 2, pp. 1312–1325, 1994.

[202] Moszynski, R., Wormer, P. E. S., Jeziorski, B., and van der Avoird, A., "Symmetry-adapted perturbation theory of nonadditive three-body interactions in van der Waals molecules. I. General theory," *J. Chem. Phys.*, vol. 103, no. 18, pp. 8058–8074, 1995.

[203] Müller-Dethlefs, K. and Hobza, P., "Noncovalent interactions: A challenge for experiment and theory," *Chem. Rev.*, vol. 100, no. 1, pp. 143–168, 2000.

[204] Mura, M. E. and Knowles, P. J., "Improved radial grids for quadrature in molecular densityfunctional calculations," *J. Chem. Phys.*, vol. 104, no. 24, pp. 9848–9858, 1996.

[205] Murray, C. W., Handy, N. C., and Laming, G. J., "Quadrature schemes for integrals of density functional theory," *Mol. Phys.*, vol. 78, no. 4, pp. 997–1014, 1993.

[206] Nadig, G., Van Zant, L. C., Dixon, S. L., and Merz, K. M., "Charge-transfer interactions in macromolecular systems: A new view of the protein/water interface," *J. Am. Chem. Soc.*, vol. 120, no. 22, pp. 5593–5594, 1998.

[207] Neese, F., "The ORCA program system," *WIREs: Comp. Mol. Sci.*, vol. 2, no. 1, pp. 73–78, 2012.

[208] Neese, F., Wennmohs, F., Hansen, A., and Becker, U., "Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. a chain–of–spheres algorithm for the HartreeFock exchange," *Chem. Phys.*, vol. 356, no. 13, pp. 98 – 109, 2009.

[209] Nozières, P., *Theory of interacting Fermi systems*. Advanced book classics, Reading, MA: Addison-Wesley, 1964.

[210] Obara, S. and Saika, A., "Efficient recursive computation of molecular integrals over Cartesian Gaussian functions," *J. Chem. Phys.*, vol. 84, no. 7, pp. 3963–3974, 1986.

[211] Olsen, J., Malmqvist, P.-A., Roos, B. O., Lindh, R., and Widmark, P.-O., "A non-linear approach to configuration interaction. The low-rank CI method (LR-CI)," *Chem. Phys. Lett.*, vol. 133, pp. 91–101, 1987.

[212] Papajak, E., Zheng, J., Xu, X., Leverentz, H. R., and Truhlar, D. G., "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," *J. Chem. Theory Comput.*, vol. 7, no. 10, pp. 3027–3034, 2011.

[213] Parker, T. M., Hohenstein, E. G., Parrish, R. M., Hud, N. V., and Sherrill, C. D., "Quantum-mechanical analysis of the energetic contributions to π stacking in nucleic acids versus rise, twist, and slide," *J. Am. Chem. Soc.*, vol. 135, pp. 1306–1316, 2013.

[214] Parker, T. M., Burns, L. A., Parrish, R. M., Ryno, A. G., and Sherrill, C. D., "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," *J. Chem. Phys.*, vol. 140, no. 9, p. 094106, 2014.

[215] Parrish, R. M., Gonthier, J. F., Corminboeuf, C. V., and Sherrill, C. D., "Practical intramolecular symmetry adapted perturbation theory via Hartree-Fock embedding," *in preparation.*

[216] Parrish, R. M., Hohenstein, E. G., Martínez, T. J., and Sherrill, C. D., "Tensor hypercontraction. II. Least-squares renormalization," *J. Chem. Phys.*, vol. 137, no. 22, p. 224106, 2012.

[217] Parrish, R. M., Hohenstein, E. G., Martínez, T. J., and Sherrill, C. D., "Discrete variable representation in electronic structure theory: Quadrature grids for least-squares tensor hypercontraction," *J. Chem. Phys.*, vol. 138, no. 19, p. 194107, 2013.

