Local Models for Strongly Correlated Molecules

by

John Anthony Parkhill II

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemistry

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Martin Head-Gordon, Chair
Professor Jhih-Wei Chu
Professor Robert A. Harris

Spring 2010

Local Models for Strongly Correlated Molecules

Abstract

Local Models for Strongly Correlated Molecules

by

John Anthony Parkhill II
Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Martin Head-Gordon, Chair

The most striking and counterintuitive consequences of quantum mechanics play out in the strong correlations of many-particle systems. The physics of these phenomena are exponentially complicated and often non-local. In chemistry, these strong correlations are vital to even qualitative pictures of chemical bonding, but they grow intractably more numerous with the number of particles and remain a significant challenge for models of chemical behavior. Luckily, the strong correlations relevant to most chemical situations can be significantly simplified and compressed using the heuristics which have been developed by chemists up to present day: ideas like bonding electron pairs, and resonance. In this thesis we present a convergent and systematically improvable series of approximations to the many-electron Schrödinger equation which exploit these patterns. Two themes dominate the work: the use of bonding electron pairs as local units for developing efficient models, and an exponential parameterization of the many-electron wave-function.

Contents

List of Figures **iv**

List of Tables **vi**

1 Introduction **1**
1.1 Context . . . 1
1.1.1 The Mean Field Approximation . . . 4
1.1.2 The Exact Solution . . . 5
1.1.3 Perturbation Theory, and the deficits of Hartree-Fock . . . 7
1.1.4 The Complete Active Space Self-Consistent Field . . . 8
1.1.5 The Coupled Cluster Anstaze . . . 10
1.1.6 Orbital-Optimized CC as an approximation of CASSCF . . . 11
1.1.7 The Perfect Pairing Model . . . 11
1.1.8 The Elephant in the room: DFT . . . 13
1.2 Outline of this Work . . . 14
1.2.1 A Sparse Automation of Many-Fermion Algebra . . . 14
1.2.2 The numerical condition of electron correlation theories when only active pairs of electrons are spin-unrestricted . . . 14
1.2.3 The Perfect Quadruples Model . . . 15
1.2.4 The Perfect Hextuples Model . . . 15
1.2.5 Dynamical Correlations: The +SD correction . . . 16
1.2.6 A Density Functional Aside . . . 16

2 Many Fermion Algebra **17**
2.1 Introduction . . . 17
2.2 Many-Electron Algebra . . . 18
2.2.1 Coupled-Cluster Theory . . . 20
2.3 Tensor Representation and Permutational Symmetry . . . 22
2.4 Contraction . . . 23
2.4.1 Rank insensitivity . . . 28
2.5 Conclusions . . . 28

3 Numerical Stability of Unrestricted Correlation Models **31**
3.1 Introduction . . . 31
3.2 Results . . . 32
3.3 Modified Equations Which are Well-Conditioned . . . 35
3.4 Heuristic Understanding of the Problem . . . 37
3.5 Conclusions . . . 39
3.6 Appendix . . . 40
3.6.1 Singularity in constrained cluster wavefunctions . . . 40

4 The Perfect Quadruples Model **44**
4.1 Introduction . . . 44
4.2 The PQ Model . . . 46
4.3 Implementation . . . 49
4.4 Results . . . 51
4.4.1 H_4 . . . 51
4.4.2 Water . . . 51
4.4.3 Ethene . . . 53
4.4.4 Nitrogen Molecule . . . 55
4.5 Discussion and Conclusions . . . 57

5 The Perfect Hextuples Model **60**
5.1 Introduction . . . 60
5.2 The perfect hextuples model . . . 63
5.2.1 Overview of VOO-CC theory . . . 63
5.2.2 Definition of the model . . . 64
5.2.3 Single excitations, orbital-optimization and exactness . . . 66
5.2.4 Implementation . . . 67
5.3 Results . . . 71
5.3.1 Benzene . . . 72
5.3.2 Cope Rearrangment . . . 73
5.3.3 Bergman Reaction . . . 74
5.3.4 F_2 . . . 76
5.4 Discussion and Conclusions . . . 76

6 The +SD Correction **78**
6.1 Introduction . . . 78
6.2 The Models . . . 80
6.3 Results . . . 82
6.3.1 H_2O . . . 82
6.3.2 F_2 . . . 83
6.3.3 BeH_2 . . . 84

6.3.4 H_4 86
6.3.5 H_8 86
6.4 Discussion and Conclusions 87

7 A Novel Range Separation of Exchange **90**
7.1 Introduction 90
7.2 The terf-attenuated LDA 92
7.3 Application to Range Separated Hybrids 94
7.4 Discussion and Conclusions 97

8 Conclusions & Outlook **98**
8.1 Summary 98
8.2 Future Research Directions 99

Bibliography **101**

List of Figures

2.1 An example of the diagram notation. Each box is a list of 4 integers (line notation) characterizing the normal-ordered many-fermion operator. 19
2.2 Number of (symmetrically unique) elements in $\hat{T}_n$ for $2 \leq n \leq 6$ for the calculations performed in Figure 3 and linear least-square fits. The number of amplitudes are constrained to grow in a cubic fashion for any rank with a three-pair constraint. Note that $\mathrm{Dim}(T_2) = \mathrm{Dim}(\langle a_1 b_1 || b_2, a_2 \rangle)$ since a PP-active space is chosen 29
2.3 Wall time consumed in calculating equation 4 for various $2 \leq n \leq 6$ using algorithm 4, and the cubically growing tensors of Figure 2. Horizontal axis is the number of occupied orbitals, and the space has the same number of virtuals. The cost exponent for $n = 2$ and $n = 6$ are estimated by the slope of the linear fits. 29

3.1 Smallest Jacobian eigenvalue for various fragment dissociations, in the 6-31G basis for all cases except Mo_2. 34
3.2 Lyapunov Exponents along the Dissociation Coordinate. 36
3.3 CCD parameters for H_2 in STO-3G basis, horizontal axis θ, vertical axis R (Å). 38
3.4 Orbital labeling for H_2 dissociation. 39
3.5 Orbital labeling for ethene dissociation. 40

4.1 Correlations in the PQ approximation. Horizontal lines represent different orbital pair spaces, occupied below and virtual above. 46
4.2 Scaling of methods in the 6-31G basis. 50
4.3 Potential energy curve for the rectangular dissociation of H_2—H_2 with the cc-pVDZ basis and restricted 2-pair doubles orbitals. (H-H distance 1 $\mathring{A}$) . 52
4.4 Potential energy curve for the symmetric dissociation of water with cc-pVDZ basis with RPP orbitals. 55
4.5 Error decomposed approximately by source for symmetric water dissociation with RPP orbitals. 56

4.6 Electronic energy of ethene (R(C–H) = 1.07 $\mathring{A}$) in (12,12) space with cc-pVDZ basis 57
4.7 Unrestricted potential energy curve for the dissociation of N_2 in the cc-pVDZ basis. 58

5.1 Scaling of amplitude iteration wall-time with system size. Calculations performed on one core of an Apple XServe (Fall 08). Largest system: t-butane (24,24) pictured. Parameters of a linear least-squares fit are inset. 70
5.2 Scaling $\hat{T}$ with system size. Same systems represented by each data point as Figure 1, as included in the supplementary information. Least-Squares fits by quartic polynomials are inset. 70
5.3 $D_{3h} \rightarrow D_{6h}$ deformation of benzene in the 6-31G basis 72
5.4 Three representative points along the path of the Bergman reaction followed in this study. From left to right, Step 0.0 (eneyne), 0.5 (near transition structure, and 1.0 (p-benzyne). 74
5.5 Electronic energy along Bergman reaction coordinate. CASSCF(8,8) in the 6-31g* basis. 75

6.1 Simultaneous dissociation of H_2O in the DZ basis. Orbitals are those of the GVB-RCC model [1]. 83
6.2 Dissociation of F_2 in the DZ basis, Errors relative to CCSDTQPH (frozen-core). PHSD employs the intermediate-pair approximation. 84
6.3 Insertion of Be into H_2. PQSD is built on the (4,4) active space. Orbitals are those of PQ. 85
6.4 Graphical depiction of H_8 model for α = -0.4, 0.4 Bohr 87
6.5 Automerization of H_8. The reference space is (8,8) and basis is DZ. 88

7.1 Various attenuators plotted for comparison. The first two are equivalent to Erf (ω = .3, .4). 93
7.2 Dissociation of He_2^+ 95
7.3 Dissociation of Ar_2^+ 95

List of Tables

3.1 Linear coupling matrix for the PP and IP exchange amplitudes for ethene ($H_2C{=}CH_2$) at a C-C bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space in the 6-31G* basis. 39
3.2 Linear coupling matrix for the PP and IP exchange amplitudes for fluoroethene ($HFC{=}CH_2$) at a C-C bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space. 41
3.3 Linear coupling matrix for the PP and IP exchange amplitudes for nitrogen (N_2) at a N-N bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space in the 6-31G* basis. 42

4.1 CASSCF energies for symmetric water dissociation (E_h), and relative errors of PQ, and non-local active-space models with the RPP orbitals. 53
4.2 Correlation Energies for symmetric water dissociation (E_h) with the RPP orbitals. 54
4.3 Correlation energies for dissociation of ethene(E_h) with restricted PP orbitals. 54
4.4 N_2 Total energies (a.u.) and the error relative to CASSCF. (* includes singles) with unrestricted PP orbitals. 58

5.1 Energetic effects of the hybrid gradient with several OO-CC models of water with the 6-31g** basis. Energies reported are Model - CASSCF(8,8) in E_h. R(O-H) = 1Å, and $\angle(HOH) = 103.1°$. 67
5.2 Typical energetic effects of the intermediate-pair approximation. Frozen PP-orbitals and 6-31g* basis employed in all cases. 69
5.3 Energies along the D_{2h} coordinate of the Cope rearrangement in the (6,6) space and 6-31g* basis. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH. 73
5.4 Study of Bergman Reaction in the (8,8) space 6-31g* basis. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH 75
5.5 F_2 dissociation in the DZ basis and (14,14) space. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH. 76

6.1 The H_4 model system. Zero is a 2-Bohr square of hydrogen atoms, and .5 is the linear configuration. The orbitals are those of PQ. The reference space is (4,4) and basis DZP. 86
6.2 Automerization of H_8. The basis is DZ. MRPH employs the intermediate pair approximation 88

7.1 Mean absolute error (kcal/mol) of G2 set atomization energies and errors of dimer cation asymptotes for various functionals. *(ωB97X), **(ωB97) †(pure Becke 88 exchange [2] and LYP [3] correlation, errors in this row are upper bounds.) 96

Acknowledgments

Many other people have been an vital part of this work. My advisor took a chance on me with little background, was supremely supportive and I'm deeply grateful. The faculty at University of Chicago were also quite inspirational during my undergraduate studies. I enjoyed the assistance of generations of fellow students (Alex Sodt, Ryan Steele, Keith Lawler, Greg Beran, and Daniel Lambrecht). I would also like to thank the professors who generously contributed their time to sit on this committee.

This work was fiscally supported by the Department of Energy through a grant under the program for Scientific Discovery through Advanced Computing (SciDAC). This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Chapter 1

Introduction

1.1 Context

"What I cannot create, I do not understand" - Richard Feynman

Near the turn of the last century it became clear that an adequate description of matter at small scales required [4–9] generalizations of the classical notions of "state" and "observable". The resulting quantum theory [10–18] revolutionized science, and barring relativity provides a comprehensive theory of all chemical phenomena *in principle*. Realistically, the extent of insight that can be afforded from these first-principles depends on how efficiently and faithfully the calculation of a molecular wave-function can be performed by the would-be theoretician. The goal of our work is to make that calculation more accurate and affordable, for the special case of electrons in molecules.

The mathematical structure of quantum mechanics postulates that observables correspond to the action of linear operators on a Hilbert space, a certain sort of inner-product vector space. The state of the physical system is postulated to correspond to a vector in this space and the corresponding dual vector guaranteed by the Riesz representation theorem. These are colloquially called a ket and bra respectively, and together a *wavefunction*. The vectors themselves are not measurable. Concerning ourselves mostly with time-independent problems we will usually imagine without explicitly noting that this wave-function is prepared as a position-space vector. In this case the bra and ket are both complex-valued, square-integrable functions of position, $|\text{ket}\rangle \sim |\Psi(r_1, r_2, \ldots, r_N)\rangle$ and $\langle\text{bra}| \sim \langle\Psi^*(r_1, r_2, \ldots, r_N)|$. The probability of observing a particle in at a given position is then given (up to normalization) by the N-1 particle integral over all but one coordinates $\rho(r) = \int dr_1 ... dr_{N-1} |\Psi|^2$.

To begin, we assume that all nuclei have negligible volume and infinite mass (the Born-Oppenheimer approximation) and their contribution to molecular energies is simple Coulombic repulsion. Even these basic approximations are too coarse for some interesting chemical problems. We solve the Schrödinger equation (SE, Eq. 1.1) for a many electron

wave-function $|\Psi(r_1, r_2..., r_N)\rangle$ given a electronic Hamiltonian parameterized by the nuclei and the absence of any other fields. We choose the atomic units and they will be omitted throughout.

$$\begin{aligned} \hat{H}_{elec}|\Psi_{elec}\rangle &= E_{elec}|\Psi_{elec}\rangle \\ \hat{H}_{elec} &= -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_i \sum_{j>i} \frac{1}{r_{ij}}. \end{aligned} \tag{1.1}$$

The first term describes the kinetic energy of the electron, the second their attraction to the nucleus and the third their repulsion from each other. Exact solutions to this equation have a complex analytic structure, with derivative singularities anywhere the positions of two electrons are the same and are not generally available.

To make progress we must guess state-vectors which we can easily express. To this end we introduce an orthonormal basis of single electron states $\{\chi_i\}$ which we assume we can grow until until complete. Ignoring relativity, we postulate that each of these states is the product $\{\chi_i(r, \sigma)\}$ of a spatial function and a spin function which takes on one of two possible orthonormal vectors with eigenvalues $(\pm\frac{1}{2})$. The spatial functions can be expanded with any sort of square-integrable functions we choose. The simplest wave-function ansatz is a simple Hartree [19–22] product of these 1-electron functions:

$$|\Psi(r_1, r_2..., r_N)\rangle \approx \prod_i \chi_i(r_i) \tag{1.2}$$

However Fermions have an eigenvalue of -1 under parity (permutation of coordinates) and this trial function does not have those statistics. The simplest possible modification which ensures anti-symmetry is a linear combination of these Hartree products, summed over the permutations $\{\sigma_i\}$ on electrons with a sign factor for each index swap in that permutation.

$$|\Psi(r_1, r_2..., r_N)\rangle \approx \sum_{\sigma_k} \mathbf{Sign}(\sigma_k) \prod_i \chi_i(r_{\sigma_k(i)}) \tag{1.3}$$

This model is called a Slater [23–25] determinant, and it has the important property that at any given position an electron excludes others of the same spin.

There is an alternative notation for an anti-symmetric product of single-particle states in terms of *creation* $\{a_i^\dagger\}$ and *annihilation* $\{a_i\}$ field operators, called *"Second Quantization"* [26, 27]. In this notation, which we will rely upon heavily for the remainder of this work, the physical state coordinates $\{r_i\}$ are abstracted away and replaced with summed indices over single-particle basis states. In this thesis these single particle states are imagined to be vectors in the position representation; for this reason we will omit resolutions over position space. The representation chosen for the single-particle states (position eigenstates, momentum eigenstates, etc.) usually doesn't alter the physicality or notation of a state in second quantization.

The creation and annihilation operators are not a part of our intuitive reality (ie: they are not Hermitian, and don't possess real spectra), but so long as they span a space which satisfies the hypotheses of our Hilbert space and obey the canonical commutation relations (we construct them to do so), they provide an allowable representation of our quantum mechanical system. We call the space they generate by action on the vacuum a Fock space. The Fermion statistics are book-kept with the *anti-commutation relations*, and normalization conditions which define the algebra on these operators:

$$\text{Commutation: } [a_i, a_j] = [a_i^\dagger, a_j^\dagger] = \delta_{i,j} \tag{1.4}$$

$$\text{Anti-Commutation: } [a_i, a_j]_+ = [a_i^\dagger, a_j^\dagger]_+ = 0 \; ; \; [a_i, a_j^\dagger]_+ = \delta_{i,j} \tag{1.5}$$

$$\text{Normalization: } a_i a_i = a_i^\dagger a_i^\dagger = 0; \; a_i a_i^\dagger = 1 \tag{1.6}$$

The eigenstates of the annihilation operators are those states which are unchanged by detection, so-called coherent states but this thesis is not concerned with their properties. Second-quantization makes the expression of many-particle states significantly easier because we don't need to constantly rely on the complex (3N variable) analytic form of the wave-function to express the operator algebra which is the essence of our quantum physical system.

To employ this notation we assume the existence of some vacuum $|0\rangle$, a state of zero particles. The Slater determinant is written as a simple product of creation operators on the left of this state; each populating a previously vacant single particle state. It is useful to introduce some conventions on the indices we will be using. Any sort of orbital may be denoted: $\{p, q, r...\}$, occupied orbitals are denoted: $\{i, j, k...\}$, and virtuals $\{a, b, c...\}$. The atomic orbital basis will be denoted: $\{\mu, \nu, \lambda...\}$. Any observable N-particle operator can be written as a tensor of dimension 2N in second quantization. Commonly we will write an operator in that fashion with a X_{bra}^{ket} ie: $\langle a_s a_r | X_{pq}^{rs} a_s^\dagger a_r^\dagger a_p a_q | a_p^\dagger a_q^\dagger \rangle$ denoted: X_{pq}^{rs}. The Slater determinant can be denoted:

$$\sum_{\sigma_k} \mathbf{Sign}(\sigma_k) \prod_i \chi_i(r_{\sigma_k(i)}) = \prod_i a_i^\dagger |0\rangle \tag{1.7}$$

This approximate form for the many-electron wave-function provides a useful starting point for quantitative calculations and so we should describe how it may be calculated in detail. There is a special class of operator strings, called *normal-ordered*, which vanish when their expectation value is taken with the vacuum. In a normal-ordered operator string all creation operators lie to the left of all annihilation operators. Although the utility of this class of operator strings isn't yet obvious it will be explored at length in Chapter 1.

1.1.1 The Mean Field Approximation

Starting with a model of 1 determinant for molecular electronic structure, we hope to approximately solve the SE. In most calculations the single particle basis is written as a linear combination of atomic orbitals ($\phi_\mu(r)$) which are themselves expanded as sums of Gaussian functions.

$$\chi_i(r) = \sum_\mu C_i^\mu \phi_\mu(r) \tag{1.8}$$

The atomic orbitals are non-orthogonal with overlap matrix $S_\mu^\nu = \langle \phi_\mu | \phi_\nu \rangle$. The first two terms of the Hamiltonian, Eq. 1.1, do not present much difficulty so long as the right basis functions are chosen they are simply integrated directly. We denote the matrix element for the "core" Hamiltonian (kinetic and electron-nuclear terms) $\hat{h}_i^j = \langle \chi_i | - \sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} | \chi_j \rangle$. The mean-field approximation is equivalent to supposing that each electron experiences an effective 1-particle (2-index) Hamiltonian parameterized by the others. The inter-electronic repulsion term has second quantized representation: $V_{pq}^{rs} = \langle sr | \frac{1}{r} | pq \rangle$. There are two two-index contributions:

$$\hat{J}_i^j = \langle ij|1/r_{ij}|ij\rangle = \int dr_1 \chi_i(r_1)[\int dr_2 |\chi_j(r_2)|^2 1/r_{ij}]\chi_i(r_1) \tag{1.9}$$

$$\hat{K}_i^j = \langle ij|1/r_{ij}|ji\rangle = \int dr_1 \chi_i(r_1)[\int dr_2 \chi_j(r_2) 1/r_{ij} \chi_i(r_2)]\chi_j(r_1). \tag{1.10}$$

So we construct the best possible effective 1-particle Hamiltonian, the Fock operator:

$$\hat{f}_i^j = \hat{h}_i^j + \hat{J}_i^j - \hat{K}_i^j \tag{1.11}$$

It should be immediately obvious that the mean-field wave-function is not an exact solution of the many-electron problem, because it doesn't even depend on the complete many-electron Hamiltonian. However in the next section we will cover the exact solution. The determinant which is the eigenfunction of this mean-field hamiltonian can be determined [28–30] by solving the generalized eigenvalue problem:

$$fC = SC\epsilon \tag{1.12}$$

ϵ is a vector containing the eigenvalues of f corresponding to each orbital. In the simplest Roothan-Hall [31, 32] approach the generalized eigenvalue problem is simply solved iteratively. Each new Fock operator generates a new determinant which generates a new Fock operator until convergence. At each iteration the N lowest eigenvalue orbitals are chosen to be occupied, and the remainder are called virtual. The Hartree-Fock energy is

the expectation value of the true Hamiltonian with the converged determinant:

$$E_{HF} = \sum_i \epsilon_i + \frac{1}{2}\sum_{i,j} J_i^j - K_i^J \tag{1.13}$$

Orbital Rotations

One can arrive at any determinant of orthogonal spin-orbitals in a given basis from any other by a series of unitary transformations of those basis vectors. Any unitary transformation can be parameterized as the exponential of an anti-Hermitian operator. In this way we can express the converged HF orbitals in terms of an orthogonal set of guess orbitals as:

$$C = C_0 U = C_0 e^{\theta} \tag{1.14}$$

$$\frac{\delta E_{HF}}{\delta U_p^i} = 2 f_p^i \tag{1.15}$$

$$\theta = -\theta^{\dagger} \tag{1.16}$$

the (N_{MO}, N_{MO}) matrix θ can be geometrically termed a rotation operator which mixes two single-particle states. Using this orbital rotation picture we can develop a steepest descent algorithm to solve the HF equations, by following the gradient of the energy with respect to θ. At each iteration the chain rule is used to construct $\frac{\delta E}{\delta \theta}$ from $\frac{\delta E}{\delta U}$ and $\frac{\delta U}{\delta \theta}$. The Hartree-Fock energy is invariant to rotations within the occupied and virtual spaces, but not between them.