[218] Parrish, R. M., Hohenstein, E. G., Schunck, N. F., Sherrill, C. D., and Martínez, T. J., "Exact tensor hypercontraction: A universal technique for the resolution of matrix elements of local finite-range n-body potentials in many-body quantum problems," *Phys. Rev. Lett.*, vol. 111, p. 132505, Sep 2013.

[219] Parrish, R. M., Hohenstein, E. G., and Sherrill, C. D., "Tractability gains in symmetry-adapted perturbation theory including coupled double excitations: CCD+ST(CCD) dispersion with natural orbital truncations," *J. Chem. Phys.*, vol. 139, no. 17, p. 174102, 2013.

[220] Parrish, R. M., Parker, T. M., and Sherrill, C. D., "Chemical assignment of symmetry-adapted perturbation theory interaction energy components: The functional-group SAPT partition," *J. Chem. Theory Comput.*, vol. 10, no. 10, pp. 4417–4431, 2014.

[221] Parrish, R. M. and Sherrill, C. D., "Quantum-mechanical evaluation of π–π versus substituent-π interactions in π stacking: Direct evidence for the Wheeler-Houk picture," *J. Am. Chem. Soc.*, vol. 136, no. 50, pp. 17386–17389, 2014.

[222] Parrish, R. M. and Sherrill, C. D., "Spatial assignment of symmetry adapted perturbation theory interaction energy components: The atomic SAPT partition," *J. Chem. Phys.*, vol. 141, no. 4, p. 044115, 2014.

[223] Parrish, R. M., Sherrill, C. D., Hohenstein, E. G., Kokkila, S. I. L., and Martínez, T. J., "Communication: Acceleration of coupled cluster singles and doubles via orbital-weighted least-squares tensor hypercontraction," *J. Chem. Phys.*, vol. 140, no. 18, p. 181102, 2014.

[224] Patkowski, K., Szalewicz, K., and Jeziorski, B., "Third-order interactions in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 125, no. 15, p. 154107, 2006.

[225] Pauling, L., "The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms," *J. Am. Chem. Soc.*, vol. 54, no. 9, pp. 3570–3582, 1932.

[226] Pauling, L. Cornell University Press, 1939.

[227] Perdew, J. P., Parr, R. G., Levy, M., and Balduz, J. L., "Density-functional theory for fractional particle number: Derivative discontinuities of the energy," *Phys. Rev. Lett.*, vol. 49, pp. 1691–1694, Dec 1982.

[228] Pipek, J. and Mezey, P. G., "A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave-functions," *J. Chem. Phys.*, vol. 90, no. 9, pp. 4916–4926, 1989.

[229] Pitoňák, M., Riley, K. E., Neogrády, P., and Hobza, P., "Highly accurate CCSD(T) and DFTSAPT stabilization energies of H-bonded and stacked structures of the uracil dimer," *Chem. Phys. Chem.*, vol. 9, no. 11, pp. 1636–1644, 2008.

[230] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Predicting structure of molecular crystals from first principles," *Phys. Rev. Lett.*, vol. 101, p. 115503, 2008.

[231] Podeszwa, R., Rice, B. M., and Szalewicz, K., "Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles.," *Phys. Chem. Chem. Phys.*, vol. 11, pp. 5512–5518, 2009.

[232] Podeszwa, R. and Szalewicz, K., "Three-body symmetry-adapted perturbation theory based on Kohn-Sham description of the monomers," *J. Chem. Phys.*, vol. 126, no. 19, p. 194101, 2007.

[233] Podeszwa, R., Cencek, W., and Szalewicz, K., "Efficient calculations of dispersion energies for nanoscale systems from coupled density response functions," *J. Chem. Theory Comput.*, vol. 8, no. 6, pp. 1963–1969, 2012.

[234] Poirier, B., "Optimal separable bases and series expansions," *Phys. Rev. A*, vol. 56, pp. 120–130, Jul 1997.