1.1.2 The Exact Solution

The mean-field solution provides an orthogonal 1-particle basis partitioned into occupied and virtual spaces, and an approximate wavefunction which lies entirely in the occupied space. Within the given 1-particle basis the space of N-electron wave-functions is spanned by a linear combination of all possible N-orbital determinants. These can be imagined as resulting from the action of normal-ordered "excitation operators" which replace a certain number of indices, occupied in the HF wave-function with previously unoccupied 1-electron states.

$$|\Psi\rangle_{\text{Full-CI}} = \sum_{p_1,p_2,\ldots p_N} C_{p_1,p_2,\ldots p_N} \prod_i a_{p_i}^{\dagger} = \sum_{n=1..N} \hat{C}_n \cdot |\Psi_{HF}\rangle \tag{1.17}$$

$$\text{Where: } \hat{C}_n = \sum_{i_k,b_k} C_{i_1,i_2\ldots i_n}^{b_1,b_2\ldots b_n} a_{b_n}^{\dagger} \ldots a_{b_1}^{\dagger} a_{i_1} \ldots a_{i_n} \tag{1.18}$$

Given this ansatz of all possible configurations, the Full-Configuration Interaction expansion (FCI) [33–40], the ground state wave-function can be determined via the Ritz variational principle. In the simplest possible version the Hamiltonian is explicitly constructed and diagonalized in the basis of $\hat{C}_n$ providing a complete manifold of bound states.

$$\hat{H} = \begin{pmatrix} E_{HF} & \langle 0|\hat{H}|\hat{C}_1 0\rangle & \cdots & \langle 0|\hat{H}|\hat{C}_n 0\rangle \\ 0 & C_1^* C_1 \langle \mu_1|\hat{H}|\mu_1\rangle & \cdots & C_1^* C_n \langle \mu_1|\hat{H}|\mu_n 0\rangle \\ \vdots & & \ddots & \vdots \\ 0 & C_n^* C_1 \langle \mu_n|\hat{H}|\mu_1\rangle & \cdots & C_n^* C_n \langle \mu_n|\hat{H}|\mu_n 0\rangle \end{pmatrix} \tag{1.19}$$

$$\text{Where: } \mu_n = \sum_{i_k, b_k} a^\dagger_{b_n} ... a^\dagger_{b_1} a_{i_1} ... a_{i_n} \tag{1.20}$$

If only a few lowest-lying states are desired the matrix can be implicitly diagonalized using a Lanzcos-type [41, 42] algorithm and only the lowest vector must be stored. However each vector must be represented in the basis of all possible n-electron determinants, a space which grows as $N_{MO}!$. To be concrete, imagine a water molecule with 8 valence electrons. Minimally we may introduce two single-particle basis functions for each electron, although chemical accuracy will not be achieved until much more than 30 are added. Even in this insufficient basis there are $\frac{16!}{8!(16!-8!)} = 12870$, 8-electron determinants. If fifteen basis functions are introduced there are roughly 782 gigabytes worth of determinants in the full expansion. If thirty basis functions are introduced for each electron there are $\approx 2.4 * 10^{14}$ expansion coefficients and so on. It is unlikely that more than 18 electrons can ever be treated in this fashion within the author's lifetime even with a basis which is very small.

The difference between the HF energy and the FCI energy defines the *correlation energy*. The remaining problem of this work is to find truncations of the FCI expansion which can be expressed in a number of variables which can be practically manipulated on a modern computer. A multitude of possible approximations have been introduced, many with significant success and application to chemistry. The variational theorem ensures that FCI energy provides a lower bound for any approximate model which is determined by projection against the *whole* SE, and so often the cheaper approximate models will *underestimate* the correlation energy. We will review some established approaches before proceeding into our own work.

Size-Consistency

The most obvious first step to a tractable approximation of the electronic structure problem is limiting the rank of $\hat{C}_n$ and solving the SE over that incomplete space. The highest-rank excitation operators are the most costly to calculate, $O(m^{2n+2})$, and often less significant than the less numerous, lower-rank contributions. Unfortunately the accuracy of

such a truncation varies when two non-interacting systems are treated as a unit rather than apart. In chemistry, where particles are redistributed during the course of reactions between reactants and products (which we can imagine to be independent), approximations which are not *size-consistent* lack a significant cancellation of errors and are simply less-useful.

Intermediate Normalization

Often the Hartree-Fock wave-function has significant weight in the exact expansion of the wave-function, and there is a convenient choice of normalization, *Intermediate Normalization*, reflecting this fact which sets the weight of determinantal reference to 1. Formally $\langle\Psi|0\rangle = 1$ where Ψ is some correlated improvement, and the energy of concern (now the correlation energy) is given: $\langle\Psi|H|0\rangle = E_c$. Throughout the remainder of this work we largely imply this choice of normalization, and if not otherwise noted assume every energy is a correlation energy.

1.1.3 Perturbation Theory, and the deficits of Hartree-Fock

The true many electron state includes information about electrons excluding each other locally, and spin-coupling information which cannot be represented in a single determinant or effective 1-particle field [43]. Conversely the exact answer has all of these effects, but defies expression and so we seek something simpler. The simplest way to correct for correlation is to perturb the HF wave-function with the missing components of $\hat{H}$ and calculate the leading order perturbative correction [44, 45]. $H|\Psi_{HF}\rangle$ is spanned by the space of double excitations above the reference, because $\hat{H}$ is a two-body operator. As in every perturbation theory we introduce a partitioning of the Hamiltonian: $\hat{H} = \hat{H}_0 + \hat{V}$, in this case (the Møller-Plesset partitioning) we choose $\hat{H}_0 = \hat{f}$. In kind we partition the space of many electron determinants into space (0) which is spanned by the eigenfunctions of $\hat{H}_0$ such that $E_0 = \langle 0|H_0|0\rangle$ and the remainder. Given these choices the correlation energy and corrected wave-function are determined as usual in Rayleigh-Schrödinger Perturbation Theory (RSPT) by collecting terms of each given order from expansion over the SE.

$$(\hat{H}_0 + \hat{V}) \sum_k |\Psi^k\rangle = (\sum_k E^k) \sum_k |\Psi^k\rangle \tag{1.21}$$

We introduce the projectors onto each space: $|0\rangle\langle 0|$ and $\hat{P} = 1 - |0\rangle\langle 0|$. We determine the amplitudes of the first order interacting space by projection onto the perturbation:

$$|\Psi\rangle_{MP1} = (1 + \hat{T}_2)|0\rangle \tag{1.22}$$
$$= -\hat{P}(\hat{H}_0 - E_0)\hat{P}\hat{V}|0> \tag{1.23}$$
$$\rightarrow \hat{T}_2 = \frac{\langle 0|\hat{V}a_b^\dagger a_a^\dagger a_i a_j|0\rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{1.24}$$

Then calculate the energy of the corrected wave-function with the SE:

$$E_{c,\mathrm{MP2}} = \langle 0|\hat{V}|\Psi_{\mathrm{MP1}}\rangle = \langle 0|\hat{V}\hat{T}_2|0\rangle \tag{1.25}$$

Strong Correlations

The MP2 wave-function significantly improves on HF, and gives nearly quantitative molecular interaction and atomization energies. However there is serious cause for concern in the details. In cases where the MP2 correction is large, one might be tempted to continue the series to higher order to achieve greater accuracy, but such an expansion often does not always converge. In fact, anytime a closed-shell molecule is dissociated into radical fragments the perturbation theory will fail. We call these PT defeating correlations "strong". The most striking consequence of quantum many-body theory is that particles can become non-locally "entangled" with one another. These are precisely the situations where no single determinant dominates the exact wave-function and a perturbation analysis will fail. Often nearly degenerate single-particle spectra are indicative of such a situation. Because this sort of correlation isn't the same as the local-exclusion, "dynamical" effect it is also sometimes called a "static" correlation problem.

A determinantal expansion of the exact wave-function in a strongly correlated situation often has a number of significant determinants which compete for dominance. The number of such determinants grows exponentially with the number of strongly correlated particles, and forms the fascinating crux of the strong correlation problem. These are often related to each other by greater-than-double replacements, and accompanied by the very large number of small-weight configurations which model the dynamical exclusion of electrons from each other. There are however models capable of capturing these strong correlation effects, to which we turn our attention. We note in passing that our problem has a converse: strongly correlated quantum systems can describe exponential complexity with linear numbers of particles. This is the force behind the growing field of quantum computation.

1.1.4 The Complete Active Space Self-Consistent Field

FCI isn't affordable for very many electrons, but usually the strong correlations of a molecular system occur between just a few single-particle states and the remaining are well

described by the Hartree-Fock model. To exploit this fact we can solve CI within a space of only a few orbitals, and make those orbitals well-defined by introducing a variational condition which determines them. The model which results from solving a CI in an *"active space"* of only a few electrons and orbitals and variationally minimizing the correlation energy with respect to perturbations of these orbitals is called multiple configuration SCF (MCSCF) [46–48]. If the CI is chosen to be complete the resulting model is called a Complete Active Space Self-Consistent Field (CASSCF) model.

$$|\Psi\rangle_{CASSCF} = \sum_{\hat{C}\in\text{Active}} \hat{C}|0\rangle \tag{1.26}$$

$$|0\rangle = \hat{U}|\Psi\rangle_{HF} \text{ where } \hat{U} = e^{\theta} \tag{1.27}$$

$$\text{s.t. } \frac{\delta E}{\delta \hat{C}} = 0, \text{ and } \frac{\delta E}{\delta \theta} = 0 \tag{1.28}$$

$$\tag{1.29}$$

Because the CI wave-function obeys a variational principle, the gradient is given trivially by the Hellman-Feynman Theorem.

$$\frac{\delta E_{MCSCF}}{\delta U} = \langle\Psi|\frac{\delta \hat{H}}{\delta U}|\Psi\rangle_{MCSCF} \tag{1.30}$$

Starting with the HF MO's a CASSCF wavefunction is determined by solving the active-space CI, forming the orbital gradient, taking a step in θ and transforming the Hamiltonian into the new basis until convergence. In addition to the occupied-virtual rotations which must be optimized in HF, rotations must be performed between active and inactive occupieds and virtuals. θ is a non-linear parameter meaning that there may be more than one set of orbitals satisfying the CASSCF condition which could be chosen to model a given electronic state. Still it is common to proceed as if there were a unique solution given a number of orbitals chosen to be active, and a number of electrons chosen to fill those orbitals. Admitting an active occupied and virtual orbital for each valence electron pair of a molecule is often an excellent choice, and called the perfect-pairing valence active space. When it is computable for the complete valence space the CASSCF wavefunction provides a smooth and qualitatively correct model of electronic structure, but too often the small active space which is affordable is not accurate.

Aside from the exponential scaling which permanently limits its range of applicability, CASSCF also systematically underestimates the correlation energy because it largely misses dynamical correlation. In most cases it has such significant overlap with the exact wave-function that a perturbative correction for these missing correlations is largely successful [47]. This divide-and-conquer approach to the correlation problem is a common element of most multi-reference correlation methods.

1.1.5 The Coupled Cluster Anstaze

Truncating the CI excitation operator at any level of excitation results in a model whose errors grow with system size, and for chemistry where particle number changes in the course of a reaction this is extremely undesirable. In statistical mechanics the exponential form of the partition function is motivated by the extensivity of the energy. Likewise if we want an extensive wave-function we should choose an exponential shape [49–51]. Suppose an ansatze of exponentiated cluster operators up to rank n :

$$\Psi_{CC}\rangle = e^{\hat{T}}|0\rangle \tag{1.31}$$

$$\hat{T} = \sum_{i=1,n} T_n \text{ and } \hat{T}_n = \sum_{i,b} T^{b_1,b_2...b_n}_{i_1,i_2...i_n} a^\dagger_{b_n}...a^\dagger_{b_1} a_{i_1}...a_{i_n} \tag{1.32}$$

This is the trial wave-function of a coupled cluster model. These models are amongst the most successful and accurate in quantum chemistry [52] and strongly influence the models developed in this thesis, although they were first developed to model the interactions between nucleons [53]. To solve for the parameters of the expression above, one might be tempted to introduce the dual of the ket, and solve using the Ritz variational method [54]. Unfortunately even if $\hat{T}$ is much smaller than the span of the Full-CI vector to do so requires an exponential amount of computational effort. An alternative set of non-linear homogeneous equations for each parameter can be derived if the ansatze above is inserted into the SE and projected on the left with each level of excitation (μ_i) present in the cluster operator:

$$\hat{H}\Psi_{CC}\rangle = Ee^{\hat{T}}|0\rangle \rightarrow \langle\mu_i|e^{-\hat{T}}\hat{H}e^{\hat{T}}|0\rangle = 0 \tag{1.33}$$

With this prescription the computational cost of a coupled-cluster model is the same as the similarly truncated CI, $O(m^{2n+2})$ where m is the dimension of the basis and n the rank of the excitation operator), but the resulting model is extensive and significantly more accurate. Unfortunately the price paid is the loss of the variational lower-bound. Most molecules at their equilibrium geometry can be quantitatively described if the cluster operator is truncated at triples. The coupled-cluster singles-doubles model (CCSD) is a ubiquitous and highly successful variant. This expansion of the wave-function makes no differentiation between weak-or-strong correlations, but fails if the truncation rank of the cluster operator is smaller than the number of strongly correlated electrons in the molecule. At the time this thesis is written computer resources are such that CCSD(T) [55] is the method-of-choice if a system can be represented with roughly fewer than 2000 basis functions, and these methods afford chemical accuracy if there is no strong correlation problem. Local variants with near-linear scaling have been developed [56, 57], although we cannot yet reach the regime where that linear scaling is cost-effective.

The cluster operator is isometric and so by similarity transforming the Hamiltonian

$e^{-\hat{T}}\hat{H}e^{\hat{T}}$ we do not alter its spectrum, although the spectral expansion of each state is shifted downwards into the lower-rank blocks of the transformed Hamiltonian matrix. The transformed Hamiltonian can be diagonalized, just like $\hat{H}$ itself providing correlated excited states which are usefully accurate for states dominated by a single excitation, even at the level of doubles. This is called Equation of Motion (EOM)-CC [58, 59]. However this transformation does create an effective Hamiltonian which is non-Hermitian, and that has some non-trivial computational and physical implications which are beyond our scope.

1.1.6 Orbital-Optimized CC as an approximation of CASSCF

CC is a size-extensive truncation of CI, and so one might wonder whether CC with an orbital optimization condition would also make a useful and size-extensive truncation of CASSCF. This direction was pursued by the Head-Gordon group roughly a decade ago. The resulting Valence-Optimized Doubles model is defined by a pseudo-variational Lagrangian and orbital optimization condition:

$$\tilde{E}_c = \langle 0|(1+\Lambda)e^{-T}|\hat{H}|e^{T}0\rangle \tag{1.34}$$

$$\text{Where: } \hat{\Lambda}_n = \sum_{i_1\ldots i_n, b_1\ldots b_n} \Lambda^{b_1\ldots b_n}_{i_1\ldots i_n} a^\dagger_{i_n}\ldots a^\dagger_{i_1} a_{b_1}\ldots a_{b_n} \tag{1.35}$$

$$\frac{d\tilde{E}}{d\hat{T}} = 0 \rightarrow \text{(Multiplier Condition)} \tag{1.36}$$

$$\frac{d\tilde{E}}{d\hat{\Lambda}} = 0 \rightarrow \text{(Amplitude Condition)} \tag{1.37}$$

$$\frac{d\tilde{E}}{d\theta^q_p} = \langle 0|(1+\Lambda)e^{-T}|[\hat{H},(a^q_p - a^p_q)]|e^{T}0\rangle = 0 \tag{1.38}$$

$$\tag{1.39}$$

When the strong valence correlations have the character of two doublets coupled to a singlet the VOD wave-function is an accurate approximation to CASSCF for non-interacting systems. The computational costs of VOD grow with the sixth order of molecular size, which is significantly smaller than $N_{MO}!$, but if more than 1 electron pair is strongly correlated (the dissociation of multiple bonds. metals, etc.), VOD is an unsatisfactory model of the electronic structure. The 6^{th} order effort required for these non-local models precludes their application to large systems and they still lack dynamical correlation.

1.1.7 The Perfect Pairing Model

Bonding electron pairs are a useful idea [60] which have found ubiquitous applications in chemistry since the invention of a Lewis dot-structure. This indicates that bonding pairs may also be a useful way to concisely express a many electron wave-function. The minimal

model which is exact for non-interacting electron pairs is called the Generalized-Valence-Bond Perfect Pairing [61, 62] method. The ansatze takes the form:

$$\Psi_{pp}\rangle = \sum_{i} \hat{T}_{i\bar{i}}^{i^*\bar{i}^*} |0\rangle \tag{1.40}$$

where $\{i, \bar{i}, i^*, \bar{i}^*\}$ are an occupied alpha, occupied beta, virtual alpha and virtual beta orbital respectively. These "pair" quartets of spin-orbitals are uniquely associated with one another by rotating the orbitals self-consistently amongst themselves. This requires additional rotations beyond CASSCF within each space, because there is no longer invariance between occupied rotations. One can solve the anstaze introduced above as a truncated coupled-cluster doubles model [63, 64] using the same equations presented above for VOD, except only allowing each amplitude and multiplier to possess indices of one electron pair. This approximate wave-function can be afforded for hundreds of electrons and usually captures between 30 and 60 percent of the valence correlation energy. Unfortunately PP exaggerates the locality of electron correlations, resulting in symmetry-breaking artifacts for systems with multiple resonance structures (benzene, allyl radical etc.). It has been the purpose of this thesis to find generalizations of this idea which provide improved accuracy, approaching the accuracy of CASSCF, and finally approximate the total electronic energy.

Models Beyond PP

To address the symmetry breaking problem our group has pursued several directions before this thesis. Attempts to reduce symmetry breaking by allowing 2-pair excitations (in the doubles space) resulted in the Imperfect-Pairing (IP) [65–67] model, which captures significantly more correlation energy but still exhibits most symmetry artifacts. The IP model's correlation energy also unphysically diverges for dissociation problems. This was repaired by introducing a truncation of the amplitude equations inspired by GVB (GVB-RCC) [68] which removes the offending terms, but is still physically lacking for multi-bond dissociation processes. Later perturbative three-pair amplitudes were combined in a hybrid Lagrangian from the IP model. The resulting three-center imperfect-pairing model (TIP) [69, 70] improved correlation energy recovery into the 90% range for most systems. A somewhat different approach attempted to add the missing projective information from greater-than doubles by truncating the extended-coupled cluster model. The resulting quadratic coupled-cluster doubles [71, 72] models improved on the performance of CCSD for dissociation problems, at the price of increased cost.

Dynamical Corrections

Parallel to work developing an affordable reference for strong correlations our group developed corrections which afford accurate total energies by adding back in the large numbers of small correlation contributions which arise from orbitals outside of the valence

virtuals. These corrections [73–77] were developed by introducing a matrix-based Löwdin [45] style partitioning of the coupled cluster Hamiltonian, in a spirit much like MP2. The numerical performance of these models relies entirely on the strength of the reference. In cases where that reference is solid, but expensive (like CCSD), the resulting models are remarkably accurate, but expensive (7^{th} order).

1.1.8 The Elephant in the room: DFT

This thesis also describes some work on density functional models of chemistry. These approaches are significantly cheaper than any other correlation method and can afford near chemical accuracy, given a simple electronic structure and empirical fitting procedures. However they achieve this efficiency by giving up the systematic improvability which is the feature of the *ab-inito* models described above. In fact they largely ignore the non-local strong correlation physics which are the thrust of this thesis.

The founding theorems [78] of this formalism state that the ground state density somehow uniquely determines the ground state energy although the functional dependence of the energy on the density is not constructively established. For a one-electron system it's relatively elementary to derive the exact functional (since there is a simple map between density and wavefunction), but for many electron systems the exact functional is not available, unless the system is otherwise exactly soluble. To continue we postulate that for each interacting many-electron system there exists a continuous map (an adiabatic connection) to a fictitious non-interacting system under a *local* external potential which possesses the same density as the true physical density. In the case of Kohn-Sham [78–81] DFT, the noninteracting system we imagine to be the HF wave-function, resulting in an effective HF-like equation for the fictitious orbitals:

$$\left(\hat{h} + \hat{T}_{\text{diff}} + \hat{J} + \hat{V}_{xc}\{\rho(r)\}\right)\phi_i(r) = \epsilon_i\phi_i(r) \tag{1.41}$$

$$\hat{V}_{xc}\{\rho(r)\} \approx \int \delta r \rho(r) F(\rho(r), \nabla\rho(r), ...) \tag{1.42}$$

Here $\hat{T}_{\text{diff}}$ is the difference between the kinetic energies of interacting and non-interacting systems (assumed 0 in our case) and $\hat{V}_{xc}$ is a *functional* of the electron density absorbing all exchange and correlation effects. The second line above reflects the functional form which is applied in the vast majority of all DFT calculations. Note that this implies an approximation that the exchange energy is semi-local (which it is not). The largest errors of most functionals come about from the missing cancellation between $\hat{J}$, and it's anti-symmetric partner, $\hat{K}$ which causes electrons to repel themselves.

To construct models for $\hat{V}_{xc}$, early work focused on analytically soluble model systems, like the uniform electron gas (UEG). As the functional is made more-and-more general these approaches have been superseded in chemistry by empirical fits of functional forms

against large sets of standardized chemical data [82, 83], sometimes preserving the uniform electron gas limit. This empiricism affords accuracy, especially for thermochemistry which is dominated by local, dynamical correlations.

So there is a possible connection between the electron density and the true energy of the ground state. Of course we can say something about how this functional must behave for toy systems. It depends on the distributions of electrons *everywhere*, and cannot be written as a semi-local gradient expansion of $\rho(r)$. Even supposing that it was some integral over two electron positions: $E_{xc} \approx \int \delta r_1 \delta r_2 F(\rho(r_1), \rho(r_2))$ (and this is not so!) the cost of such a construction would grow with the size of a molecule like MP2. No functional in common use is exact for even a single-electron atom. If we had an expression for the exact functional, such an expression would be intractable for large systems, and likely no more appropriate a starting point than FCI. Nonetheless DFT is an enormously successful model of dynamical correlations in chemistry. No thesis towards the correlation problem is complete without some understanding of how DFT achieves this massive feat.