[235] Poirier, B., "Quantum reactive scattering for three-body systems via optimized preconditioning, as applied to the O + HCl reaction," *J. Chem. Phys.*, vol. 108, no. 13, pp. 5216–5224, 1998.

[236] Poirier, B. and Miller, W. H., "Optimized preconditioners for Green function evaluation in quantum reactive scattering calculations," *Chem. Phys. Lett.*, vol. 265, no. 12, pp. 77 – 83, 1997.

[237] Pople, J. A., Gill, P. M., and Johnson, B. G., "Kohn-Sham density-functional theory within a finite basis set," *Chem. Phys. Lett.*, vol. 199, no. 6, pp. 557 – 560, 1992.

[238] Pulay, P., "Convergence acceleration of iterative sequences. the case of SCF iteration," *Chem. Phys. Lett.*, vol. 73, pp. 393–398, 1980.

[239] Purvis, G. D. and Bartlett, R. J., "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.*, vol. 76, pp. 1910–1918, 1982.

[240] Raghavachari, K., Trucks, G. W., Pople, J. A., and Head-Gordon, M., "A 5th-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989.

[241] Raha, K., van der Vaart, A. J., Riley, K. E., Peters, M. B., Westerhoff, L. M., Kim, H., and Merz, K. M., "Pairwise decomposition of residue interaction energies using semiempirical quantum mechanical methods in studies of proteinligand interaction," *J. Am. Chem. Soc.*, vol. 127, no. 18, pp. 6583–6594, 2005.

[242] Raju, R. K., Bloom, J. W. G., and Wheeler, S. E., "Broad transferability of substituent effects in π-stacking interactions provides new insights into their origin," *J. Chem. Theory Comput.*, vol. 9, no. 8, pp. 3479–3490, 2013.

[243] Reed, A. E., Weinstock, R. B., and Weinhold, F., "Natural population analysis," *J. Chem. Phys.*, vol. 83, no. 2, pp. 735–746, 1985.

[244] Rekharsky, M. V., Mori, T., Yang, C., Ko, Y. H., Selvapalam, N., Kim, H., Sobransingh, D., Kaifer, A. E., Liu, S., Isaacs, L., Chen, W., Moghaddam, S., Gilson, M. K., Kim, K., and Inoue, Y., "A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy–entropy compensation," *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 52, pp. 20737–20742, 2007.

[245] Rendell, A. P. and Lee, T. J., "Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks," *J. Chem. Phys.*, vol. 101, pp. 400–408, 1994.

[246] Reynolds, G., Martinez, T. J., and Carter, E. A., "Local weak pairs spectral and pseudospectral singles and doubles configuration interaction," *J. Chem. Phys.*, vol. 105, pp. 6455–6470, 1996.

[247] Ringer, A. L. and Sherrill, C. D., "First principles computation of lattice energies of organic solids: The benzene crystal," *Chem. Eur. J.*, vol. 14, pp. 2542–2547, 2008.

[248] Ringer, A. L. and Sherrill, C. D., "Substituent effects in sandwich configurations of multiply substituted benzene dimers are not solely governed by electrostatic control," *J. Am. Chem. Soc.*, vol. 131, pp. 4574–4575, 2009.

[249] Ringer, A. L., Sinnokrot, M. O., Lively, R. P., and Sherrill, C. D., "The effect of multiple substituents on sandwich and T-shaped π-π interactions," *Chem. Eur. J.*, vol. 12, pp. 3821–3828, 2006.

[250] Ringnalda, M. N., Belhadj, M., and Friesner, R. A., "Pseudospectral Hartree-Fock theory: Applications and algorithmic improvements," *J. Chem. Phys.*, vol. 93, pp. 3397–3407, 1990.

[251] Ringnalda, M. N., Won, Y., and Friesner, R. A., "Pseudospectral Hartree-Fock calculations on glycine," *J. Chem. Phys.*, vol. 92, pp. 1163–1173, 1990.

[252] Riplinger, C., Sandhoefer, B., Hansen, A., and Neese, F., "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," *J. Chem. Phys.*, vol. 139, no. 13, p. 134101, 2013.

[253] Robledo, L. M., "Separable approximation to two-body matrix elements," *Phys. Rev. C*, vol. 81, p. 044312, Apr 2010.

[254] Roeggen, I. and Wisloff-Nilssen, E., "On the Beebe-Linderberg 2-electron integral approximation," *Chem. Phys. Lett.*, vol. 132, pp. 154–160, 1986.

[255] Rolik, Z., Szegedy, L., Ladjánszki, I., Ladóczki, B., and Kállay, M., "An efficient linear-scaling CCSD(T) method based on local natural orbitals," *J. Chem. Phys.*, vol. 139, no. 9, pp. –, 2013.

[256] Ruth Le Sueur, C. and Stone, A. J., "Localization methods for distributed polarizabilities," *Mol. Phys.*, vol. 83, no. 2, pp. 293–307, 1994.

[257] Rybak, S., Jeziorski, B., and Szalewicz, K., "Many-body symmetry-adapted perturbation theory of intermolecular interactions. H2O and HF dimers," *J. Chem. Phys.*, vol. 95, no. 9, pp. 6576–6601, 1991.

[258] Rys, J., Dupuis, M., and King, H. F., "Computation of electron repulsion integrals using the Rys quadrature method," *J. Comput. Chem.*, vol. 4, no. 2, pp. 154–157, 1983.

[259] Saebø, S. and Pulay, P., "Local configuration interaction: An efficient approach for larger molecules," *Chem. Phys. Lett.*, vol. 113, pp. 13–18, 1985.

[260] Saebø, S. and Pulay, P., "Local treatment of electron correlation," *Annu. Rev. Phys. Chem.*, vol. 44, pp. 213–236, 1993.

[261] Salonen, L. M., Ellermann, M., and Diederich, F., "Aromatic rings in chemical and biological recognition: Energetics and structures," *Angew. Chem., Int. Ed. Engl.*, vol. 50, pp. 4808–4842, 2011.

[262] Schneider, H.-J., "Binding mechanisms in supramolecular complexes," *Angew. Chem., Int. Ed. Engl.*, vol. 48, no. 22, pp. 3924–3977, 2009.

[263] Schreiner, P. R., Chernish, L. V., Gunchenko, P. A., Tikhonchuk, E. Y., Hausmann, H., Serafin, M., Schlecht, S., Dahl, J. E. P., Carlson, R. M. K., and Fokin, A. A., "Overcoming lability of extremely long alkane carbon-carbon bonds through dispersion forces," *Nature*, vol. 477, pp. 308–311, 09 2011.

[264] Schrödinger, E., "Quantisierung als eigenwertproblem," *Ann. Phys.*, vol. 384, no. 4, pp. 361–376, 1926.

[265] Schrödinger, E., "An undulatory theory of the mechanics of atoms and molecules," *Phys. Rev.*, vol. 28, pp. 1049–1070, Dec 1926.

[266] Schütz, M., Lindh, R., and Werner, H.-J., "Integral-direct electron correlation methods," *Mol. Phys.*, vol. 96, no. 4, pp. 719–733, 1999.

[267] Scuseria, G. E., Janssen, C. L., and Schaefer III, H. F., "An efficient reformulation of the closed–shell coupled cluster single and double excitation (CCSD) equations," *J. Chem. Phys.*, vol. 89, no. 12, pp. 7382–7387, 1988.

[268] Scuseria, G. E., Scheiner, A. C., Lee, T. J., Rice, J. E., and Schaefer, H. F., "The closedshell coupled cluster single and double excitation (CCSD) model for the description of electron correlation. a comparison with configuration interaction (CISD) results," *J. Chem. Phys.*, vol. 86, no. 5, pp. 2881–2890, 1987.