1.2 Outline of this Work

1.2.1 A Sparse Automation of Many-Fermion Algebra

Algorithms useful in the construction of electron correlation models are collected alongside new developments for cases of high rank and sparsity. In the first part of this paper a Brandow diagram manipulation program is presented. The complementary second section describes a general-rank sparse contraction algorithm which exploits the permutational symmetries of many-fermion quantities. Several recently published local correlation models (perfect quadruples and perfect hextuples) were built using these codes. This paper should facilitate reproduction and extension of high-rank electron correlation models that combine truncation by level of substitution with truncation by locality, such as the number of entangled electron pairs. This Chapter has been published as a paper in Molecular Physics [84].

1.2.2 The numerical condition of electron correlation theories when only active pairs of electrons are spin-unrestricted

The use of spin-unrestriction with high-quality correlation theory, like coupled-cluster (CC) methods, is a common practice necessary to obtain high quality potential energy surfaces. While this typically is a useful approach, we find that in the unrestricted limit of ROHF fragments the CC equations are singular if only the strongly correlated electrons are considered. Unstable amplitudes which don't represent the physics of the problem are easily found and could be unwittingly accepted without inspection. We use CCD stability analysis and the condition number of the coupled-cluster doubles Jacobian matrix to examine the

problem, and present results for several molecular systems with a variety of unrestricted cluster models. Finally a regularization of the CC equations is proposed which allows us to apply CC, and Lagrangian gradient formulas even with completely unrestricted orbitals. This Chapter has been published as a paper in The Journal of Chemical Physics [85].

1.2.3 The Perfect Quadruples Model

A local approximation to the Schrödinger equation in a valence active-space is suggested, based on coupled cluster (CC) theory. Working in a pairing active space with one virtual orbital per occupied orbital, this Perfect Quadruples (PQ) model is defined such that electrons are strongly correlated up to "four-at-a-time" in up to 2 different (occupied-virtual) electron pairs. This is a truncation of CC theory with up to quadruple substitutions (CCSDTQ) in the active space, such that the retained amplitudes in PQ are proportional to the fourth root of the number of CCSDTQ amplitudes. Despite the apparently drastic nature of the PQ truncation, in the cases examined this model is a very accurate approximation to Complete Active Space Self-Consistent Field (CASSCF). Examples include deformations of square H_4, dissociation of two single bonds (water), a double bond (ethene), and a triple bond (nitrogen). The computational scaling of the model (4^{th} order with molecule size) is less than integral transformation, so relatively large systems can be addressed with improved accuracy relative to earlier methods such as perfect and imperfect pairing which are truncations of CCSD in an active space. This Chapter has been published as a paper in The Journal of Chemical Physics [86].

1.2.4 The Perfect Hextuples Model

We present the next stage in a hierarchy of local approximations to complete active space self-consistent field model (CASSCF) in an active space of one active orbital per active electron, based on the valence orbital-optimized coupled-cluster (VOO-CC) formalism. Following the perfect pairing (PP) model, which is exact for a single electron pair and extensive, and the perfect quadruples (PQ) model, which is exact for two pairs, we introduce the perfect hextuples (PH) model, which is exact for three pairs. PH is an approximation to the VOO-CC method truncated at hextuples containing all correlations between three electron pairs. While VOO-CCDTQ56 requires computational effort scaling with the 14^{th} power of molecular size, PH requires only 6^{th} power effort. Our implementation also introduces some techniques which reduce the scaling to fifth order, and has been applied to active spaces roughly twice the size of the CASSCF limit. Because PH explicitly correlates up to six electrons at a time, it can faithfully model the static correlations of molecules with up to triple bonds in a size-consistent fashion and for organic reactions usually reproduces CASSCF with chemical accuracy. The convergence of the PP, PQ, PH hierarchy is demonstrated on a variety of examples including symmetry breaking in benzene, the Cope

rearrangement, the Bergman reaction and the dissociation of fluorine. This Chapter has been submitted as a paper in The Journal of Chemical Physics.

1.2.5 Dynamical Correlations: The +SD correction

The multi-reference cluster approach based on single-reference formalism(SRMRCC) is combined with paired, active space treatments of static correlation to produce a satisfyingly simple cluster truncation amenable to strongly correlated problems. An implementation of the method is compared to benchmark results for F_2 and H_2O dissociation problems, the H_4 and H_8 model systems, and the insertion of beryllium into hydrogen. The model demonstrates the simplicity, accuracy and compactness offered by orbital-optimized coupled cluster models (OO-CC), and the possibility of a local method for strong correlation. This Chapter has been submitted as a paper in The Journal of Chemical Physics.

1.2.6 A Density Functional Aside

The exchange energy of a uniform electron gas which experiences a novel 2-parameter separation of the Coulomb interaction is derived as a local functional of the electron density. The 2 parameter range separating function allows separate control of where and how rapidly the Coulomb interaction is switched off, as opposed to conventional 1-parameter error function attenuators. The usefulness of the functional is briefly assessed by combination with a recently published pair of exchange and correlation functionals. The self-interaction error (SIE) of noble-gas dimer dissociation is found to be reduced while thermochemistry is relatively unperturbed. These results suggest that attenuator shape is a direction by which range-separated exchange functionals may be further improved. This Chapter has been published as a paper in Chemical Physics Letters [87].

Chapter 2

Many Fermion Algebra

2.1 Introduction

There are no mysteries in the first-principles of quantum chemistry, but many problems in realizing these principles as usable models. Often the premise of an idea can be expressed easily, but the path to code overwhelms human endurance. Shifting the effort of many-Fermion algebra onto computers is not a new idea [88]. Some fundamental early techniques like string-based configuration interaction [89, 90] already shifted symbolic efforts onto the computer. These ideas had some early impacts exploring the performance of high-orders of many-body perturbation theory [91]. Extensions of [92–94] similar ideas were later used to create exponential operator strings and consequently an (unfactorized) Coupled-Cluster(CC) code. Alongside these developments the diagrammatic bookkeeping techniques [95] which are so useful many-body theory were pushed to their limits [96, 97].

A symbol manipulation program requires an investment of time, and it isn't much of an "end in itself", and so until recent years these tools have been uncommon. The general CC program of Kallay and Surjan [98], was a substantial step forwards. Near the same time an ambitious project to use symbolic algebra to generate optimized correlation models began [99]. In recent years interesting demonstrations of these techniques have begun to yield models which would not have been imaginable without them. In particular it became possible to formulate multi-reference state-specific cluster theories of very high-order [98, 100], some local static correlation models [101], and implement explicitly correlated cluster models [102] which approach spectroscopic accuracy amongst other applications [59, 103, 104]. Outside of the quantum chemical community similar ideas are also being pursued [105].

In this paper we document a framework for the automated derivation and subsequent numerical evaluation of many-Fermion expressions. The algorithms presented feature a diagrammatic (Antisymmetrized-Brandow [106]) formalism which yields explicit equations, implicit treatment of permutational symmetry and a general contraction algorithm. The contraction algorithm introduced leverages sparsity on its arguments in a general way. This

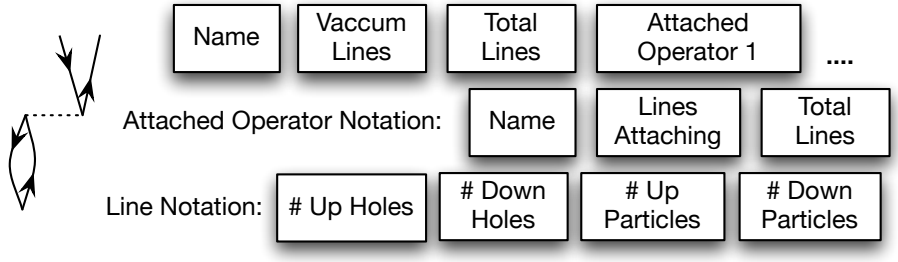
novel feature is especially important for high-rank correlation models, which would otherwise be simply too expensive for all but the smallest applications. Virtually any sort of localization scheme can be exploited, and we demonstrate that the cost of the resulting algorithm is insensitive to tensor rank. Subjects which have been the focus of other work [107–109] (factorization and loop structure) are not investigated, but the ideas described in those treatments can be leveraged in this scheme directly [108]. This work also presents the details which make our recently published local correlation models [101] reproducible, and hopefully extensible by other groups.

2.2 Many-Electron Algebra

Wick's theorem [27], which governs the construction of virtually any electron correlation model can be coded as a few recursive replacement rules: the Fermion anti-commutation relations. In *Mathematica* this can be achieved in less than 10 lines of code, and in principle it is enough to determine algebraic expressions for matrix elements of any operator. Beginning with sums over every index of a string of operators, Wick's theorem exploits a reordering of the creation and annihilation operators so that the terms which vanish because $a_p^\dagger a_p^\dagger = 0$ or $a_p a_p = 0$ can be easily eliminated. δ functions which restrict ("contract") two indices on different operators to be the same are introduced by the repeated application of the anti-commutators. In quantum chemistry "contraction" sometimes means making these restricted sum expressions and often means evaluating these restricted sum expressions. We will describe algorithms for both problems.

With a simple recursive sort implementation of Wick's theorem as described above one would not get very far because an exponential number of algebraically equivalent terms are generated by such a procedure. We offer the ten-line Wick's theorem in the supporting information to illustrate that. If one desires a truly general formalism which can handle high-rank tensors, contractions must be represented in a topologically unique way, only recording the pattern of connectivity between the tensors introduced by the δ functions. This amounts to employing a diagrammatic formalism, where the δ functions become lines between vertices. This work is based on the antisymmetrized Brandow-type diagrams which dominate the literature of coupled cluster theory and have been described many times [52, 110]. The resulting models are built on spin-orbitals.

Many readers may be unfamiliar with diagrams altogether, and this paper cannot replace the introductory reviews mentioned above, but we can try to express the basic idea [26, 111] which is conceptually simple. The four sorts of operators that occur in a second quantized picture of electrons are mapped onto 4 sorts of directed lines drawn away from a vertex. Creation ($a^\dagger$) lines are drawn above (up) and annihilation (a) lines are drawn down. If the operator involves a virtual ($a_b^\dagger$ or a_b) the line is given an arrow going upwards. Correspondingly if an occupied level is involved ($a_j^\dagger$ or a_j, the arrow is downwards. At the



$$\sum_{ck} \hat{T}_k^c \hat{V}_{ik}^{ac} \rightarrow \hat{R}_1, \text{ Denoted: } \{\hat{V}, (1,0,1,0), (1,1,1,1), (\hat{T}_1, (1,0,1,0), (1,0,1,0))\}$$

$$\text{Where: } \hat{T}_1 = \hat{T}_k^c = T_k^c a_c^\dagger a_k \tag{2.1}$$

Figure 2.1: An example of the diagram notation. Each box is a list of 4 integers (line notation) characterizing the normal-ordered many-fermion operator.

base of the diagram you imagine the reference $|ket\rangle$ and above the $\langle bra|$. The advantage of the diagram is that the "contractions" between indices, which can be written so many ways in a formula, are just one connection between matching arrows. The whole elegant notation is somewhat isomorphic to the popular children's toy called "Lego" where the arrows are the part of the brick which holds on to the one below.

Kallay and Surjan [112] introduced a very clear and compact notation for a diagram of the coupled cluster amplitude equations, which strongly influences the notation employed in this work. Operators are described by a name, a sequence of 4 integers (the number of up and down hole and particle lines they possess), and another quartet for the number of lines connecting to the Fermi vacuum (determinant $|0\rangle$). Diagrams are described by lists of operators. Inside each operator the other tensors to which this tensor is connected are listed by name, and another quartet of numbers for each sort of connection between them. Figure 1 depicts this process. A diagram can be written more than one way with this scheme, but contraction is fast.

Given a string of normal ordered Fermion operators represented in this fashion a matrix element between two vacuums can be obtained from the complete set of unique diagrams formed by the constituent operators as described in Algorithm 1. Unique in this context means: "Has the same number of connections of each sort between each pair tensors". This set is built up in the obvious fashion: operator-by-operator and line-by-line. Algorithm 1 is best thought of pictorially. Imagine two piles of Lego bricks left and right, which correspond to summed operators $\hat{H} = (\hat{F}_{oo} + \hat{V}_{oooo}....)$ and $e^{\hat{T}} = \hat{T}_1 + \hat{T}_2 + \hat{T}_1\hat{T}_2 + ...$ We are going to make the expressions for $\{\hat{H}e^{\hat{T}}\}_c$ which are all the unique structures that sit an $\hat{H}$

brick on top of a brick from the right. Select a brick a from the left pile (say $\hat{V}_{oovv}$). Next we pick up a brick from the right pile, say $\hat{T}_1$. $\hat{V}_{oovv}$ has an open down virtual which can connect with $\hat{T}_1$'s open up virtual when we place $\hat{V}$ on top of $\hat{T}$. We connect those bricks and set them aside. We grab another copy of $\hat{V}_{oovv}$ from the left pile and another copy of $\hat{T}_1$ from the right, then ask if there are any other connections we could sit $\hat{H}$ on $\hat{T}_1$. There are not any such new connections and so we pick up a new brick from the right and repeat the process until there are no new bricks on the right or left. In the supporting information Algorithm 1 is given as a *Mathematica* routine called **Kontract**[].

for Diagram $i \in Pool_2$ **do**
  for Line type $t \in (o^\dagger, o, v^\dagger, v)$ **do**
    if a has open t line **and** i has $t^\dagger$ open **then**
      Attach a with i by t and append to output
    end if
  end for
  Take unique union of pool and growing output;
end for

Algorithm 1: Contraction of Diagram a with $Pool_2$

2.2.1 Coupled-Cluster Theory

The coupled-cluster(CC) theory has distinguished itself [52] amongst approaches to the many-electron problem. This paper will apply the algorithm above to automate the derivation and implementation of these equations. It is useful to cast CC theory in a pseudo-variational language [58, 113, 114] even though it is not economical to optimize $e^{\hat{T}}|0\rangle$ as a variational trial wave-function [115] without the projective ansatze. In order to derive the projective amplitude equations from a variational principle we introduce a de-excitation operator $\hat{\Lambda} = \lambda^{a\cdots}_{i\cdots} a_i^\dagger ... a_a$, as a Lagrange multiplier for each amplitude equation. Given a Lagrangian pseudo-energy $\tilde{E}_c = \langle 0|(1+\hat{\Lambda})\{\hat{H}e^{\hat{T}}\}_c|0\rangle$ a set of non-linear equations for $\hat{T}$ can be derived by assuming this Lagrangian is stationary to variations in $\hat{\Lambda}$ (equation 2). The notation $\{\}_c$ means that $\hat{T}$ must share at least one line with $\hat{H}$ if the diagram is to be included, and $\{\hat{H}e^{\hat{T}}\}_c = e^{-\hat{T}}\hat{H}e^{\hat{T}}$.

$$\frac{\partial \tilde{E}_c}{\partial \hat{\Lambda}} = \langle 0|(a_i ... a_a^\dagger)\{\hat{H}e^{\hat{T}}\}_c|0\rangle = 0 \tag{2.2}$$

We have already walked through the generation of diagrams for $\{\hat{H}e^{\hat{T}}\}_c$. To make an amplitude equation from this pool we simply select any diagrams which would be closed when projected against the desired excitation manifold. In the notation of the code described in Figure 1, this means selecting any diagram whose vacuum lines (the second quartet of

indices) are that manifold, for singles: $\{1,0,1,0\}$, for doubles: $\{2,0,2,0\}$ etc.. A routine that performs these manipulations to produce the CC amplitude diagrams with the **Kontract**[] routine is also given in the supporting information as **AmplDiag**[].

With $\{\hat{H}e^{\hat{T}}\}_c$ in-hand generation of the coupled cluster Λ(multiplier) diagrams, effective density matrices [113], Jacobian [59] etc. are straightforward matters of making the correct pools of diagrams and operating this algorithm on them. For example, let us consider how to generate the Λ equations which arise from the following:

$$0 = \frac{d\tilde{E}_c}{dT} \rightarrow 0 = \langle 0|(1+\Lambda)[\{\hat{H}e^{\hat{T}}\}_c, a_p^\dagger ... a_q]0\rangle \tag{2.3}$$

$\hat{H}$ is contracted with $e^{\hat{T}}$ to construct $\{\hat{H}e^{\hat{T}}\}_c$, as described above and symbolically in algorithm 2. Imagining the "pile of bricks", we have made for $\{\hat{H}e^{\hat{T}}\}_c$ we now need to sit $(1+\hat{\Lambda})$ on top of all the pieces in this pile in every unique way. The algorithm (2) is made a little simpler to write down by noting at most quadratic amplitudes and linear multipliers occur in Eq(3). In the language of the brick analogy it means that our $e^{\hat{T}}$ pile contains structures with at most 4 bricks. We don't need to worry about the 5 brick piece $\hat{T}_1\hat{T}_1\hat{T}_1\hat{T}_1\hat{T}_1$ because it cannot make a connected diagram with $\hat{H}$ that only has 4 lines to touch. So we make every unique product of amplitudes up to the 4^{th} order in the first part of algorithm 2, contract these with $\hat{H}$ and keep only the unique ones. Then we loop over possible $\hat{\Lambda}_n (0 \leq n \leq N)$ and contract each. This pool is finally then projected on the right for the desired excitation level of multiplier, ie: $\{0,1,0,1\}$ for the $\hat{\Lambda}_1$ equation etc. This algorithm is given in the supporting information as **LamDiag**[]

One need not generate the CC Lagrangian, $(1+\Lambda)\{\hat{H}e^{\hat{T}}\}_c$, in its entirety to construct the multiplier diagrams, and for very high-rank equations that might not be advisable. We have coded and described it in this way because with this quantity in hand most interesting CC equations [116, 117] are just projections against the appropriate manifold. Generalization to analytic second derivatives is also possible [118]. For some response properties it is useful to write the Lagrangian [58, 119] as a dot product between $\hat{H}$ and effective one(γ_p^q) and two (Γ_{pq}^{rs}) particle density matrices. In our brick analogy, we look at every structure from $(1+\Lambda)\{\hat{H}e^{\hat{T}}\}_c$ which has no open lines; meaning it contributes to the energy when we put the reference on either side. We can see the hamiltonian brick inside. If we pull that hamiltonian brick out, the shell we leave behind is Γ, which now has the lines which were shared with the Hamiltonian open.

$$\tilde{E}_c = \hat{F}_p^q \gamma_p^q + \hat{V}_{pq}^{rs} \Gamma_{pq}^{rs} \tag{2.4}$$

To obtain diagrams for those density matrices from $(1+\Lambda)\{\hat{H}e^{\hat{T}}\}_c$, one only needs to first select the diagrams which are closed on both ends, order them by the $\hat{H}$ block they contain (if desired), and then delete the $\hat{H}$ operator and leave the lines dangling to the vacuum. This is also automated in the supporting information in routine: **DMDiag**[].

Converting a pool of diagrams into the usual strings of sums over basis functions is

Create all disconnected diagrams: $T_{n_1}^{p_1} T_{n_2}^{p_2} T_{n_3}^{p_3} T_{n_4}^{p_4}$ s.t. $0 \leq p_i \leq N$ **and** $\sum p_i \leq 4$
for diagram $h_i \in \hat{H}$ **do**
  for diagram $D_j \in$ amplitude pool **do**
    Contract(h_i, D_j)
    Discard disconnected diagrams
    Take unique union of results and growing $\{\hat{H}e^{\hat{T}}\}_c$.
  end for
end for
for $\Lambda_k, 0 \leq k \leq N$ **do**
  for diagram $D_j \in \{\hat{H}e^{\hat{T}}\}_c$ **do**
    Contract(Λ_k, D_j) and gather to form $(1 + \Lambda)\{\hat{H}e^{\hat{T}}\}_c$
  end for
end for
for $0 \leq l \leq N$ **do**
  Project $(1 + \Lambda)\{\hat{H}e^{\hat{T}}\}_c$ onto $|\mu_l\rangle$ (the Λ_l residual equation)
end for

Algorithm 2: Construction of Multiplier Diagrams, rank N

ironically much more tedious to automate, but can be important for the beginner. To begin, construct a skeletal list of tensors with dummy indices. Then for each internal line on each internal tensor collapse two dummy indices from the appropriate tensors into a single summed index. Sign is easily accomplished by listing the second-quantized version of a diagram and counting the number of intersections between lines connecting the same index. In *Mathematica*, one can just call **Signature**[] on a list of indices. The factor arising from the number of equivalent tensors/lines is computed after the diagram has been expressed as indexed tensors because its easier to establish subgroups of tensors and indices. Permutation symbols aren't generated at all because anti-symmetry of tensors is assumed by the contraction algorithm. In the supporting information these tasks are performed by **DtA**[] with some examples. It remains to describe how the resulting sums of high rank tensors will be achieved, and we now turn to this question.

2.3 Tensor Representation and Permutational Symmetry

Computational performance is heavily dependent on the way data is arranged in memory [107]. Ideally contraction can be mapped onto matrix multiplication with minimal reordering, and for situations where tensors may represented as simple arrays in memory there are codes which can optimize tiling for peak performance [108, 120]. However we are

interested in situations where dense representations of data are not tractable, and we also need to deal with high permutational symmetry. One solution is to represent tensors as strings of second quantized operators [90]. In this way symmetry concerns are easily dealt with, but vectorization is sacrificed. We make a similar compromise, but choose a slightly different format: a data structure which handles its own symmetries.

For the lesser ranks there are some standard formats [121] interfaced to sparse linear algebra packages which generalize the common matrix formats. However for high ranks there is an ordering problem which results if there is some "leading dimension" which can be iterated over more rapidly than the others. Our first approach was to introduce an n-dimensional net in which each node knew the location of its nonzero neighbors in each dimension. Sparse contraction can then be effected quite simply by traveling in the correct direction along this net. Memory bandwidth limited the capability of this idea insofar as we implemented it because element addition requires traversing each dimension of the grid to find the appropriate niche. Nevertheless this structure was used to initially debug the amplitude equations of the PQ [86] model.

If a sparsity pattern exists within the data because of spatial locality [56, 122, 123] or a pair constraint [101, 124–126], or some other local model [127] this must also be leveraged in the storage scheme. Desiring the most general code possible, we adopt the simple sparse coordinate representation: $\{i_1, i_2, ..., \text{Value}\}$. The Fermion index symmetries are easily exploited in this representation by representing only those coordinate-value pairs which are sorted with respect to an ordering on these symmetries. Each tensor data object is given an attached symmetry object which lists the sets of dimensions which are antisymmetric to one another. Because they are now a natural part of the data structure, contraction actually becomes simpler because there is no need to determine lengthy permutation symbols. Before the evaluation of a contraction the symmetries of the two argument tensors are used to establish the symmetry of the result if the result is an intermediate. If the result is an amplitude, these symmetries are already known, and only the unique output coordinates obeying the exclusion principle are incremented.

The storage of the coordinates adds substantial memory overhead, but in the most general case there is no alternative. The costs of these symbolic operations are poised to decrease, as computer scientists develop optimized, vectorized [128, 129] and parallelized sort algorithms. Random-access writes are made cheapest with a hash-map data structure. The data is usually written to a simple array when it must be sorted for the algorithms of the next section. As development in high-rank methods continues, tensor decomposition approximations will likely be used by ourselves and others to reduce the cost of storage and contraction.

2.4 Contraction

We assume a pair-wise factorization of each term of the electron correlation theory, both because this has been given excellent attention by other authors [108,120] and because one may achieve the "correct" cost exponent with virtually any pairwise factorization of a nonlinear contraction. An example is useful, and so consider this term of the CCSDTQ $\hat{\Lambda}_2$ residual equation:

$$\frac{1}{16}\hat{\Lambda}^{a,b,p_{11},p_{12}}_{i,j,h_7,h_8}\hat{T}^{p_9,p_{10},p_{11},p_{12}}_{h_5,h_6,h_7,h_8}\hat{V}^{p_9,p_{10}}_{h_5,h_6} \rightarrow \frac{1}{16}\hat{\Lambda}^{a,b,p_{11},p_{12}}_{i,j,h_7,h_8}I^{p_{11},p_{12}}_{h_7,h_8} \rightarrow \hat{\Lambda}^{ab}_{ij}. \tag{2.5}$$

We will use the contraction of $\hat{\Lambda}$ (called Tensor 1) with the intermediate (Tensor 2) as our example. Now to complete the calculation of an electron correlation model all that remains is to evaluate an anti-symmetrized sum over internal dimensions with matching indices. An anti-symmetrized sum means that the $\hat{\Lambda}^{ab}_{ij}$ must be antisymmetric with respect to permutations of the index labels (because it represents a Fermionic wavefunction), and our contraction algorithm is made responsible for ensuring this. It quickly becomes impossible to perform the products which are identical to each other by antisymmetry more than once because these grow exponentially with tensor rank. We will now discuss how these can be avoided using the storage scheme from the previous section. We often must talk about whether an index is summed over or mapped onto the result. The former are called "internal" and the latter "external".

Ignoring anti-symmetry the "contraction" itself is uniquely defined (Alg number, line number) [Alg.3:1] by a signed factor, $C = \frac{1}{16}$, a matched list of contracted dimensions $\{(d^k_1, d^k_2)\}$, and another which determines the source dimension of a result dimension $\{(r^k, s^k)\}$. If we number the dimensions of a tensor: $T^{n+1,n+2\ldots}_{0,1,\ldots n}$ then these lists would be denoted: $\{(2,0),(3,1),(6,2),(7,3)\}$ and $\{(0,\Lambda_0),(1,\Lambda_1),(2,\Lambda_4),(3,\Lambda_5)\}$ in our example. Because only elements with sorted indices have been stored and the other elements related to these by permutation are implied, the dimensions in these lists must be permuted for the symmetries of each argument. This process is described generally in algorithm 3 and specifically for our example below.

In our example both tensors have anti-symmetries between the contracted occupied and virtual indices. If we simply applied every permutation of both tensors to $\{(d^k_1, d^k_2)\}$ some multiplications would be performed more than once. For example the permutations $I^{p_{11},p_{12}}_{h_7,h_8} = I^{p_{12},p_{11}}_{h_8,h_7}$ and $\hat{\Lambda}^{a,b,p_{11},p_{12}}_{i,j,h_7,h_8} = \hat{\Lambda}^{a,b,p_{12},p_{11}}_{i,j,h_8,h_7}$ would only permute the contraction list and give, $\{(3,1),(2,0),(7,3),(6,2)\}$. Likewise there are several $\{(d^k_1, d^k_2)\}$ which would be empty because of the sorting on the indices, like $\{(2,1),(3,0),(6,2),(7,3)\}$. In order to leverage symmetry we must maintain the order of $\{d^k_1\}$ and $\{d^k_2\}$.

For the special case of the CC equations this step is simplified by the fact that one tensor's symmetries will always contain the symmetries of the other. For the sake of the routine that follows assume that the symmetries of $\{d^k_2\}$ map to a subset of the symme-

tries on $\{d_1^k\}$ (as they do in our example) and in the other situation reverse the labeling. First, obtain the number of redundant contractions from tensor 1 and multiply the factor by the inverse [Alg. 3:9]. This number, $Order(S_2)$, is the product of the factorials of the number of contracted dimensions which fall in each symmetrical set. In our example there are two occupieds which are contracted and antisymmetrical to one another and likewise two virtuals so the factor is $\frac{1}{2!2!}$. Next, apply all the permutational symmetries of tensor 1, denoted σ_k, to $\{d_1^k\}$ and keep only those results $\{c_1^k\}$ which are ordered, adjusting the sign accordingly. For example beginning with $\{d_1^k\} = \{2,3,6,7\}$ the permutation $\sigma_1 \cdot \hat{\Lambda}_{i,j,h_7,h_8}^{a,b,p_{11},p_{12}} \rightarrow (-1) * \hat{\Lambda}_{i,j,h_7,h_8}^{a,p_{11},b,p_{12}}$ produces another unique $\{c_1^k\} = \{2,3,5,7\}$ [Alg. 3:9]. Often σ produces a redundant $\{c_1^k\}$, and these are not added to the contraction list [Alg 3: 6,16]. When this process is finished for the example there are 36 unique $\{c_1^k\}$ out of the 576 permutations which exist on Λ_4. Likewise apply all the symmetries of tensor 2 to $\{d_2^k\}$ and keep only those which are ordered in tensor 2 and map to a $\{c_1^k\}$ which is sorted in the order of tensor 1, adjusting the sign accordingly [Alg. 3:19]. In our example there is only one such $\{d_2^k\}$: $\{0,1,2,3\}$. It's immediately obvious how important the symmetries are. We would be doing 576*4 = 2304 loops in the non-symmetric algorithm, instead of 36. If our algorithm ignored symmetry and stored the tensors in a redundant fashion we would be wasting the same amount of effort. Given loops over the lists of contracted dimensions generated by algorithm 3, $\{c_1^k\}$ and $\{c_2^k\}$, symmetry has been taken care of, and exploited. We need only evaluate a simple sum when these contracted dimensions match and sort the index of the result, which makes the greater task of sparse contraction somewhat simpler.

In local electron correlation models [57,122,130] amplitudes are so sparse that the polynomial scaling exponent of cost with respect to system size can often be decreased by several orders if floating point effort is not squandered on zero multiplications. One must somehow identify the nonzero sets of matching internal dimensions, $\{d_1^k\}$ and $\{d_2^k\}$, on each partner without explicitly looping over both tensors or assuming every possible $\{d_1^k\}$ is nonzero. In yet another phrasing, alignment must be achieved between the internal dimensions of the two contracted tensors to realize sparse contraction. Previous work on rank 3 and 4 tensors has employed sorting [131] or simulated annealing [57] to achieve this alignment. Let's make this concrete by returning to the example:

$$\sum_{h_7,h_8,p_{11},p_{12}} \frac{1}{16} \hat{\Lambda}_{i,j,h_7,h_8}^{a,b,p_{11},p_{12}} I_{h_7,h_8}^{p_{11},p_{12}} \tag{2.6}$$

In a non-sparse correlation model this sum would be achieved as four nested loops over the summed indices, and an 8^{th} order amount of effort would be expended regardless of the number of nonzero elements. A first guess at the sparse analogue might be to loop over nonzero Λ and nonzero I, performing multiplication when they match on $\{h_7,h_8,p_{11},p_{12}\}$. This algorithm's cost grows with the product of number of nonzero entries in each tensor, even if there are no non-zero multiplications to be performed. If the number of nontrivial elements in each grows in a greater-than-linear fashion, then this algorithm will also grow

Input: (C, $\{d_1^k\}$,$\{d_2^k\}$, $\{(r^k, s^k)\}$)
for $\sigma_1 \in$ One's anti-symmetries **do**
  Permute $\{c_1^k\} = \sigma_1 \cdot \{d_1^k\}$
  if $\{c_1^k\}$ disobeys sorting of One or Two **then**
    Continue loop
  else if $\{c_1^k\} \in$ result **then**
    Continue loop
  else
    add (Sign(σ_1)$\frac{C}{Order(S_2)}$, $\{c_1^k\}$), to result list 1.
  end if
end for
for $\sigma_2 \in$ Two's anti-symmetries **do**
  Permute $\{c_2^k\} = \sigma_2 \cdot \{d_2^k\}$
  if $\{c_2^k\}$ disobeys sorting of Two **then**
    Continue loop
  else if $\{c_2^k\} \in$ result **then**
    Continue loop
  else
    add (Sign(σ_2), $\{c_2^k\}$), to result list 2.
  end if
end for

Algorithm 3: Exploitation of Permutational Symmetry

unphysically with system size. However if the two tensors are given an order before the summation, then by following the elements in sequence we can establish if there is a non-zero partner without any effort.

The indices of nonzero elements on the dimensions $\{h_7, h_8, p_{11}, p_{12}\}$ in Λ and I, called $\{c_1^k\}$ and $\{c_2^k\}$ respectively, which we initially have stored as a simple unordered list, are sorted in a lexical order. Suppose that $\{d_1^k\}$ was $\{2, 3, 6, 7\}$ as for our example. Then after this sorting the list of elements representing Λ are strictly ordered as: $(X, X, 1, 2, X, X, 3, 4) < (X, X, 1, 2, X, X, 3, 5) < (X, X, 6, 2, X, X, 3, 5)$ etc. This ordering is exploited by the contraction code to avoid a nested loop over the elements of both tensors, as described in the inner while() loop of algorithm 4. The loop begins with two pointers at the least elements of both tensors. If the two elements do not match on the contracted dimensions the loop advances the pointer which lexically precedes the other on the contracted dimensions and otherwise the floating point effort of contraction is performed as usual. *By construction the number of comparison/skip operations cannot exceed the sum of the number of nonzero elements in both tensors.* Regardless of the sparsity pattern this means that the cost of algorithm 4 only grows with the number of non-trivial products in each argument. In the supporting information the inner while() loop is implemented for simple coordinate tensors

as a C++ routine, **SparseContract**(). Quantum chemists are used to high-performance matrix algebra packages, but there is a growing literature of computer science devoted to more symbolic tasks [132] like sorting. Some codes have even been developed to perform these operations on highly vectorized graphics processors, and these should be exploited in future work.

At this point most of the notation in the algorithm 4 has been described. We also use a shorthand for the $\{c^k\}$ indices of the i^{th} element in a block : $\{c^k\}_i$. For example if $Block_1$ were two elements: $\{(7,6,1,2,8,9,3,5),(3,1,6,2,9,0,3,5)\}$ and $\{c_1^k\}$ was $\{2,3,6,7\}$ then $\{c_1^k\}_1$ is $\{1,2,3,5\}$. It is often useful to separate the elements of the tensor into blocks because large numbers of indices are incompatible with one another, for example because of spatial or spin symmetries. Algorithm 4 mentions this idea because in the application which follows we have enforced a sparsity on tensors by blocking them into local groups of orbitals. We assume that the reader can group elements of tensors into blocks based on the dimensions which are contracted, and can decide whether two blocks will contribute to the result with some compatibility rule.

for each $\{c_1^k\}$ from Algorithm 3 **do**
  Block One on $\{d_1^k\}$
  for each $\{c_2^k\}$ from Algorithm 3 **do**
    Block Two on $\{c_2^k\}$
    for $Block_1 \in$ Blocks of One **do**
      Extract all compatible elements of Two into $Block_2$.
      (Note: $Block_1$, $Block_2$ raster sorted on $\{c_1^k\}$, $\{c_2^k\}$)
      Index in $Block_1$, $i_1 = 0$ **and** Index in $Block_2$ $i_2 = 0$
      while $i_1 <$ Size($Block_1$) **and** $i_2 <$ Size($Block_2$) **do**
        if $\{c_1^k\}_{i_1} < \{c_2^k\}_{i_2}$ (Raster sense of $<$) **then**
          $++i_1$
        else if $\{c_1^k\}_{i_1} > \{c_2^k\}_{i_2}$ **then**
          $++i_2$
        else if $\{c_1^k\}_{i_1} = \{c_2^k\}_{i_2}$ **then**
          Sort result index and keep the sign.
          Result += (Value at i_1)*(Value at i_2)*(Sign from index sort)*C*(Sign and symmetry factor from $\{c_1^k\}$, $\{c_2^k\}$)
        end if
      end while
    end for
  end for
end for

Algorithm 4: Sparse contraction loop

2.4.1 Rank insensitivity

To give a brief demonstration of the usefulness of this contraction algorithm consider the following term which occurs in the CC/CI equations of any order ≥ 2:

$$\hat{T}^{a_3,...a_n,b_1,b_2}_{i_1,...i_n}\langle a_1 b_1||b_2,a_2\rangle \rightarrow \hat{R}^{a_1,...a_n}_{i_1,...i_n} \tag{2.7}$$

The naive cost of such a contraction is $O(o^n v^{n+2})$, and so already at quadruples this term makes CC intractable for all but the smallest systems. The feature of the contraction algorithm developed in this section is that the CPU cost of this term can be limited by defining some subsets of $E \subset \{i_1,...i_n,a_1,...a_n\}$, $I \subset \{b_1,b_2\}$ and then the CPU cost of the contraction is limited to $O(E*I)$. In our pair-based models [86,125], we limit the growth of these subsets using a particular labeling, but the contraction algorithm only requires that one be able to quickly compute if an index belongs in E or I, so we could imagine labeling them by spatial domain etc. For the purposes of demonstration we will limit the growth of $\{E\},\{I\}$ with system size to some simple polynomial with a pair-labeling. The indices of the amplitudes, integral and residual are each spanned by a certain number of mutually exclusive orbital sets $\{o_\alpha,o_\beta,v_\alpha,v_\beta\}$ which together span an entire orbital active space.

Supposing that $\hat{T}$, $\hat{R}$ and $\hat{V}$ are restrained to two pairs, the scaling of this contraction is reduced to third order for $n \geq 2$. In the three-pair case this summation takes on at most 4 pair labels, and so it ideally scales with the 4^{th} order of system size regardless of rank. For perspective we provide the size of the three-pair amplitudes in Figure 2. The nonzero dimensions of all amplitude ranks between 2 and 6 vary over roughly an order of magnitude between the (22,22) and (40,40) active spaces (where (X,Y) denotes the space formed by X electrons in Y orbitals). For a growing number of pairs we separately timed this contraction for amplitudes between doubles and hextuples with a three-pair approximation.

As the rank grows, the pre-factor due to coordinate storage grows, as does the cost of sorting $\hat{T}$ and $\hat{V}$ on $\{b_1,b_2\}$ for each permutationally unique contraction. Furthermore the sorting cost is not strictly linear; it is sub-quadratic (NLogN). Still using algorithm 4 we achieve virtually the same cost-exponent for the hextuples as we do for the doubles (Figure 3). Interestingly, the performance is less predictable in the large-space, high-rank region of the plot. We can say with certainty this is not due to any peculiarities of a chemical system (all tensors have been filled with noise).

2.5 Conclusions

The computational resources available to scientists interested in physical modeling continue to grow in a non-linear fashion (albeit now in parallel), and so there are ever-more interesting applications of computable theories. However the number of man-hours which can be devoted to such projects are much the same today as they were a long time ago, and automation of the relevant mathematics will continue to spur new developments. To

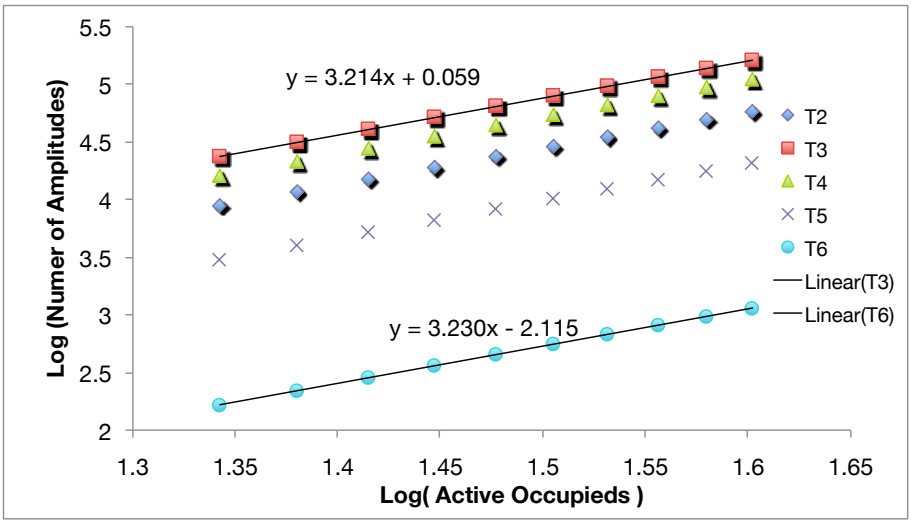


Figure 2.2: Number of (symmetrically unique) elements in $\hat{T}_n$ for $2 \leq n \leq 6$ for the calculations performed in Figure 3 and linear least-square fits. The number of amplitudes are constrained to grow in a cubic fashion for any rank with a three-pair constraint. Note that $\mathrm{Dim}(T_2) = \mathrm{Dim}(\langle a_1 b_1 || b_2, a_2 \rangle)$ since a PP-active space is chosen

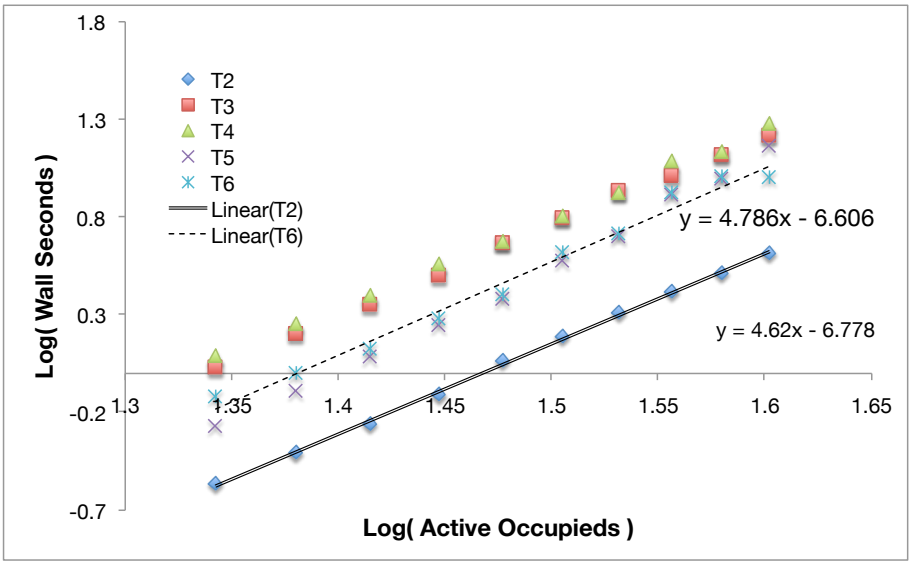


Figure 2.3: Wall time consumed in calculating equation 4 for various $2 \leq n \leq 6$ using algorithm 4, and the cubically growing tensors of Figure 2. Horizontal axis is the number of occupied orbitals, and the space has the same number of virtuals. The cost exponent for $n = 2$ and $n = 6$ are estimated by the slope of the linear fits.

further reduce the amount of time spent by others seeking to reproduce these results we include the routines described in *Mathematica*/C++ code as we have implemented them. Extensions of these routines to other sorts of diagrams that integrate-out spin, or extended normal ordering [133] to generate a multi-reference cluster theory would be quite useful and interesting.

The algorithms presented were conceived with an emphasis on generality, and the goal of correct scaling of cost (up to an exponent). They have proven useful for prototyping electron correlation theories which would otherwise be intractable, but a large additional pre-factor could be recovered through close attention to the introduction and re-use of intermediates, memory layout and vectorization. The growing divide between highly symbolic and multilinear correlation models, and the standard libraries at a computer scientist's disposal will need to be bridged.

It would also be quite valuable to develop a general scheme for representing sparse, high-rank data without the memory bandwidth costs of the coordinate representation. If the sparsity has at least some structure (like a pair-constraint), then these can be used to address loops, but such a structure is not always available or so simple to order. Multi-linear compression technologies [134–136] should prove useful in this respect.

Chapter 3

Numerical Stability of Unrestricted Correlation Models

3.1 Introduction

Many electron models often sidestep the complicated structure of separating electron pairs by allowing approximate wavefunctions to break spin symmetry. The spin-unrestricted wavefunction [28] has the strength that it matches the energy of non-interacting molecular fragments in the limit of dissociation. As a consequence, spin-unrestricted Hartree-Fock has become a standard reference for many sorts of correlated treatments including high-quality coupled-cluster (CC) methods [52]. When correlated models are used to optimize unrestricted orbitals [137] strong correlations between paired electrons are described redundantly, and the two competing descriptions can cause difficulties. Multiple solutions are one manifestation, and another which we have unfortunately encountered while developing orbital-optimized cluster models [138] is singular behavior of the amplitude equations.