[269] Shao, Y., Molnar, L. F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S. T., Gilbert, A. T., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P., DiStasio Jr, R. A., Lochan, R. C., Wang, T., Beran, G. J., Besley, N. A., Herbert, J. M., Yeh Lin, C., Van Voorhis, T., Hung Chien, S., Sodt, A., Steele, R. P., Rassolov, V. A., Maslen, P. E., Korambath, P. P., Adamson, R. D., Austin, B., Baker, J., Byrd, E. F. C., Dachsel, H., Doerksen, R. J., Dreuw, A., Dunietz, B. D., Dutoi, A. D., Furlani, T. R., Gwaltney, S. R., Heyden, A., Hirata, S., Hsu, C.-P., Kedziora, G., Khalliulin, R. Z., Klunzinger, P., Lee, A. M., Lee, M. S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E. I., Pieniazek, P. A., Min Rhee, Y., Ritchie, J., Rosta, E., David Sherrill, C., Simmonett, A. C., Subotnik, J. E., Lee Woodcock III, H., Zhang, W., Bell, A. T., Chakraborty, A. K., Chipman, D. M., Keil, F. J., Warshel, A., Hehre, W. J., Schaefer III, H. F.,

Kong, J., Krylov, A. I., Gill, P. M. W., and Head-Gordon, M., "Advances in methods and algorithms in a modern quantum chemistry program package," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 3172–3191, 2006.

[270] Shavitt, I. and Bartlett, R., *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*. Cambridge Molecular Science, Cambridge University Press, 2009.

[271] Shavitt, I., "The method of configuration interaction," in *Methods of Electronic Structure Theory* (Schaefer, H. F., ed.), pp. 189–275, New York: Plenum Press, 1977.

[272] Sherrill, C. D., Takatani, T., and Hohenstein, E. G., "An assessment of theoretical methods for nonbonded interactions: Comparison to complete basis set limit coupled-cluster potential energy curves for the benzene dimer, the methane dimer, benzene-methane, and benzene-H_2S," *J. Phys. Chem. A*, vol. 113, pp. 10146–10159, 2009.

[273] Sherrill, C. D., "Energy component analysis of π interactions," *Acc. Chem. Res.*, vol. 46, no. 4, pp. 1020–1028, 2013.

[274] Sherrill, C. D. and Schaefer, H. F., "The configuration interaction method: Advances in highly correlated approaches," in *Advances in Quantum Chemistry* (Löwdin, P.-O., ed.), vol. 34, pp. 143–269, New York: Academic Press, 1999.

[275] Silvi, B. and Savin, A., "Classification of chemical bonds based on topological analysis of electron localization functions," *Nature*, vol. 371, pp. 683–686, 10 1994.

[276] Sinnokrot, M. O. and Sherrill, C. D., "Unexpected substituent effects in face-to-face π-stacking interactions," *J. Phys. Chem. A*, vol. 107, pp. 8377–8379, 2003.

[277] Sinnokrot, M. O. and Sherrill, C. D., "Substituent effects in π-π interactions: Sandwich and T-shaped configurations," *J. Am. Chem. Soc.*, vol. 126, pp. 7690–7697, 2004.

[278] Skyrme, T., "The effective nuclear potential," *Nuclear Physics*, vol. 9, no. 4, pp. 615 – 634, 1958.

[279] Slater, J. C., "The theory of complex spectra," *Phys. Rev.*, vol. 34, pp. 1293–1322, Nov 1929.

[280] Slater, J. C., "A simplification of the Hartree-Fock method," *Phys. Rev.*, vol. 81, pp. 385–390, Feb 1951.

[281] Smilde, A. K., Bro, R., and Geladi, P., *Multi-way analysis with applications in the chemical sciences*. Chichester, West Sussex, England/Hoboken, NJ: Wiley, 2004.

[282] Šponer, J., Jurečka, P., Marchan, I., Luque, F. J., Orozco, M., and Hobza, P., "Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps," *Chem. Eur. J.*, vol. 12, pp. 2854–2865, 2006.