Because of their great physical impact, the existence and character of solutions to the coupled cluster equations are of interest in themselves. Despite the non-linearity and high dimension of these equations much is now known about their solutions thanks to the efforts of several groups [139–146]. Our focus in this paper is much more quotidian, we simply explore why we were unable to combine valence-space CC with completely unrestricted orbitals. We find, to our surprise, that the CC equations are quite generally singular in the valence space if the one-particle basis is totally spin-polarized in an ROHF-sense, meaning that the occupied space is comprised of "valence" spin-orbitals perfectly localized on fragments and "core" spin-orbitals whose α and β parts are identical.

Poor numerical condition of the CC amplitude equations at restricted dissociation is not new to any practitioner, but with unrestricted orbitals they are usually well-behaved.

This case is noteworthy for a few reasons: it is a general feature of combining unrestricted orbitals for dissociation with a correlation model in the unrestricted space, it is easy to find "false solutions", and we can offer a possible solution. The Jacobian and closely related stability matrix [147] of the CC equations will be examined for these purposes. The former has been examined before for the case of restricted linear [140] and multi-reference [142] cluster theories.

Our analysis begins with a curious set of calculations on the N_2 molecule. Orbitals were prepared such that the only the 6 valence electrons were unrestricted and localized on each fragment so that a dissociation curve could be followed inwards from the correct asymptote at a separation of a few Angstrom. We attempted to apply CCD in the active space, and with simple amplitude iterations convergence was sluggish. Inspection of the amplitudes revealed that they were a scaled unit vector. Employing a standard DIIS [148] solver the correction vector was zeroed in a few iterations, but again the amplitudes appeared unphysical. The gradient obtained from these amplitudes took strange orbital optimization steps, and the same results were found independently in our two totally independent implementations of the theory, and so we proceeded to examine them further.

3.2 Results

In all that follows CC calculations were performed in the active space formed by the pairs relevant to a bond-dissociation process. The orbitals were unrestricted such that each fragment reproduces the ROHF limit, this will be called the unrestricted limit. It is important to stress that our conclusions only hold for an cluster model where all occupied spin-orbitals are localized on high-spin fragments. They do not hold for the usual UHF orbitals because of the spin-polarization of pairs which would be restricted in the ROHF case.

Many attempts were made to obtain a solution by continuation. At N-N separations of less than $1.7\mathring{A}$ the S^2 of the reference determinant is well below the spin-polarized limit (3) even for simple Hartree-Fock orbitals and the coupled cluster equations can be solved easily, but this solution cannot be followed to the dissociation limit. Even at the dissociation limit, one can easily solve the CC equations if a single term linear in the amplitudes is neglected. We attempted to continue this solution by introducing a simple continuous deformation parameter λ as the coefficient of a term and solving the equations along the real axis between $\lambda = [0, 1]$, but were unsuccessful. Convergence stagnated, but no single element of the amplitude diverged. A similar situation has been observed in some other studies of singular CC equations [141]. The same attempts were made for several other dissociation problems (Ethene, H_2, etc.) with the same results. Further analysis of homotopies [143–146] can completely characterize the solutions of non-linear equations should they exist, and can establish the precise identity of a non-linear singularity (pole, branch,

pinch etc.), but any such distinction is of mathematical, not physical, concern as was established the pioneering work of others [139] (c.f. Section IV(d)). We will instead focus on firmly establishing the singularity of the Jacobian, the scope of the problem, developing a similar set of well-conditioned equations, and heuristic understanding of this situation.

The Jacobian characterizes the response of the coupled cluster equations to a linear perturbation:

$$\frac{\partial^2 \tilde{E}}{\partial T \partial \Lambda} = \frac{\partial \langle \Phi_{uv}^{\beta\gamma} | \{ \hat{H} e^T \}_c | \Phi_0 \rangle}{\partial t_{lm}^{\lambda\mu}}. \tag{3.1}$$

This matrix is the size of the amplitude vector squared; if one of it's eigenvalues should become non-positive either the equations have no well-behaved solution, or solutions meet at this point. In either of the previous cases the amplitude vector to which the the Jacobian belongs should not be regarded a good approximation to a physical ground state. Furthermore one may say that as a linear approximation of an amplitude iteration step, the "condition number" (the ratio of the Jacobian's smallest and largest eigenvalues) measures how rapidly the equations can be solved by iteration. This non-hermetian matrix has been derived and coded into programs many times [149–151] in the history of quantum chemistry. For the purposes of this paper we produced a computer-generated implementation, as have others [152, 153]. The results of the automatic implementation are complemented by a separate program derived and coded by hand for the cases of our local cluster models with which all stability matrix calculations were performed. In all cases the Jacobian was explicitly constructed and diagonalized to avoid the art of guess construction.

Figure X depicts the lowest eigenvalue of the CC Jacobian for the case of ethene dissociation as obtained by explicit diagonalization. The orbitals were prepared at the ROHF dissociation limit, then allowed to restrict as the fragments coalesce. Several sets of amplitudes were examined: the null guess (alternatively this can be considered the Linear Coupled Cluster (LCC) Jacobian), the MP2 guess, the best amplitudes which can be reached by simple iteration (as seen in the figure), those produced by DIIS, and even amplitudes iterated from random noise, all to the same effect. The increase of the lowest eigenvalue on the inclusion of single excitations is noteworthy. Performing the same exercise for analogous dissociation problems produces the same results. By $5\mathring{A}$ the condition numbers of all of these dissociation processes are so large that convergence seems impossible with double precision arithmetic. Of the cases examined Mo_2 is the most stable, with a smallest Jacobian eigenvalue of roughly $1 * 10^{-4}$ at $7\mathring{A}$.

If the active space is expanded so that not all the orbitals are spin-polarized, the equations become immediately well-behaved; the resulting Jacobian eigenvalues are strictly positive at any displacement. The reader is undoubtedly familiar with the reasonable condition of UCC calculations and so this should be evidence enough. Having converged

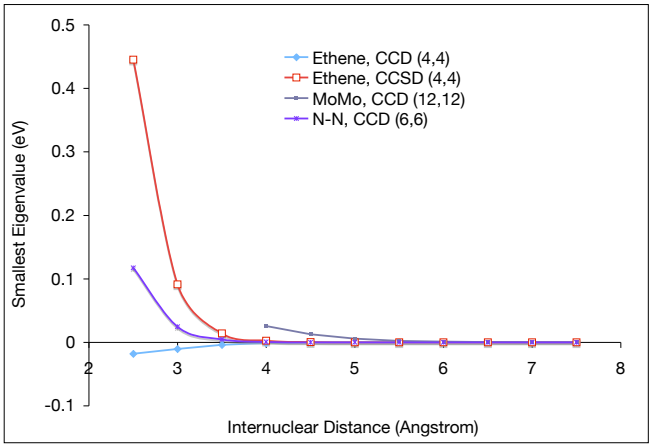


Figure 3.1: Smallest Jacobian eigenvalue for various fragment dissociations, in the 6-31G basis for all cases except Mo_2.

these amplitudes for the case of ethene (12,12), they were projected on the completely unrestricted (4,4) active space as a guess (the orbitals are unchanged between these two calculations only the amplitude space) and the singularity remains whether one tries to converge from this guess, or immediately diagonalizes the Jacobian. Based on these results and the previous observations we argue that the singularity can be understood with the linear part of the Jacobian which does not depend on the amplitudes. Inspection of the fragment orbitals provides another simple argument, all non-linear CCD terms depend on integrals of the sort $\langle oo||vv\rangle$ and for these fragment localized spin-orbitals (Figure X) these are separated across space and vanishing.

The coupled cluster stability matrix contains information very similar to the Jacobian, but can be used to understand the convergence properties of the iterative process we rely upon to solve these equations. Surján [147] and coworkers recently published work examining this matrix for several solutions of the CC equations along dissociation curves. Their results showed that the CC equations may exhibit a diverse range of iterative behavior (convergence, chaos and divergence) if manipulated by a denominator shift, and that extrapolative methods like DIIS [148] can be misleading as they seem to converge on what appear to be stable fixed points. We reiterate the formulas for this matrix given CCD and the usual partitioning. One can see that it is essentially the Jacobian dressed by factors which reflect the conventional form we use to solve the CC equations.

$$J^{22}_{uv\beta\gamma,lm\lambda\mu} = \delta_{ul}\delta_{vm}\delta_{\beta\lambda}\delta_{\gamma\mu} - \frac{\partial\langle\Phi^{\beta\gamma}_{uv}|\{\hat{H}e^{T}\}_{c}|\Phi_{0}\rangle/\partial t^{\lambda\mu}_{lm}}{f^{\beta}_{\beta} + f^{\gamma}_{\gamma} - f^{u}_{u} - f^{v}_{v}} \tag{3.2}$$

The Lyapunov exponents are the central object of this analysis, the logarithm of the Stability Matrix's eigenvalues. If these should equal or exceed zero iterations are not convergent. We turned to this tool because we wanted to understand what was occurring when simple iterations would stagnate at very small residual values. Table X depicts the results for several small molecules obtained with a hybrid of our local CC methods: The Perfect Pairing ($\hat{T}_{PP} = \sum t^{i^*\bar{i}^*}_{i\bar{i}}\hat{a}^{\dagger}_{i^*}\hat{a}^{\dagger}_{\bar{i}^*}\hat{a}_{i}\hat{a}_{i}$) [61, 137, 154, 155], and Imperfect Pairing ($\hat{T}_{IP} = \hat{T}_{PP} + \sum_{i\neq j} t^{i^*j^*}_{ij}\hat{a}^{\dagger}_{j^*}\hat{a}^{\dagger}_{i^*}\hat{a}_{j}\hat{a}_{i} + t^{j^*i^*}_{ij}\hat{a}^{\dagger}_{i^*}\hat{a}^{\dagger}_{j^*}\hat{a}_{j}\hat{a}_{i}$) [65, 66, 68, 124] models. Note that even if the species is asymmetrical the same trend is observed.

These results firmly establish the singularity of the CC equations in the unrestricted limit, but what meaning, if any, does this have for UCC as it is practiced when complete spin polarization is almost never the case? We can imagine one situation when this should be kept in mind, if the restricted core electrons are cut from a calculation by a pseudopotential the results are essentially the same as those which would be obtained from the minimal active space. Figure X depicts the CCD Jacobian's lowest eigenvalue for the dissociation of Mo_2 with the CRENBS basis and matching pseudo-potential.

3.3 Modified Equations Which are Well-Conditioned

Since the instability and ill-conditioning of the CC amplitude equations at the unrestricted limit has been established, we seek to restore solubility to these cluster models when the appropriate ROHF fragment orbitals are desired. Here we will discuss a few possible solutions based on the idea of regularization [156]. We will rate them based on a simple set of criteria: a) does it stabilize the CC amplitudes, b) is it simple to define and implement, c) how heavily does it affect the energetics of the molecular system at equilibrium, and d) will it allow us to optimize orbitals with active space CC Lagrangian methods.

The first and simplest correction is to add a constant denominator shift. As shown by Surján et al. [147], this enables us to make the amplitude equations stable. However, it requires a constant shift of at least 12 kcal/mol to be able to optimize the orbitals along the entire ethene dissociation curve. This is a very strong penalty near equilibrium bond lengths where the amplitude equations are usually well-conditioned. We also took a non-linear equation solving approach and tried to identify the source of the singularity, eliminate it, solve the system when it is non-singular, and follow a homotopy back closely [157]. In the section below we attribute the singularity to the structure of the block

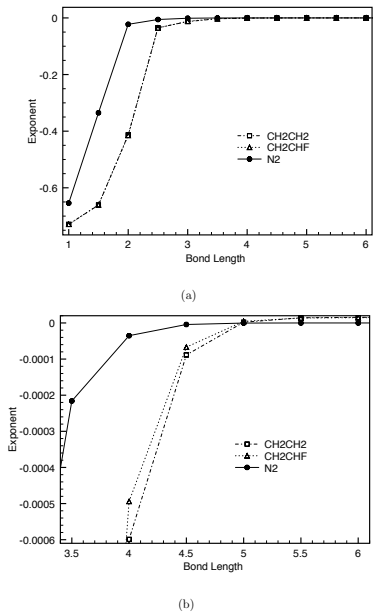


Figure 1: The change in the largest Lyapunov exponent for a stability analysis PP and IP exchange type amplitudes for 3 example molecules using the IP energy Ansatz on the PP orbitals in a minimal active space with the 6-31G* basis (a). (b) is an enhancement around the unrestriction point.

1

Figure 3.2: Lyapunov Exponents along the Dissociation Coordinate.

of the linear coupling matrix containing the PP and the IP exchange type amplitudes, and created a stable nonlinear system and solution by eliminating the off-diagonal matrix elements. The CC amplitude equations are easily solved in that diagonal representation, and the homotopy can be followed in very closely to the original problem. However the homotopy could only be followed in to a scaling parameter of at maximum around 80.0% for the IP level of correlation, and it is prohibitively expensive to follow the homotopy properly.

Another approach can be taken from our recent work [70]. A denominator shift reminiscent to amplitude regularization that has the form $-\gamma(e^{(t/t_c)^{2n}} - 1)$ can be constructed to affect the amplitudes similarly to a static shift. Unlike the static denominator shift the penalty function approach is flexible enough to be very small before the unrestriction point

on the dissociation PES where amplitudes are typically small, and large when amplitudes become large (typically on non-variational surfaces the amplitudes are greater than one). Of course to apply a gradient in the presence of such a penalty we must propagate this modification through the derivatives of the Lagrangian, but the resulting gradient "works" in the sense that we can use it to optimize orbitals and geometries.

The simple choice for the critical value of t parameter, t_c, is one, since that prevents a complete inversion of the reference and the doubly excited state. The other parameters (γ and n) should be chosen to balance making the corrections small at equilibrium, with ensuring that orbitals can be optimized towards dissociation. The shift we propose is $-8(e^{t^6}-1)$ E_h, and a corresponding shift of $-8(e^{t^6}(1+6t^6)-1)$ E_h for the Lagrange multipliers. There is only a cost in correlation energy of 9.1 μ E_h for ethene at its equilibrium geometry with the IP+DIP method, and a mean absolute error of 11.5 μ E_h for the PP active space G2a and G2b sets [158,159] done with 6-31G*. A choice of a γ of 2 and a power of 4 also works quite well, but it bears an energy cost around twice as large (both RMS and MAE) to our suggested parameters. Although the CC equations need to be penalized to be solved, this approach seems to be preferable as only modification to the amplitude and lagrange multipliers need to be made, and there is only a very slight cost in terms of correlation energy.

3.4 Heuristic Understanding of the Problem

There is a very simple argument for linear dependence in the basis composed of excitations from the orbitals of the unrestricted limit. For the usual UHF method at equilibrium one can imagine spin-polarization as a degree of freedom provided by the basis of excitations (and orbital rotations), and it is plain that in the unrestricted limit with only the spin-polarized orbitals excitations do not span this degree of freedom but the number of parameters is the same. One can also ask how this is manifest in the representation of the hamiltonian in this basis explicitly, and this is much less straightforwards. The results demonstrate that generally this singularity is present in the coupled cluster equations and strongly suggest that it lies in the linear part of the Jacobian. There is a possibility that both the linear and non-linear equations are singular but for different reasons, but we will assume that their conditions stem from the same problem.

The simplest possible case is the dissociation of a H_2 molecule in the minimal STO-3G basis. There is only one unique amplitude in the wavefunction, and one unique orbital unrestriction parameter which reflects the rotation of the beta bonding MO into the beta anti-bonding MO. The coupled cluster equations for the amplitude have the simple form of a quadratic equation, whose coefficients can be constructed for any distance, R and any

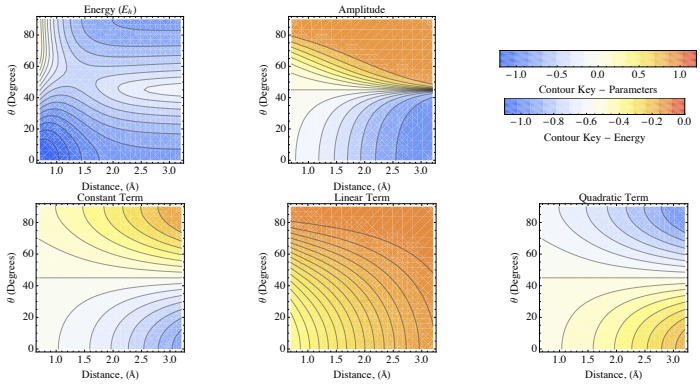


Figure 3.3: CCD parameters for H_2 in STO-3G basis, horizontal axis θ, vertical axis R (Å).

orbital rotation, θ, these surfaces are plotted in Figure X.

$$0 = A + B\hat{T} + C\hat{T}^2 \tag{3.3}$$

First notice on the plot of energy that variational determination of the orbitals would not unrestrict at any distance. The simple 1-amplitude CCD expansion can handle the open-shell singlet. Next focus on the line passing through 45^o, both the constant and quadratic terms vanish, but the linear term doesn't, so one of the two roots diverges in this unrestricted limit, while the other goes to zero. We can dissect the linear term further and find that there is a single diagram responsible for the non-vanishing term with the usual algebraic form:

$$\sum_{kc} < ka||ic > \hat{T}_{cb}^{kj} \Rightarrow B\hat{T}_2 \tag{3.4}$$

In the numbering of Figure X the 2-electron integral that appears in this term is $\langle 1\overline{1}^*||1\overline{1}^*\rangle$. The coulomb operator lies between orbitals lying on the same atom, and so it fails to decay as the bond is broken.

Considering a more general case, such as ethene dissociation, the CCD equations are now of much higher dimension, but we maintain that the essential features of this singularity are the same. The linear part of the equations is now multidimensional, and the coefficient of the linear term is a matrix which is precisely the well-known LCCD Jacobian [140], alternatively the CCD Jacobian for the null guess, also known as the matrix $\hat{U}$ in the notation of some other papers [138]. For complete CCD the dimension is too large to permit direct inspection, but we can examine truncated block of amplitudes (Table X) spanned by a local model for physical insight, remembering that they exhibit the same behavior seen in the full active-space CC models examined in section 2, and were the root

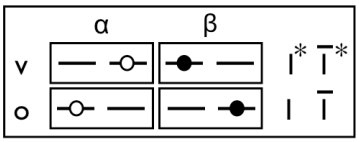


Figure 3.4: Orbital labeling for H_2 dissociation.

	$t_{1\bar{1}}^{1^*\bar{1}^*}$	$t_{1\bar{2}}^{2^*\bar{1}^*}$	$t_{2\bar{1}}^{1^*\bar{2}^*}$	$t_{2\bar{2}}^{2^*\bar{2}^*}$
$t_{1\bar{1}}^{1^*\bar{1}^*}$	0.093	-0.047	-0.047	0.000
$t_{1\bar{2}}^{2^*\bar{1}^*}$	-0.047	0.093	0.000	-0.047
$t_{2\bar{1}}^{1^*\bar{2}^*}$	-0.047	0.000	0.093	-0.047
$t_{2\bar{2}}^{2^*\bar{2}^*}$	0.000	-0.047	-0.047	0.093

Table 3.1: Linear coupling matrix for the PP and IP exchange amplitudes for ethene ($H_2C{=}CH_2$) at a C-C bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space in the 6-31G* basis.

of our interest in this problem. A figure labeling the relevant 1 particle functions is also included for clarity (Figure X).

Ironically, orbitals that lack any spin-symmetry impart a fragment symmetry onto this matrix which causes it's determinant to vanish. The off-diagonal matrix elements are found to arise in the same diagram examined above and the troublesome $\langle ov \| ov \rangle$ integral. We can provide further argument for the structure of this matrix for our local models because the number of amplitudes is manageable, and general formulas for the determinant given this structure, and the interested reader is referred to the appendix.

3.5 Conclusions

Recently an interesting study was directed at the question, "Do broken-symmetry references contain more physics than the symmetry adapted ones?" [160]. In that case the

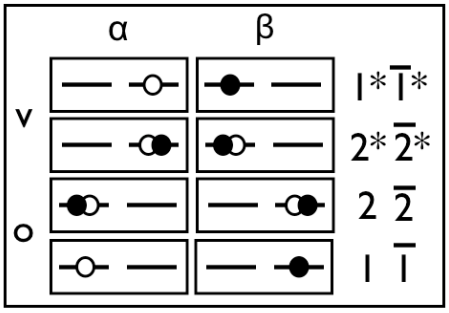


Figure 3.5: Orbital labeling for ethene dissociation.

concern was for RHF orbitals with broken spatial symmetries, and it was found that it was difficult for single-reference CC theories to correct the symmetry-broken wavefunction. In a similar vein, we have found that in the case of completely broken spin symmetry, the physics of spin correlation is removed from the cluster equations such that they are singular. This manifests itself in poor numerical condition of the cluster equations which prevents us from finding physical solutions and hampers orbital optimization. Ad-hoc manipulation of the Lagrangian using a penalty of the form $-\gamma(e^{(t/t_c)^{2n}} - 1)$ can render the equations soluble, even in this situation, and it seems that the remaining correlations are relatively unperturbed. Philosophically, our results suggest that CC based on UHF references succeeds (in the sense that most dissociation curves are reasonable even in the intermediate region) largely through the independence of strong spin correlations from others. This bodes well for correlation models designed on the principle of dividing these problems.

3.6 Appendix

3.6.1 Singularity in constrained cluster wavefunctions

In the arguments that follow, a pair correspondence between alpha and beta spin orbitals will be assumed by our notation. We enforce such a correspondence on our constrained CC models, however we note that by the pairing theorem [161] it is valid to apply such language to unrestricted wave-functions generally, as they may be cast into this form.

	$t_{1\overline{1}}^{1^*\overline{1}^*}$	$t_{1\overline{2}}^{2^*\overline{1}^*}$	$t_{2\overline{1}}^{1^*\overline{2}^*}$	$t_{2\overline{2}}^{2^*\overline{2}^*}$
$t_{1\overline{1}}^{1^*\overline{1}^*}$	0.102	-0.047	-0.056	0.000
$t_{1\overline{2}}^{2^*\overline{1}^*}$	-0.047	0.102	0.000	-0.056
$t_{2\overline{1}}^{1^*\overline{2}^*}$	-0.056	0.000	0.102	-0.047
$t_{2\overline{2}}^{2^*\overline{2}^*}$	0.000	-0.056	-0.047	0.102

Table 3.2: Linear coupling matrix for the PP and IP exchange amplitudes for fluoroethene (HFC=CH_2) at a C-C bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space.

The electronic Hamiltonian is separable into operators affecting each fragment separately $\hat{H} = \hat{H}_1 + \hat{H}_2$, and because of the high-spin structure of our reference any opposite-spin doubles amplitude doesn't alter the number of electrons on a fragment (only their spin), and spatial symmetry dictates that the spatial parts of the orbital in each pair should be the same on both fragments.

The Two Electron-Pair Case.

The situation of the 2-pair case of two-atom dissociation (an active space with 4 spatial, 8 spin orbitals) is similar physically to that of the 1-pair case. However, instead of two degenerate states that we can use the amplitudes to decide between, there is a larger degenerate subspace leading to far more zeroes in our Hamiltonian. The couplings between these degenerate states again leads to potentially infinite amplitudes, and the necessity of solving the CC equations iteratively leads to quite a problem here. A simple description of these degenerate doubly excited references can obtained with the pairing notion, as there are amplitudes that show coupling within a pair (PP), ie: $\hat{T}_{1\overline{1}}^{1^*\overline{1}^*}$, and opposite spin (OS) exchange-type amplitudes that couple the two pairs (IP), ie: $\hat{T}_{1\overline{2}}^{2\overline{1}^*}$. Because the determinants generated by all of these amplitudes only differ from the reference by flipping the spins of two of the spin-orbitals, the diagonal of the doubles-doubles block of the Hamiltonian is some value, a. Breaking down the off-diagonal matrix element coupling a PP amplitude to an IP amplitude, $\langle\Phi_{1\overline{1}}^{1^*\overline{1}^*}|\hat{H}|\Phi_{1\overline{2}}^{2^*\overline{1}^*}\rangle$, over the two fragments, we see that on one fragment the projections of these determinants are the same as generated on the diagonal of $\hat{H}$. Because the two determinants differ by 2 spin-orbitals the matrix element is a single integral, $\langle 1\overline{1}^*|2\overline{2}^*\rangle$ which may be identified with the integral examined in the 1-pair case. These matrix elements which share a single index may be denoted b. Matrix elements which share no indices, ie. PP of one pair with PP of the other, differ by more than two spin orbitals and are thus zero. The result is that the Hamiltonian has the following structure (in the basis: $\hat{T}_{1\overline{1}}^{1^*\overline{1}^*}$, $\hat{T}_{1\overline{2}}^{2^*\overline{1}^*}$, $\hat{T}_{2\overline{1}}^{1^*\overline{2}^*}$, $\hat{T}_{2\overline{2}}^{2^*\overline{2}^*}$):

	$t_{1\bar{1}}^{1^*\bar{1}^*}$	$t_{1\bar{2}}^{2^*\bar{1}^*}$	$t_{1\bar{3}}^{3^*\bar{1}^*}$	$t_{2\bar{1}}^{1^*\bar{2}^*}$	$t_{2\bar{2}}^{2^*\bar{2}^*}$	$t_{2\bar{3}}^{3^*\bar{2}^*}$	$t_{3\bar{1}}^{1^*\bar{3}^*}$	$t_{3\bar{2}}^{2^*\bar{3}^*}$	$t_{3\bar{3}}^{3^*\bar{3}^*}$
$t_{1\bar{1}}^{1^*\bar{1}^*}$	0.142	-0.036	-0.036	-0.036	0.000	0.000	-0.036	0.000	0.000
$t_{1\bar{2}}^{2^*\bar{1}^*}$	-0.036	0.142	-0.036	0.000	-0.036	0.000	0.000	-0.036	0.000
$t_{1\bar{3}}^{3^*\bar{1}^*}$	-0.036	-0.036	0.142	0.000	0.000	-0.036	0.000	0.000	-0.036
$t_{2\bar{1}}^{1^*\bar{2}^*}$	-0.036	0.000	0.000	0.142	-0.036	-0.036	-0.036	0.000	0.000
$t_{2\bar{2}}^{2^*\bar{2}^*}$	0.000	-0.036	0.000	-0.036	0.142	-0.036	0.000	-0.036	0.000
$t_{2\bar{3}}^{3^*\bar{2}^*}$	0.000	0.000	-0.036	-0.036	-0.036	0.142	0.000	0.000	-0.036
$t_{3\bar{1}}^{1^*\bar{3}^*}$	-0.036	0.000	0.000	-0.036	0.000	0.000	0.142	-0.036	-0.036
$t_{3\bar{2}}^{2^*\bar{3}^*}$	0.000	-0.036	0.000	0.000	-0.036	0.000	-0.036	0.142	-0.036
$t_{3\bar{3}}^{3^*\bar{3}^*}$	0.000	0.000	-0.036	0.000	0.000	-0.036	-0.036	-0.036	0.142

Table 3.3: Linear coupling matrix for the PP and IP exchange amplitudes for nitrogen (N_2) at a N-N bond length of 7.50 Åwith unrestricted PP orbitals in the minimal active space in the 6-31G* basis.

$$\hat{H} = (\hat{H}_1 + \hat{H}_2) = \begin{pmatrix} a & b & b & 0 \\ b & a & 0 & b \\ b & 0 & a & b \\ 0 & b & b & a \end{pmatrix} \tag{3.5}$$

This matrix is singular if $b = -a/2$ which is empirically found to be the case, and explained below. Of course this is only a sub-block of the linear coupling matrix, U, which would occur in complete CCD, and we affirm with calculations that in unrestricted CCD as it is usually practiced this is not a concern.

The Many-Pair Case.

In general, for a given PP amplitude there are three classes of possible OS exchange IP amplitudes: those which share no indices, those which excite from the same α orbital (we will call this one the fragment 1 corresponding IP amplitude) and those which excite from the same β orbital (fragment 2). The matrix elements of these combinations were examined above. If two IP amplitudes excite from a common index they will also have a nonzero matrix element b, and with these rules in hand one may construct $\hat{H}$. Consider a blocking of $\hat{H}_{IJ}$ such that a PP amplitude I and its corresponding fragment 1 IP amplitudes (indexed by J) are grouped together in a block which is ordered by their beta indices. In conventional spin-orbital notation this ordering is written: $\{\hat{T}_{i\bar{1}}^{1^*i^*}, \hat{T}_{i\bar{2}}^{2^*i^*}, \hat{T}_{i\bar{i}}^{i^*i^*}, ..., \hat{T}_{i\bar{n}}^{n^*i^*}\}$. Symbolically one may construct the diagonal block ($\underline{A}$) for the general case and see that it has the shape below, and determinant $(a-b)^{n-1}(a+(n-1)b)$. The off diagonal blocks ($\underline{B}$) are themselves diagonal, with value b and determinant b^n. The complete matrix has the

following block structure with this ordering, and the determinant of $\hat{H}$ for n pairs ($n > 2$) is $(a-2b)^{(n-1)^2}(a+(n-2)b)^{2(n-1)}(a+2(n-1)b)$. The matrix will be singular when $b = -a/2$, or if $(a + (n-2)b) = 0$ or $(a + 2(n-1)b) = 0$. In the full space, correlations of spin-paired basis functions often lift this linear dependence by coupling to this block, explaining the robust strength of unrestricted coupled cluster demonstrated in the literature.

$$\underline{A} = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & & \ddots & \vdots \\ b & b & \cdots & a \end{bmatrix} ; \underline{B} = \begin{bmatrix} b & 0 & \cdots & 0 \\ 0 & b & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & b \end{bmatrix} ; \hat{H} = \begin{bmatrix} \underline{A} & \underline{B} & \cdots & \underline{B} \\ \underline{B} & \underline{A} & \cdots & \underline{B} \\ \vdots & & \ddots & \vdots \\ \underline{B} & \underline{B} & \cdots & \underline{A} \end{bmatrix} \tag{3.6}$$

Chapter 4

The Perfect Quadruples Model

4.1 Introduction

The chemical community benefits greatly from systematically improvable single reference ab initio electronic structure methods that can be applied to molecules when Hartree-Fock (HF) is a valid approximation. Recent advances in computational technology have made methods such as coupled-cluster (CC) singles and doubles (CCSD) routinely applicable to dozens of heavy atoms [56, 57, 162, 163]. Some post-HF methods are becoming appreciably cheaper than HF itself [127, 164], and likewise the cost of hybrid DFT. When the mean-field HF reference doesn't provide an accurate basis for perturbation theory or low-rank coupled cluster techniques such as CCSD, the situation is quite different. The most commonly applied approach for coping with strong correlation is complete active space self-consistent field (CASSCF) [46, 165], but the growth of cost with the size of the correlated space is prohibitive and unphysical considering the dimension of the electronic Hamiltonian [166]. Moreover, CASSCF cannot provide chemically accurate energetics, and remains an exponential bottleneck behind dynamically-corrected treatments built upon it like CASPT2 [167] and multi-reference cluster theories [168–172]. There is a need for a tractable, black-box, approximation which can be used in lieu of CASSCF for large systems.

Many clever models have been proposed to overcome the CASSCF bottleneck for the problem of strongly correlated electrons. Broadly they can be divided into two groups : one that tries to overcome the complexity of the wavefunction by resorting to alternative descriptions of the system often based on the density matrices (density matrix renormalization group (DMRG) [173, 174], variational reduced density matrix (RDM) [175–177], etc.), and another group that tries to build the strong correlations into an accurate zero order wavefunction (valence bond [60,63,178,179], the related geminal models [180], renormalized cluster methods [116], spin-flip models [181–183], etc.). Despite a large body of work no single model has become attractive and reliable enough to consider the problem solved, and

the direct consequence is that quantitative studies of reactivity are usually only possible when a handful of important (strongly correlated) electrons can be modeled in a CASSCF calculation.

The importance of local correlation methods to the future of ab-initio many-electron theory cannot be exaggerated, because in virtually all cases of chemical interest we already have procedures which, if computable, would be exact. The greatest deficiencies in our understanding of the many electron problem are reflected in the way our physical expressions become unbearably complicated if only a handful of new bodies are added. To combat this problem local methods were introduced early in the history of quantum chemistry [122,184–186], and continue to reduce the cost of useful calculations [123,130,162,187]. For dynamical correlation problems like dispersion, the locality of the Coulomb interaction has been appreciated for a long time, but this information is rarely exploited in non-dynamical correlation methods. DMRG is a notable exception, but it is naturally best suited to exploit locality in one dimension.

For several years our group has developed a family of orbital-optimized coupled-cluster (OO-CC) [188] models for static correlation. Locality [125] is incorporated into these models naturally through the definition of chemically relevant electron pairs, via a pairing active space in which each electron pair is described by 2 orbitals, one nominally occupied or bonding, and the other nominally empty or antibonding. For a given electron pair, providing one double substitution amplitude for promotion of two electrons from the bonding to the antibonding level defines the perfect pairing model [60, 61, 63, 189], which is exact within the active space for a single electron pair, and is size-consistent. Providing additional amplitudes that simultaneously promote two electrons from two different pairs defines the imperfect pairing (IP) model [1, 65], which recovers a significant fraction of the interpair correlation energy. It is also possible to include the pair correlations that transfer electrons between 2 pairs [125] or couple three different pairs [190], which gives an accurate approximation to the complete treatment of all active space double substitutions. However, both PP and IP cluster amplitudes are truncated at double substitutions, usually rendering these methods unsuitable for more than two simultaneously strongly correlated electrons. This can lead to non-variational breakdowns with restricted orbitals. They can be partially addressed by modifying either the energy expression, as in quadratic coupled cluster doubles [71, 72], or the amplitude equations, as in the GVB-RCC [191] variant of IP. However, neither of these improved approaches can exactly solve the problem of even 4 strongly coupled electrons (e.g. a double bond breaking) in the pairing active space , because they still correlate only 2 electrons at a time.

Here we lift this restriction, by defining the truncated cluster model that exactly reproduces the wavefunction of two electron pairs coupled to each other. In a paired cluster framework this is possible through the addition of a single quadruple amplitude to couple each pair of PP amplitudes. Additionally one requires the triple substitutions that couple each pair of amplitudes. The idea is made concrete in the diagram below which enumerates the correlations contained in the resulting model, which we term Perfect Quadruples (PQ).

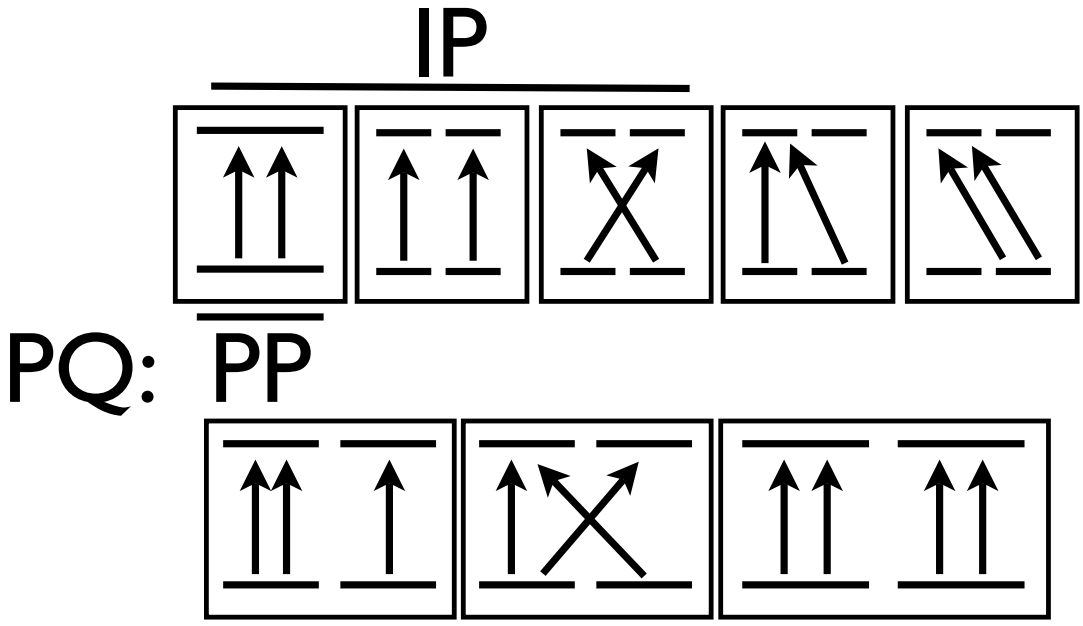


Figure 4.1: Correlations in the PQ approximation. Horizontal lines represent different orbital pair spaces, occupied below and virtual above.

The PQ model exactly treats the problem of 4 electrons in 4 orbitals (2 pairs) and is size-consistent. By restricting ourselves to include only electron correlations between two pairs with a cluster operator limited at quadruples, we obtain a model which is also balanced from a perspective of computational cost: PQ is a truncation of active space CCSDTQ which as a function of molecular size, M, retains only $O(M^2)$ of the $O(M^8)$ amplitudes of the parent theory. This model is realized and its performance is evaluated below.

4.2 The PQ Model

Coupled cluster theory has a storied history in quantum chemistry that has been well reviewed in recent literature [52, 110], so we will only make our ansatz clear and refer the reader for further information. The formulas below assume that the 1-particle basis has been divided into active and inactive subspaces in a well-defined way; the orbitals we name are exclusively active. Occupied orbitals are denoted $i, j, ...$ virtuals are denoted $a, b, ...$ the symbols $p, q, ...$ represent any sort of index. One begins by expressing the ket of the CC wavefunction (1) as the exponential of an excitation operator operating on the reference, which in this case will be the orbitals of a perfect pairing calculation. It is this exponential

that ensures a size-extensive treatment of correlation as is required for chemical systems. If a complementary de-excitation operator, Λ, is introduced then a set of equations that determine the amplitudes and the gradient may be derived by zeroing the partial derivatives of a quasi-energy Lagrangian, $\tilde{E}$. If the excitation operator $\hat{T}$ is truncated at some rank the expansion of the exponential contributing to an amplitude equation ends after a certain number of terms and at most quartic powers of an amplitude occur.

$$
\begin{aligned}
|\Psi_{\rm cc}> = e^T\big|\,0> \;\;;\hat{T}=\sum_n \hat{T}_n \;;\;\; ;\hat{\Lambda}=\sum_n \hat{\Lambda}_n \\
\hat{T}_n = \hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} = t^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} a^\dagger_{a1}\ldots a^\dagger_{a_n}\ldots a_{i_n}\ldots a_{i_1} \;;\; \hat{\Lambda}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} = \lambda^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} a_{i_n}\ldots a_{i_1}\ldots a^\dagger_{a_1}\ldots a^\dagger_{a_n} \\
\tilde{E} =< 0|(1+\hat{\Lambda})(\hat{H}e^T)_c|0> \;\;;\;\; \frac{\partial \tilde{E}}{\partial \Lambda}=0 \;;\;\; \frac{\partial \tilde{E}}{\partial T}=0
\end{aligned}
\tag{4.1}
$$

To couple four electrons exactly, amplitudes up to those in $\hat{T}_4$ must be included, defining the CCSDTQ method. The resulting triples and quadruples equations are summarized in (2) below. Complete symbolic expressions for the resulting matrix elements have been available in the literature for many years [96] [97], but their long derivation is equalled by the difficulty of programming these equations into a computer. For this reason we automated the diagrammatic derivation of second quantized expressions as have others [103, 109, 112].

$$
\begin{aligned}
0 =< \mu_3|H(\hat{T}_2+\hat{T}_3+\hat{T}_4+\hat{T}_1\hat{T}_2+\hat{T}_2\hat{T}_3+\hat{T}_1\hat{T}_3+\hat{T}_1\hat{T}_4+\frac{1}{2!}\hat{T}_2^2+\frac{1}{2!}\hat{T}_2^2\hat{T}_1 \\
+\frac{1}{2!}\hat{T}_1^2\hat{T}_2+\frac{1}{3!}\hat{T}_1^3\hat{T}_2+\frac{1}{2!}\hat{T}_1^2\hat{T}_3)|_c 0> \\
0 =< \mu_4|H(\hat{T}_3+\hat{T}_4+\hat{T}_2\hat{T}_3+\hat{T}_1\hat{T}_3+\hat{T}_1\hat{T}_4+\hat{T}_2\hat{T}_4+\hat{T}_1\hat{T}_2\hat{T}_3+\frac{1}{2!}\hat{T}_2^2+\frac{1}{2!}\hat{T}_3^2 \\
+\frac{1}{3!}\hat{T}_2^3+\frac{1}{3!}\hat{T}_3\hat{T}_1^3+\frac{1}{2!}\hat{T}_1^2\hat{T}_4+\frac{1}{2!}\hat{T}_1^2\hat{T}_3+\frac{1}{2!2!}\hat{T}_1^2\hat{T}_2^2+\frac{1}{2!}\hat{T}_1\hat{T}_2^2)|_c 0>
\end{aligned}
\tag{4.2}
$$

Computing the matrix elements of the second term of the quadruples equations without approximation takes $O(M^{10})$ computational operations, and $O(M^8)$ data storage as a function of molecular size, M. This rapidly becomes impractical for all but the smallest systems so we adopt a local ansatz based on electron pairs. Within a pairing active space, each pair is described by four orbitals, alpha occupied, alpha virtual, beta occupied, beta virtual, which in principle should be optimized to minimize the total energy. If this is done consistently then because orbital variations are nearly redundant with single substitutions, the singles amplitudes become very small. However, if one uses pairing orbitals optimized at a lower level of theory (for instance we shall employ PP orbitals in the tests reported

later of the PQ model), then single excitations within the active space can serve usefully as a surrogate for orbital optimization within the active space when the higher level wave function differs significantly from the one employed to optimize the orbitals. Therefore we retain single excitations.

Turning to pair-based truncation of CCSDTQ, at least two pairs must be used to construct a quadruples amplitude, and retaining the S,D,T,Q substitutions that couple only up to pairs of electron pairs defines a quadratic sparsity pattern that we call the Perfect Quadruples (PQ) approximation. This will be the fewest terms necessary to be exact for 4 electrons in the pairing active space. Symbolically this means that the spin orbital indices of an amplitude may originate in at most two distinct pairs as summarized as a diagram (Figure 1) and a definition below (3).

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \hat{T}_4 \text{ ; with } \hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} := s.t.\{a_k, i_k\} \subset \{\text{Pair}_1\} \times \{\text{Pair}_2\} \tag{4.3}$$

The amplitude equations are simply those of CCSDTQ [96, 97] when one neglects all amplitudes which belong to more than two pairs. If one neglects the triple and quadruple substitutions, a two pair approximation is very similar energetically to the Imperfect Pairing (IP) [1] [65] wavefunction previously developed in our group, but includes amplitudes that do not conserve the number of electrons in a pair [125]. These singly and doubly ionic amplitudes were omitted from the IP model to avoid the generation of some classes of integrals, but these are included here.

The coupled cluster energy depends on at most double substitutions, and so the triples and quadruples are necessary only insofar as they correct the doubles amplitudes. It was realized some years ago that the poor dissociation behavior of coupled cluster active space methods is rooted in their failure to obey a Pair Exclusion Principle [68]. Terms nonlinear in the amplitudes introduce excitations that recouple the spins of pairs to themselves and exaggerate the correlation energy. Deleting these terms as was done in the GVB-RCC model can repair the dissociation behavior of the model, but is deleterious near equilibrium. The quadruples fill this role, but one may imagine that they are turned-on naturally by the amplitude equations when important so equilibrium performance is conserved.