[283] Stone, A. J., "Distributed multipole analysis, or how to describe a molecular charge-distribution," *Chem. Phys. Lett.*, vol. 83, pp. 233–239, 1981.

[284] Stone, A. J., "Distributed multipole analysis: Stability for large basis sets," *J. Chem. Theory Comput.*, vol. 1, pp. 1128–1132, 2005.

[285] Stone, A. J. and Alderton, M., "Distributed multipole analysis - methods and applications," *Mol. Phys.*, vol. 56, pp. 1047–1064, 1985.

[286] Stone, A. J., *The Theory of Intermolecular Forces*. Oxford: Oxford University Press, 1996.

[287] Strain, M. C., Scuseria, G. E., and Frisch, M. J., "Achieving linear scaling for the electronic quantum Coulomb problem," *Science*, vol. 271, no. 5245, pp. 51–53, 1996.

[288] Stratmann, R., Scuseria, G. E., and Frisch, M. J., "Achieving linear scaling in exchange-correlation density functional quadratures," *Chem. Phys. Lett.*, vol. 257, no. 34, pp. 213 – 223, 1996.

[289] Svozil, D., Hobza, P., and Sponer, J., "Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. can we determine correct order of stability by quantum-chemical calculations?," *J. Phys. Chem. B*, vol. 114, pp. 1191–1203, 2010.

[290] Szabo, A. and Ostlund, N. S., *Modern Quantum Chemistry*. New York: McGraw Hill, 1982.

[291] Szalewicz, K., "Symmetry-adapted perturbation theory of intermolecular forces," *WIREs Comput. Mol. Sci.*, vol. 2, no. 2, pp. 254–272, 2012.

[292] Takatani, T., Hohenstein, E. G., and Sherrill, C. D., "Improvement of the coupled-cluster singles and doubles method via scaling same- and opposite-spin components of the double excitation correlation energy," *J. Chem. Phys.*, vol. 128, p. 124111, 2008.

[293] Takatsuka, A., Ten-no, S., and Hackbusch, W., "Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller–Plesset perturbation theories," *J. Chem. Phys.*, vol. 129, no. 4, p. 044112, 2008.

[294] Talman, J. D., "Some properties of three-dimensional harmonic oscillator wave functions," *Nuclear Physics A*, vol. 141, no. 2, pp. 273 – 288, 1970.

[295] Taube, A. G. and Bartlett, R. J., "Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory," *Collect. Czech. Chem. Commun.*, vol. 70, pp. 837–850, 2005.

[296] Ten-no, S., "Initiation of explicitly correlated Slater-type geminal theory," *Chem. Phys. Lett.*, vol. 398, no. 13, pp. 56–61, 2004.

[297] Treutler, O. and Ahlrichs, R., "Efficient molecular numerical integration schemes," *J. Chem. Phys.*, vol. 102, no. 1, pp. 346–354, 1995.

[298] Turney, J. M., Simmonett, A. C., Parrish, R. M., Hohenstein, E. G., Evangelista, F. A., Fermann, J. T., Mintz, B. J., Burns, L. A., Wilke, J. J., Abrams, M. L., Russ, N. J., Leininger, M. L., Janssen, C. L., Seidl, E. T., Allen, W. D., Schaefer, H. F., King, R. A., Valeev, E. F., Sherrill, C. D., and Crawford, T. D., "Psi4: an open-source ab initio electronic structure program," *WIREs: Comput. Mol. Sci.*, vol. 2, no. 4, pp. 556–565, 2012.

[299] Vahtras, O., Almlöf, J., and Feyereisen, M. W., "Integral approximations for LCAO-SCF calculations," *Chem. Phys. Lett.*, vol. 213, pp. 514–518, 1993.

[300] Valeev, E. F., "Improving on the resolution of the identity in linear R12 ab initio theories," *Chem. Phys. Lett.*, vol. 395, no. 46, pp. 190–195, 2004.