Finally, it is known that low-scaling correlation models based on localized orbitals can be affected by symmetry breaking artifacts. As the imposed sparsity on the amplitudes is relaxed, these artifacts are lifted by the inherent orbital-invariance of coupled cluster theory at the expense of steeper scaling of computational cost. This local truncation error is exemplified by the preference of benzene for D_{3h} rather than D_{6h} geometry at the PP and IP levels, a problem which can be almost entirely removed by additionally including double substitutions that couple 3 electron pairs at a time [126]. This local truncation error leads to a slight energetic preference for localized valence structures over delocalized ones, as in benzene. We mention the role of 3-pair couplings for the sake of completeness, but focus in this publication on the role of the new high-rank excitations. Any chemist

can recognize situations where two resonance structures are important to the electronic structure of a molecule, and these are the cases when the more costly three pair model should be employed.

4.3 Implementation

The lengthy coupled cluster equations truncated at quadruples were symbolically derived in closed form using a diagram-based symbolic manipulator developed by one of the authors (JAP) within the Mathematica package. The framework developed includes routines to automatically construct C++ codes for the amplitude equations that were integrated in a developmental version of the Q-Chem electronic structure package [192]. All contractions are performed as a series of pair-wise contractions with appropriately introduced intermediates so that the resulting cluster models scale with the familiar $O(M^{2s+2})$ rule, where s is the highest retained substitution (4 here). The very large (24^2) pre-factor associated with ignoring the antisymmetry of T_4 would make the PQ method impractical for all but the smallest systems, but the contraction algorithm developed exploits the permutational symmetries of all amplitudes and intermediates. The correctness of the coupled cluster implementation was verified by comparing with benchmark results for the CCSDT and CCSDTQ models. Local paired models like PP, IP, and the complete valence-optimized doubles were also re-implemented in the course of testing.

A method that preserves only a quadratic number of quadruples amplitudes should scale formally much more cheaply than the traditional CCD. To achieve that is a challenge because sparsity imposed on T_4, an eighth rank tensor, must be completely exploited. To meet this challenge we have developed a tensor library and contraction algorithm capable of performing these high rank contractions with the correct scaling. A structure is kept on the tensor such that any subset of dimensions may be fixed and the non-zero entries of the others may be iterated over without sorting or re-alignment. The effort of maintaining and operating on the sparse structure is spent entirely during element addition. Convergence of the amplitudes is accelerated with the DIIS algorithm [193].

The intermediates introduced for the nonlinear terms must also be generated with consideration for the sparsity imposed on the amplitude results. Because of the large number of intermediates this was done by simply limiting the number of pairs on the external dimensions of the intermediate destined for the result amplitude. This is sufficient to realize formal M^4 scaling with molecule size, which is the same as a local doubles model where ionic couplings are retained. Formally the slowest step of our algorithm for large systems is the production of the complete set of two-electron integrals in the active space. Table 2 collects the average number of floating point operations performed in contraction per iteration of the coupled cluster equations for some of the models considered here as imple-

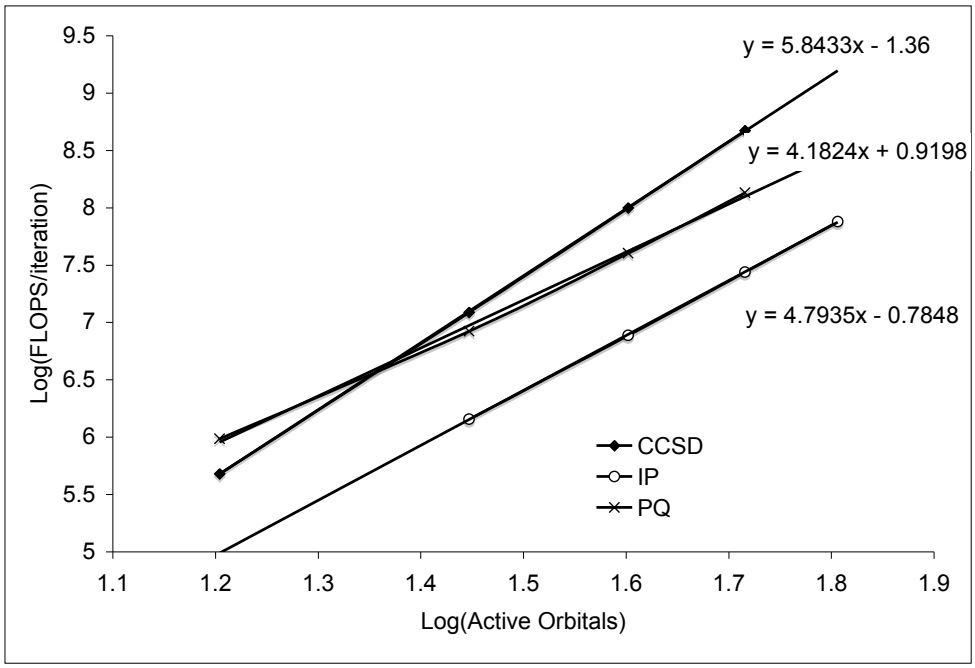


Figure 4.2: Scaling of methods in the 6-31G basis.

mented in this pilot code. The model system is a set of alkanes of increasing size in the geometry of an adamantane molecule in a 6-31G basis with full-valence PP active spaces (ie: methane, ethane, etc.). No spatial symmetry information was employed.

As a pilot implementation it is premature to judge by our program what the limits of a nearly optimal code would be. Even so, the pilot code is economical enough such that active spaces much larger than those usually approachable with CASSCF can be explored in modest amounts of time and memory. An optimal implementation would employ an integral approximation such as the resolution of the identity, or Cholesky decomposition. These approximations were recently applied to the CASSCF model [194], however in that case the exponential cost of solving for the amplitudes in the active space means that integral approximations cannot significantly extend the range of systems that can be explored. Because only a quadratic number of amplitudes are retained, the PQ model is well suited to take advantage of integral approximations, and this will most likely be done in future development.

4.4 Results

For the model systems considered below, the PQ approximation reproduces valence CASSCF within less than ≤ 10 kcal/mol of absolute error. Energies obtained parallel the CASSCF results over the entire surface within ≤ 5 kcal/mol. The local pair approximation and the limitation placed on the level of excitation both introduce limited errors into our model. Further improvement depends on disentangling and quantifying these errors for systems which are well understood. CASSCF calculations were performed with the aid of the GAMESS [195] package. Unless otherwise noted the cc-pVDZ basis and perfect pairing optimized orbitals were employed in the PP valence active space. All models except PP included single excitations.

4.4.1 H_4

The rectangular arrangement of four hydrogen atoms can exhibit strong static correlations [196], with complete configurational degeneracy at the square geometry. It is perhaps the simplest system for which the local doubles models PP, IP and their parent method CCD possess severe artifacts: an underestimated correlation energy and unphysical cusp at the square geometry. Note that in the figure CCD was performed without an active space, and so it includes dynamical correlation whereas the other methods do not. H_4 is also perhaps the simplest model one could construct for a polyradicaloid transition state. The PQ model exactly reproduces CASSCF (4,4) for this system by construction as is seen in Figure 3, while the PP and IP methods yield barriers that are too high by about 20 kcal/mol, using unrestricted orbitals.

4.4.2 Water

The simultaneous dissociation of water is a common performance benchmark for models of non-dynamical correlation [169]. If only double excitations are included from reference the restricted coupled cluster expansion becomes qualitatively incorrect away from equilibrium. In Figure 4 the PQ model is compared with CASSCF, and the same model without triples and quadruples (2P) (which is slightly better than the IP model). The corresponding energies are shown in Table 1. The cc-pVDZ basis was employed with the restricted orbitals of the PP wavefunction in the (8,8) active space. Near equilibrium, the doubles model and PQ both provide a reasonable approximation to CASSCF. Beginning at 2 $\mathring{A}$ the quality of the PQ approximation is made plain, as it continues to track CASSCF, while the 2P doubles curve falls to an incorrect asymptote. If the model were allowed to break spin symmetry, doubles models would capture the dissociation potential curve qualitatively, though in the intermediate regime they remain quantitatively inaccurate because they lack the configurations which complete the spin eigenfunction.

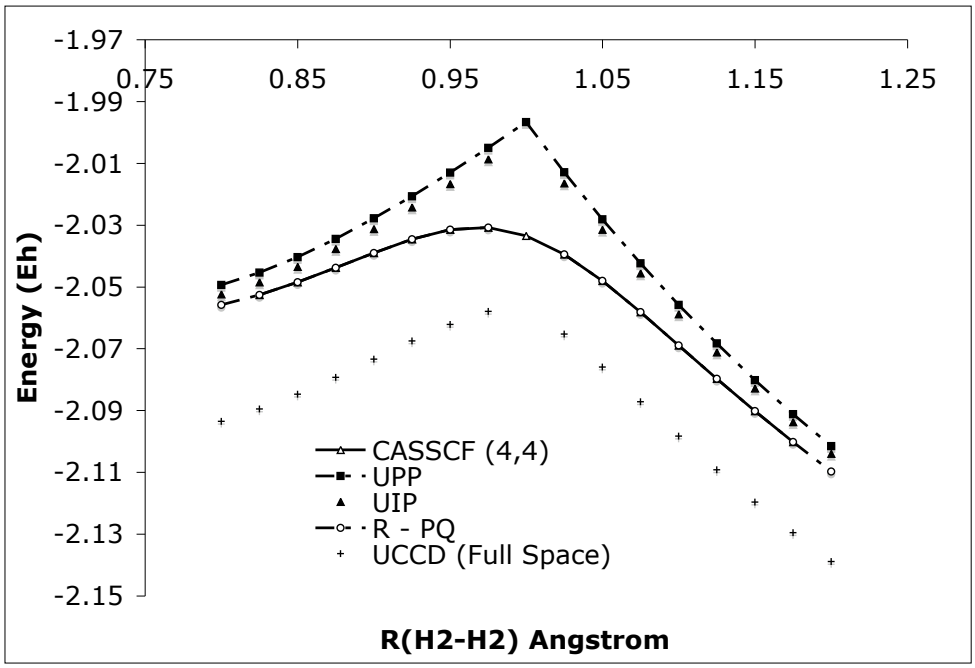


Figure 4.3: Potential energy curve for the rectangular dissociation of H_2—H_2 with the cc-pVDZ basis and restricted 2-pair doubles orbitals. (H-H distance 1 $\mathring{A}$)

To decompose the difference relative to CASSCF, the local models are compared in Table 2 against their non-local counterparts CCD, and CCSDTQ within the active space. We want to distinguish local truncation error (due to coupling only 2 pairs) from excitation level truncation at quadruples. Within the doubles space, 3 and 4-pair correlations absent in the 2-pair models (and PQ) account for 24 percent of the doubles correlation near equilibrium, which is roughly the same percentage of correlation energy that is missing from PQ vs. CCSDTQ at equilibrium even though the two pair locality constraint neglects a much larger fraction of triples and quadruples than doubles. At dissociation the correlations missing from the two pair models no longer contribute appreciably to the correlation energy and the local models both approximate their nonlocal counterparts with $\leq 1mE_h$ of error.

Insofar as correlation energy is additive with respect to the spaces of amplitudes included

R(O–H)$\mathring{A}$	CASSCF(8,8)	2p (8,8)	PQ (8,8)	CCD	CCSDTQ
1.1	-76.12104205	0.0087	0.0084	0.0056	0.0048
1.3	-76.04248739	0.0096	0.0091	0.0059	0.0045
1.5	-75.96406087	0.0085	0.0076	0.0056	0.0037
1.7	-75.90366501	0.0061	0.0053	0.0044	0.0023
1.9	-75.86333002	0.0042	0.0058	0.0028	0.0032
2.1	-75.83961132	-0.0046	0.0046	-0.0054	0.0028
2.3	-75.82726866	-0.0161	0.0040	-0.0167	0.0027
2.5	-75.82129839	-0.0259	0.0036	-0.0263	0.0027
2.7	-75.81841092	-0.0330	0.0033	-0.0334	0.0027
2.9	-75.81696701	-0.0380	0.0031	-0.0384	0.0027
3.1	-75.81622269	-0.0414	0.0030	-0.0418	0.0028
3.3	-75.81583025	-0.0438	0.0029	-0.0443	0.0028

Table 4.1: CASSCF energies for symmetric water dissociation (E_h), and relative errors of PQ, and non-local active-space models with the RPP orbitals.

(empirically this is often true), one might imagine decomposing the errors of PQ by their origin. In Figure 5 the quadruples truncation error (CASSCF - CCSDTQ), doubles locality error (Error(2P) - Error(CCD)), and triple/quadruple locality error (Error(PQ)-quadruples truncation error - doubles locality error) are plotted. The encouraging conclusion is that higher excitations in the valence space are often even more effectively captured by the local approximation than the doubles. This contrasts with the case of dynamical correlation where local models for higher excitations have been a more difficult challenge [197,198]. At 2.5 $\mathring{A}$ in the nonlocal quadruples the dimension of T_2, T_3 and T_4 are respectively 328, 1184 and 1810, and in the PQ model 100, 48 and 6.

4.4.3 Ethene

The dissociation of ethene and simultaneous dissociation of water are very similar static correlation problems, but in the former case the strongly correlated pairs are nearer to one another and the valence correlation space is three times larger. As shown in Figure 6, PQ reproduces CASSCF (12,12) faithfully using the RPP orbitals with an NPE of 9 mE_h and the resulting wave-function is spin-pure. Whereas IP falls to an incorrect over-correlated asymptote. If we allow IP to employ spin-contaminated orbitals it will meet the correct asymptote, but it under-correlates in the intermediate region as will be shown in the next section.

R(O–H)$\mathring{A}$	2p (8,8)	PQ (8,8)	CCD	CCSDTQ	Reference Energy
1.1	-0.1254	-0.1257	-0.1286	-0.1294	-75.98691672
1.3	-0.1470	-0.1476	-0.1508	-0.1521	-75.88583379
1.5	-0.1786	-0.1795	-0.1815	-0.1834	-75.77699934
1.7	-0.2208	-0.2217	-0.2226	-0.2246	-75.67668994
1.9	-0.2713	-0.2697	-0.2727	-0.2723	-75.58780651
2.1	-0.3339	-0.3246	-0.3347	-0.3265	-75.51033499
2.3	-0.3997	-0.3795	-0.4003	-0.3808	-75.44373647
2.5	-0.4598	-0.4303	-0.4602	-0.4313	-75.38736158
2.7	-0.5111	-0.4748	-0.5115	-0.4754	-75.34032152
2.9	-0.5534	-0.5123	-0.5538	-0.5126	-75.30158291
3.1	-0.5876	-0.5432	-0.5881	-0.5434	-75.27001032
3.3	-0.6152	-0.5685	-0.6157	-0.5686	-75.24442247

Table 4.2: Correlation Energies for symmetric water dissociation (E_h) with the RPP orbitals.

R(CH_2-CH_2)	CASSCF (12,12)	PQ Error	2P Error	CCSD Error
1.2	-78.145341	0.0110	0.0112	0.0053
1.4	-78.179446	0.0109	0.0112	0.0059
1.6	-78.142116	0.0108	0.0114	0.0065
1.8	-78.087149	0.0103	0.0115	0.0072
2	-78.035113	0.0096	0.0115	0.0079
2.2	-77.993032	0.0090	0.0109	0.0079
2.4	-77.962413	0.0086	0.0081	0.0056
2.6	-77.942196	0.0084	0.0022	0.0003
2.8	-77.929896	0.0083	-0.0049	-0.0065
3	-77.922803	0.0081	-0.0114	-0.0127
NPE		0.0029	0.0225	0.0130

Table 4.3: Correlation energies for dissociation of ethene(E_h) with restricted PP orbitals.

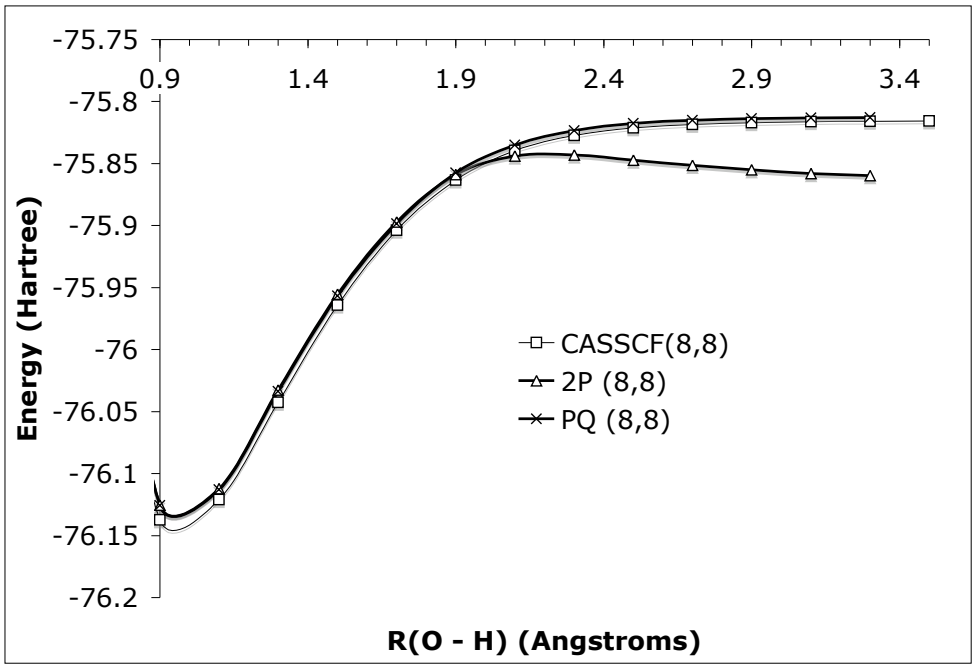


Figure 4.4: Potential energy curve for the symmetric dissociation of water with cc-pVDZ basis with RPP orbitals.

4.4.4 Nitrogen Molecule

The deceptively simple nitrogen molecule is a very rigorous test for models of electron correlation [199], and even very recently continues to be a benchmark for costly and sophisticated [200] models of static correlation [201]. Because it lacks hextuple excitations PQ "turns over" towards an over-correlated asymptote when restricted PP orbitals are employed. This can likely be ameliorated (without relaxing the spin symmetry constraint) by either including six-particle components of $\bar{H}$ or observing a "Pair Exclusion Principle" [191]; we hope to pursue these ideas in future papers, but for now, we employ unrestricted orbitals to avoid this issue.

With unrestricted orbitals, PQ should reproduce the entire CASSCF curve qualitatively, and the agreement should be quantitative at the dissociation limits. U-CCSDTQ's energies

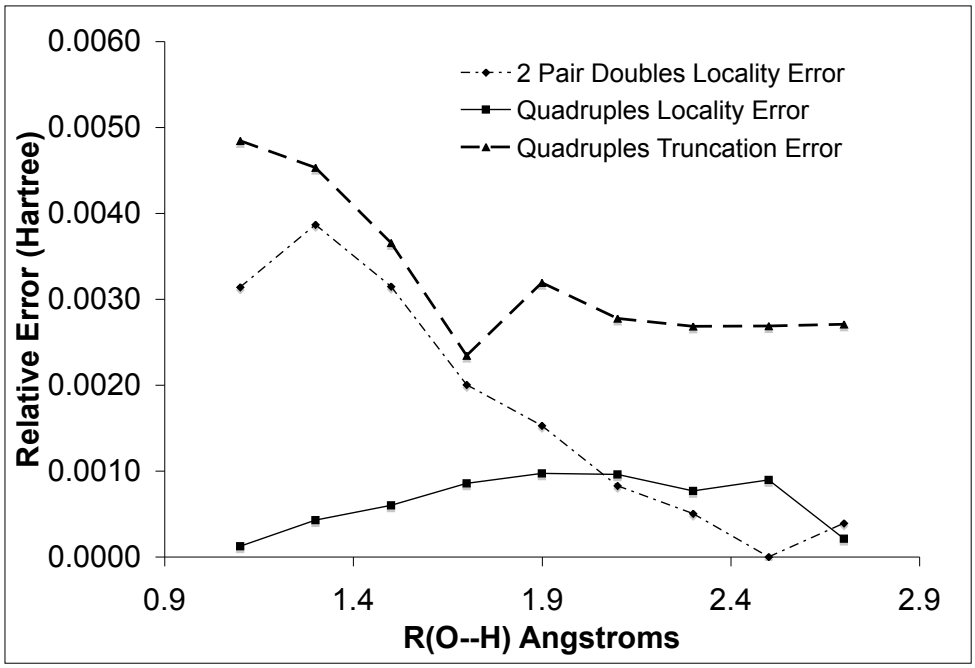


Figure 4.5: Error decomposed approximately by source for symmetric water dissociation with RPP orbitals.

are known to approximate full-CI very well for this system with only a few kcal/mol of NPE. In the highly correlated intermediate regime we may gauge the accuracy of PQ by comparing it against CASSCF. The switch to unrestricted orbitals complicates our assessment of the PQ model because the UPP's orbitals are more unrestricted (ie. unrestricted at shorter internuclear distances) than PQ's optimized orbitals would be. Single excitations assist in this capacity by relaxing the PP orbitals, and the resulting energies compare well with those of the restricted model before non-variational collapse.