[301] van der Vaart, A., Bursulaya, B. D., Brooks, C. L., and Merz, K. M., "Are many-body effects important in protein folding?," *J. Phys. Chem. B*, vol. 104, no. 40, pp. 9554–9563, 2000.

[302] van der Vaart, A. and Merz, K. M., "Divide and conquer interaction energy decomposition," *J. Phys. Chem. A*, vol. 103, no. 17, pp. 3321–3329, 1999.

[303] van der Vaart, A. and Merz, K. M., "The role of polarization and charge transfer in the solvation of biomolecules," *J. Am. Chem. Soc.*, vol. 121, no. 39, pp. 9182–9190, 1999.

[304] Van Lenthe, J. H., Zwaans, R., Van Dam, H. J. J., and Guest, M. F., "Starting SCF calculations by superposition of atomic densities," *J. Comp. Chem.*, vol. 27, no. 8, pp. 926–932, 2006.

[305] Varganov, S. A. and Martínez, T. J., "Variational geminal-augmented multireference self-consistent field theory: Two-electron systems," *J. Chem. Phys.*, vol. 132, no. 5, p. 054103, 2010.

[306] Vautherin, D. and Brink, D., "Hartree-Fock calculations with Skyrme's interaction. I. Spherical nuclei," *Phys. Rev. C*, vol. 5, no. 3, pp. 626 – 647, 1972.

[307] Verstraelen, T., Ayers, P., Speybroeck, V. V., and Waroquier, M., "The conformational sensitivity of iterative stockholder partitioning schemes," *Chem. Phys. Lett.*, vol. 545, no. 0, pp. 138 – 143, 2012.

[308] Vydrov, O. A. and Voorhis, T. V., "Improving the accuracy of the nonlocal van der Waals density functional with minimal empiricism," *J. Chem. Phys.*, vol. 130, no. 10, p. 104105, 2009.

[309] Vydrov, O. A. and Voorhis, T. V., "Implementation and assessment of a simple nonlocal van der Waals density functional," *J. Chem. Phys.*, vol. 132, no. 16, p. 164113, 2010.

[310] Wall, M. R., Dieckmann, T., Feigon, J., and Neuhauser, D., "Two-dimensional filter-diagonalization: spectral inversion of 2D NMR time-correlation signals including degeneracies," *Chem. Phys. Lett.*, vol. 291, no. 56, pp. 465 – 470, 1998.

[311] Watt, M., Hardebeck, L. K. E., Kirkpatrick, C. C., and Lewis, M., "Face-to-face arene-arene binding energies: Dominated by dispersion but predicted by electrostatic and dispersion/polarizability substituent constants," *J. Am. Chem. Soc.*, vol. 133, pp. 3854–3862, 2011.

[312] Weigend, F., "A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.*, vol. 4, pp. 4285–4291, 2002.

[313] Weigend, F., Köhn, A., and Hättig, C., "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," *J. Chem. Phys.*, vol. 116, pp. 3175–3183, 2002.

[314] Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R., Schütz, M., and others, "MOLPRO, version 2012.1, a package of ab initio programs," 2012. see http://www.molpro.net.

[315] Wheatley, R. J. and Gopal, A. A., "Covalent bond orders and atomic anisotropies from iterated stockholder atoms," *Phys. Chem. Chem. Phys.*, vol. 14, pp. 2087–2091, 2012.

[316] Wheeler, S. E., "Local nature of substituent effects in stacking interactions," *J. Am. Chem. Soc.*, vol. 133, pp. 10262–10274, 2011.

[317] Wheeler, S. E., "Understanding substituent effects in noncovalent interactions involving aromatic rings," *Acc. Chem. Res.*, vol. 46, pp. 1029–1038, 2013.

[318] Wheeler, S. E. and Houk, K. N., "Substituent effects in the benzene dimer are due to direct interactions of the substituents with the unsubstituted benzene," *J. Am. Chem. Soc.*, vol. 130, pp. 10854–10855, 2008.

[319] Wheeler, S. E., McNeil, A. J., Mueller, P., Swager, T. M., and Houk, K. N., "Probing substituent effects in aryl-aryl interactions using stereoselective Diels Alder cycloadditions," *J. Am. Chem. Soc.*, vol. 132, pp. 3304–3311, 2010.

[320] Wheeler, S. E. and Bloom, J. W. G., "Towards a more complete understanding of non-covalent interactions involving aromatic rings," *J. Phys. Chem. A*, vol. 118, no. 32, pp. 6133–6147, 2014.

[321] White, C. A. and HeadGordon, M., "A J matrix engine for density functional theory calculations," *J. Chem. Phys.*, vol. 104, no. 7, pp. 2620–2629, 1996.

[322] White, C. A., Johnson, B. G., Gill, P. M., and Head-Gordon, M., "The continuous fast multipole method," *Chem. Phys. Lett.*, vol. 230, no. 12, pp. 8 – 16, 1994.

[323] White, C. A., Johnson, B. G., Gill, P. M., and Head-Gordon, M., "Linear scaling density functional calculations via the continuous fast multipole method," *Chem. Phys. Lett.*, vol. 253, no. 34, pp. 268 – 278, 1996.

[324] Whitten, J. L., "Coulombic potential-energy integrals and approximations," *J. Chem. Phys.*, vol. 58, pp. 4496–4501, 1973.

[325] Williams, H. L. and Chabalowski, C. F., "Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," *J. Phys. Chem. A*, vol. 106, p. 646, 2001.

[326] Williams, H. L., Szalewicz, K., Moszynski, R., and Jeziorski, B., "Dispersion energy in the coupled pair approximation with noniterative inclusion of single and triple excitations," *J. Chem. Phys.*, vol. 103, no. 11, pp. 4586–4599, 1995.

[327] Woodward, R. B. and Hoffmann, R., "Stereochemistry of electrocyclic reactions," *J. Am. Chem. Soc.*, vol. 87, no. 2, pp. 395–397, 1965.

[328] Wu, Q. and Van Voorhis, T., "Direct optimization method to study constrained systems within density-functional theory," *Phys. Rev. A*, vol. 72, p. 024502, Aug 2005.

[329] Wu, Q. and Van Voorhis, T., "Constrained density functional theory and its application in long-range electron transfer," *J. Chem. Theory Comput.*, vol. 2, no. 3, pp. 765–774, 2006.

[330] Zhang, J., Chen, P., Yuan, B., Ji, W., Cheng, Z., and Qiu, X., "Real-space identification of intermolecular bonding with atomic force microscopy," *Science*, vol. 342, no. 6158, pp. 611–614, 2013.

VITA

Rob Parrish was born in Miami, Florida in 1989, and has lived in Tampa since 1993. He graduated Chamberlain High School in May 2007. In August 2007 he started as a Penny and E. Roe Stamps IV Leadership Scholar at Georgia Tech. He met Prof. David Sherrill in Spring 2009, and was assigned the traditional introductory exercise of writing a Hartree-Fock code, which he still has yet to fully complete. He obtained a Bachelor's of Science in Mechanical Engineering in May 2010, and won the Senior Design Expo and the Richard K. Whitehead Prize. He took the weekend off and has been working full time for Dr. Sherrill ever since. He was awarded the Department of Energy Computational Science Graduate Fellowship in 2011, was invited to the Lindau Meeting of Nobel Laureates in Chemistry in 2012, and was selected to *Forbes'* 30 Under 30 in Science in 2015. In 2012, he performed a practicum in nuclear structure theory under Dr. Nicolas Schunck of Lawrence Livermore National Lab. Following a very productive and enjoyable Ph.D. course under Dr. Sherrill, he will begin a postdoc with Prof. Todd Martínez at Stanford in June 2015.