Figure 7 and Table 4 compare PQ and the related doubles models with CASSCF in the valence PP active space. PQ outperforms the related doubles models, and maintains a very modest NPE error of only 3.83 kcal/mol (6.1 mEh). Predictably the underestimation of correlation energy is concentrated in the bond breaking region where the hextuples play an important role in configuration space, but unlike UPP this doesn't result in an unphysical barrier to dissociation. To put this result in perspective, an $O(M^6)$ scaling method based

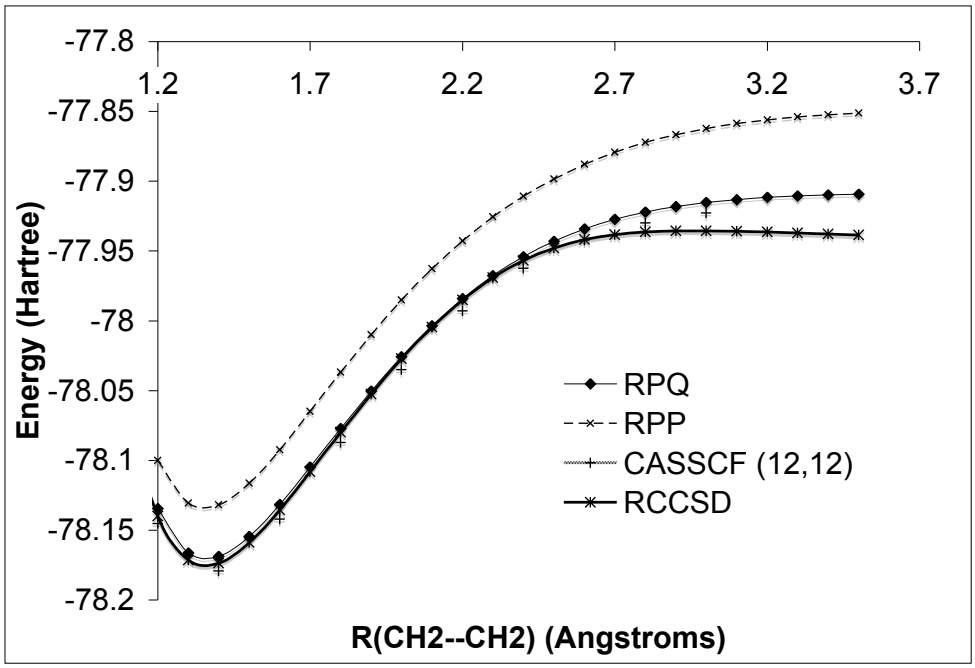


Figure 4.6: Electronic energy of ethene (R(C–H) = 1.07 $\AA$) in (12,12) space with cc-pVDZ basis

on variational (but approximately N-representable) reduced density matrices was recently applied to this same system in the same basis set and a (6,6) active space with an NPE of 18.9 mE_h [202]. At 1.8 $\AA$ in the nonlocal quadruples the dimension of T_2, T_3 and T_4 are respectively 825, 4360 and 8070, and in the PQ model 165, 80 and 10.

4.5 Discussion and Conclusions

We have introduced a model called Perfect Quadruples (PQ) for static electron correlation in a pairing active space. It's worthwhile to compare PQ with other efficient models of static correlation. The recent progress made in DMRG is impressive given the short history of the theory [203] [174], but chemical applications have all employed a decimation operation which is essentially one-dimensional. The even treatment of excitation regardless of rank is appealing, and the cost per sweep of the method is on the order of PQ. It will

R[N-N] (Å)	CASSCF (10, 10)	PP	IP	PQ
1.1	-109.12969	0.0876	0.0207	0.0077
1.2	-109.11145	0.0907	0.0216	0.0069
1.3	-109.06683	0.0935	0.0229	0.0068
1.4	-109.0148	0.0958	0.0253	0.0082
1.5	-108.96273	0.0950	0.0265	0.0089
1.6	-108.91571	0.0856	0.0252	0.0096
1.7	-108.87631	0.0682	0.0224	0.0102
1.8	-108.84984	0.0553	0.0217	0.0129
1.9	-108.82789	0.0412	0.0165	0.0104
2	-108.81368	0.0308	0.0125	0.0083

Table 4.4: N_2 Total energies (a.u.) and the error relative to CASSCF. (* includes singles) with unrestricted PP orbitals.

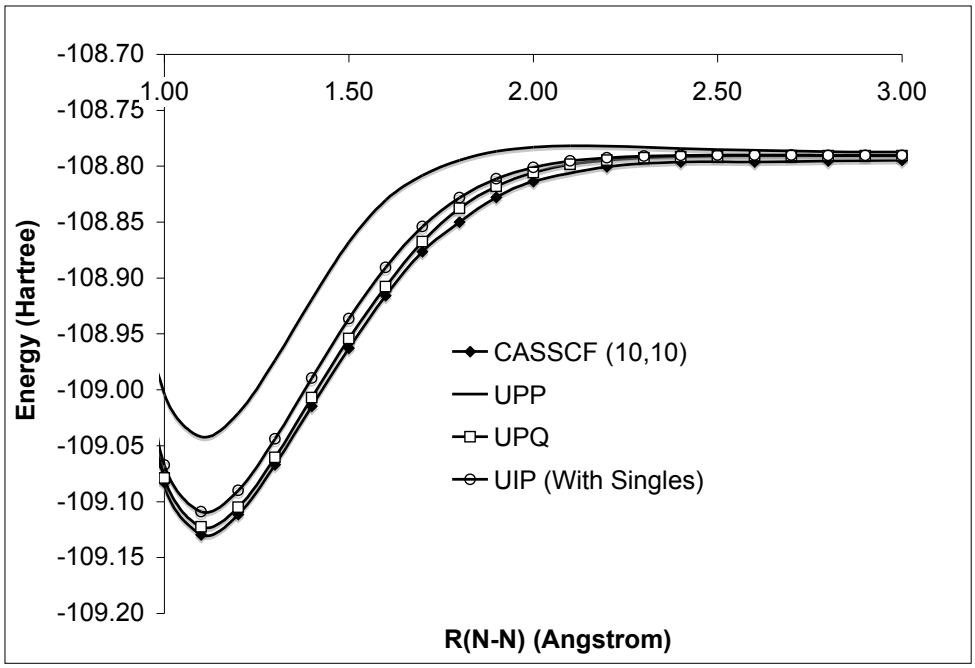


Figure 4.7: Unrestricted potential energy curve for the dissociation of N_2 in the cc-pVDZ basis.

be exciting to see with further numerical testing how this method performs as a reference for an efficient total energy calculation. Methods like canonical transformation [204] could provide a fair assessment of different references as models for the complete electronic energy.

Two modifications of valence bond theory [180,205] have been recently published which produce spin eigenfunctions remarkably with costs comparable to PQ. The valence-bond approaches are physically pleasing because they satisfy a property of the exact wave-function, but their correlation energies are much smaller than CASSCF near equilibrium. These methods represent a significant advance in the way we understand correlated spins. Spin-purity may be important for some applications, but likewise may not be most important for reproducing some sorts of data and one may imagine complementary niches for both VB and MO based methods.

There is also a model based on the spin-flip formalism [206], which describes quadruply radical systems and their excited states again with roughly the cost of the method just presented. The spin-flip approach and PQ both treat 4 electron problems well, the former with the strength that excited states are trivially obtained, and the drawback that it depends on an ROHF reference and doesn't treat the non-interacting ensembles of strongly correlated systems that occur in large molecules.

Of all these methods, PQ is distinguished as the one which most strongly resembles the CASSCF approach quantitatively and qualitatively. It is a size-extensive truncation of CASSCF within the active space, which is exact for two electron pairs in that space. This means that the perfect quadruples model is appropriate for molecules with tetraradical character and can describe such systems over their entire surface. At the same time, because correlations only connect pairs of electron pairs, the overall cost scales only as (M^4) with molecule size. This means all valence electrons can be taken as active without the CASSCF problem of trying to choose a very small (feasible) active space that is still reasonably accurate. The preliminary implementation and benchmark tests suggest that the PQ method is a useful step forward over existing lower rank truncations.

There are many interesting possible extensions, a number of which we hope to develop in the future. One conceptually obvious, though practically very challenging extension is to generalize the paired concept to still higher excitations (hextuple excitations coupling 3 different electron pairs would be the next member of the family defined by PP and PQ, and would be exact for 3 strongly coupled electron pairs). Considering appropriate dynamical correlation corrections is obviously important for practical applications requiring chemical or near-chemical accuracy. Algorithmic improvements that increase efficiency and permit direct orbital optimization are desirable.

Chapter 5

The Perfect Hextuples Model

5.1 Introduction

Quantum chemists distinguish two sorts of correlation, one which is captured easily by perturbation theory called dynamic, and another called "static", or "non-dynamical" which afflicts molecules whose bonds are nearly broken [168, 196], or which are otherwise poorly described by methods that depend strongly on a mean-field reference. Quantitative ab-initio studies of chemical reactivity are most straightforwardly possible when a computational method capable of tackling static correlation can be applied to the entire system of interest. However the standard approach to static correlation, the complete active space self-consistent field (CASSCF) method [46, 207, 208], is exponentially costly as a function of the number of active electrons. As a result, a number of fundamental reactions of interest are exceedingly difficult to address with CASSCF, such as the Cope rearrangement [209–212], due to the need to choose a truncated active space. A wide variety of on-going work is directed at developing more feasible methods for treating strongly correlated molecular problems [116, 174, 175, 177, 178, 180, 199, 202, 213–219].

It is desirable to have a systematically improvable hierarchy of approximations for static correlation whose cost increases with accuracy and system size in as tractable a way as possible, to eliminate CASSCF's drastic limitations on the number of active orbitals. In other words, we need active space analogs of the large body of work directed at developing approximations to the Schrodinger equation, as exemplified by the coupled cluster (CC) hierarchy of models: CCSD (truncation at double substitutions from the reference), CCSDT (triples), CCSDTQ etc [52, 110, 220, 221]. One possibility is a hierarchy of valence active space orbital-optimized cluster theories (VOO-CC). VOO-CC methods were first proposed some years ago [188], and implemented at the CCD level, and later extended to the quadratic CCD model [72]. VOO-CCD is exact for isolated pairs of electrons in the active space, and is extensive. Higher VOO-CC methods (VOO-CCDT, VOO-CCDTQ etc) would constitute a systematically improvable hierarchy of approximations to CASSCF,

as we desire. However, direct implementation of higher VOO-CC methods (VOO-CCDT, VOO-CCDTQ etc) seems unpromising because their scaling with the number of active electrons mirrors that of conventional CC theory, and it is possible that even higher excitations may be required for strongly correlated problems.

An alternative approach that appears much more promising replaces the VOO-CCD starting point with one of the simplest models of strong correlations, perfect pairing (PP). The PP model [60,61,179] describes the j^{th} electron pair by two configurations, made from two orbitals, one bonding, ϕ_j and one antibonding, ${\phi_j}^*$, and a single amplitude, t_j:

$$G_j = |\phi_j \bar{\phi}_j\rangle + t_j |\phi_j^* \bar{\phi}_j^*\rangle \tag{5.1}$$

For a system of multiple electron pairs, the intermediate-normalized PP wave function is then an antisymmetrized product of the geminal corresponding to each pair, G_j, which is alternatively a local active space variant of CCD [63] where only one-pair amplitudes are retained:

$$|\Psi_P P\rangle = \exp(\hat{T}_{PP})|\Phi_0\rangle \tag{5.2}$$

Here $|\Phi_0\rangle$ is the reference with all bonding orbitals occupied, and,

$$\hat{T}_{PP}|\Phi_0\rangle = \sum_j t_j |\Phi_{j\bar{j}}^{j^*\bar{j}^*}\rangle \tag{5.3}$$

where $|\Phi_{j\bar{j}}^{j^*\bar{j}^*}\rangle$ is a doubly substituted determinant where the j^{th} bonding level has been replaced by the antibonding level, reflecting the second configuration of Eq. (5.1). The PP energy is minimized with respect to variations in the amplitudes and the orbitals. The number of non-redundant orbital degrees of freedom is increased relative to VOO-CCD, because rotations of the active occupied (or virtual) orbitals amongst themselves also affect the PP energy. Within the active space, PP is exact for isolated electron pairs, and extensive, just like VOO-CCD. And, as it is a cluster operator, it can be improved. A variety of augmentations that include the most important inter-pair correlations at the level of double excitations have been suggested [65, 190], which approach the accuracy of VOO-CCD in the pairing active space at much lower computational cost [222].

How should the PP starting point (exact for a single pair of electrons and extensive) best be extended to more strongly correlated systems in the CC framework? One very promising approach is to seek exactness for larger numbers of electrons at minimum cost. The next model should achieve exactness for two pairs (4 electrons), meaning that the cluster operator must be truncated only at quadruples (i.e. a version of VOO-CCDTQ). To keep the complexity as low as possible subject to this goal, we chose keep only those active space amplitudes that couple either 1 or 2 electron pairs together. We have recently proposed and implemented this idea, which we call the perfect quadruples model (PQ) [86]. PQ is an enormous simplification over VOO-CCDTQ which can couple up to 8 different

pairs together. As a function of the number of valence electrons, o, VOO-CCDTQ involves $O(o^8)$ amplitudes and $O(o^{10})$ computation, while PQ retains only $O(o^2)$ amplitudes with $O(o^4)$ computation. Numerical tests were generally very encouraging – for instance the PQ absolute error relative to CASSCF(10,10) for N_2 dissociation was less than 0.33 eV, while the non-parallelality error was less than 0.2 eV. Of course this performance must degrade for problems where the number of strongly correlated electrons increases, which makes extensions of PQ desirable.

In this work, we therefore implement and explore the next member of this systematically improvable pair-based hierarchy of approximations to CASSCF. The next level should be exact for isolated systems of 3 electron pairs (6 active electrons), which makes it a subset of the VOO-CCDTQ56 method. To preserve exactness for 3 pairs, we choose to retain only those VOO-CCDTQ56 amplitudes that involve 3 electron pairs or fewer. We will refer to this truncation of VOO-CCDTQ56 as the Perfect Hextuples (PH) model. PH will reduce the $O(o^{12})$ amplitudes and $O(o^{14})$ computation of VOO-CCDTQ56 to only $O(o^3)$ amplitudes with $O(o^6)$ computation. Indeed, as will be discussed later, the computation can be further reduced to $O(o^5)$ with a further approximation. PP, PQ and PH together comprise the lowest 3 levels of a hierarchy of pair approximations which cost $O(o^{2p})$ before further approximations, where p is the number of electron pairs for which we demand exact agreement with CASSCF. Each step of the hierarchy improves upon the previous one in two respects: first, by inclusion of two additional levels of substitution, and second, by the inclusion of correlations that entangle one additional electron pair at the existing lower level of substitutions.

The restriction on the number of electron pairs that are coupled in the correlation amplitudes is essentially a local correlation approximation [122] that is inherent in this hierarchy. The strongest correlations entangle one or relatively few pairs, and balanced truncation by excitation level ($2p$) and number of coupled pairs (p) enables the cost to increase so much more gradually than truncation by excitation level only. Therefore improved results at each new level reflect error reduction from one or both of these two aspects of truncation. It is also possible to separate the two types of truncation, by separately choosing the maximum level, n, of substitutions and the maximum number of different pairs, p whose orbitals are allowed to comprise such an index, to define an (n, p) pairing approximation. We shall occasionally make use of this additional flexibility. For instance, symmetry breaking in benzene in PP is known to be a result of neglecting 3-pair double substitutions, rather than neglect of higher substitutions [190].

CASSCF itself requires a companion treatment of the neglected dynamical correlation, such as by multireference perturbation theory [64, 167, 223, 224], in order to approach quantitative accuracy. So too will any approximation in the PP, PQ, PH hierarchy, and the reader may be concerned that it will be more difficult to construct such a model for PH. Even with only a density matrix there are now very promising methods which can provide dynamical correlation [204, 225–229]. There are further options for PH because we possess a good 0^{th}-order wavefunction and a well-defined partitioning of the Hamiltonian. Our

group has developed perturbation theories for active-space cluster models based on Löwdin partitioning [73], which have been applied to the PP starting point [230]. It is also possible to reformulate the state-universal multi-reference cluster theory (MRCC) [201] for incomplete active spaces. It is especially appealing to adapt the state-specific single-reference MRCC [171] formalism for use with the PP, PQ, PH hierarchy since it produces a unified model which amounts to an alternative truncation of the CC hierarchy which accounts for total correlation, and we will soon publish results along this line.

5.2 The perfect hextuples model

5.2.1 Overview of VOO-CC theory

Our goal is to employ coupled cluster (CC) theory to approximate CASSCF within a perfect pairing active space, optimizing both correlation amplitudes and orbitals [188]. We begin with the usual CC ansatz which parameterizes a many-electron state, $|\Psi\rangle$, as the exponential of a correlation operator, $\hat{T}$, acting on a reference determinant, $|0\rangle$ (written as $|\Phi_0\rangle$ in the Introduction): $|\Psi\rangle = e^{\hat{T}}|0\rangle$. This correlation operator is restricted to an active space of orbitals, generally one occupied (in the reference) for each valence electron totaling o and the same number of unoccupied single-particle states. We denote active occupied orbitals as i_k, and active virtual orbitals a_k.

Following usual CC theory [52, 110], the operator $\hat{T}$ is chosen to include every possible orbital substitution, $\hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n}$,

$$\hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} = t^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n} \hat{a}^\dagger_{a_1} \ldots \hat{a}^\dagger_{a_n} \hat{a}_{i_1} \ldots \hat{a}_{i_n} \tag{5.4}$$

up to some maximum level of substitution, n. These sets of all possible $\hat{T}$ for substitutions of a given level n, are called $\hat{T}_n$. The equations that determine the amplitudes corresponding to each retained active space orbital substitution, t_s are obtained by projecting the Schrödinger equation with the corresponding substituted determinant, $\langle\mu_s| = \langle 0| \left(\hat{a}^\dagger_{a_1} \ldots \hat{a}^\dagger_{a_s} \hat{a}_{i_1} \ldots \hat{a}_{i_s}\right)^\dagger$:

$$\langle\mu_s|e^{-\hat{T}} \hat{H} e^{\hat{T}}|0\rangle = \langle\mu_s|\{\hat{H}e^{\hat{T}}\}_c|0\rangle = 0 \tag{5.5}$$

The VOO-CC energy that approximates CASSCF follows from projecting with the reference:

$$E = \langle 0|e^{-\hat{T}} \hat{H} e^{\hat{T}}|0\rangle = \langle 0|\{\hat{H}e^{\hat{T}}\}_c|0\rangle \tag{5.6}$$

In order to have an energy that can be varied with respect to orbital rotations, it is convenient to define a pseudo-energy, $\tilde{E}$ that augments the VOO-CC energy, Eq.(6), with Lagrange multipliers, λ_s multiplying each CC amplitude equation, Eq.(5). Defining a de-excitation operator $\hat{\Lambda}$ analogous to the excitation operator $\hat{T}$, the pseudo-energy can be

written as:

$$\tilde{E} = \langle 0|(1+\hat{\Lambda})e^{\hat{-T}}\hat{H}e^{\hat{T}}|0\rangle = \langle 0|(1+\hat{\Lambda})\{\hat{H}e^{\hat{T}}\}_c|0\rangle \tag{5.7}$$

The non-linear equations which determine the amplitudes, energy and gradient can be derived by making the pseudo-energy stationary with respect to variations of $\hat{T}$ and $\hat{\Lambda}$, and finally with respect to orbital variations, θ:

$$\frac{\partial\tilde{E}}{\partial\hat{T}} = 0 \; ; \; \frac{\partial\tilde{E}}{\partial\hat{\Lambda}} = 0 \; ; \; \frac{\partial\tilde{E}}{\partial\theta} = 0 \tag{5.8}$$

These $\hat{\theta} = \theta_p^q(a_p^q - a_q^p)$ parametrize a unitary transformation $U = e^{\hat{\theta}}$, mapping the guess orbitals to the optimal orbitals $\hat{H} \rightarrow U^\dagger\hat{H}U$. In the traditional OO-CC orbital variations include occupied-virtual mixings (θ_i^a), as well as active-inactive mixings in both the occupied and virtual spaces. Equation 8 has some appealing formal consequences: active space optimization is possible; the response theory of the model does not exhibit spurious poles, and some properties are rendered Gauge invariant [231]. However, it should be noted that if single excitations are neglected [232] Equation 8 does not exactly recover recover the CASSCF energy. Including the singles would require some improved algorithms to obtain orbital convergence in a reliable and practical manner, and as of yet this problem remains unsolved, but can be side-stepped if the orbital gradient is modified. We discuss this further below, and address this technicality such that the model reproduces CASSCF for 3 electron pairs.

With the truncation of $\hat{T}$ at $\hat{T}_n$, Eq. (5.8) must be solved iteratively with $O(o^{2n+2})$ operations per iteration, as follows from contractions associated with the $O(o^{2n})$ amplitudes. A quantitative picture of an n-electron singlet dissociation problem often requires some amplitudes in the n^{th} rank cluster approximation, which means $O(o^{2n+2})$ effort. For chemical systems undergoing 6-electron processes (e.g. the Cope rearrangement) this corresponds to $O(o^{14})$ computation. This effort, while far better than exponential, is scarcely affordable for anything beyond toy calculations at present. We must therefore follow an alternative path that does not simply truncate the VOO-CC hierarchy by substitution level alone.

5.2.2 Definition of the model

As already mentioned in the Introduction, instead of truncating $\hat{T}$ by substitution level alone, in the PP, PQ, PH hierarchy, amplitudes are chosen so that a certain number of electron pairs, p, can be treated exactly, while amplitudes that couple more than p pairs are discarded. With this simultaneous truncation by the maximum excitation level, $2p$ for the p-pair model, and by the number of pairs coupled (p), the p-pair dissociation problem can be solved with vastly lower cost than simply truncating $\hat{T}$ at $\hat{T}_{2p}$.

The PP model was already reviewed as motivation in the Introduction, and is a subset [63] of VOO-CCD [188]. VOO-CCD uses $\hat{T}_2$ as the cluster operator, which contains all

$O(o^4)$ double substitutions, $\hat{T}^{a_1 a_2}_{i_1 i_2}$. PP replaces $\hat{T}_2$ by $\hat{T}_{PP}$, Eq. (5.3), which contains only $O(o)$ amplitudes, t_j. Each amplitude belongs to a pair of two electrons, labelled as j, and described by 2 orbitals, as in Eq. (5.1). Therefore all 4 indices in each double substitution refer to only 1 particular pair: $\hat{T}^{j^* \bar{j}^*}_{j\bar{j}}$. This association between orbitals and pairs is made well-defined (and informative) when the energy $\tilde{E}_{\rm PP}$ is minimized by varying the orbitals to find bonding and anti-bonding levels that best describe pairs of electrons. The resulting orbitals often localize in physically meaningful ways, as is mathematically required in order to exactly describe the behavior of truly isolated electron pairs. From that PP starting point, we recently then defined and implemented the PQ model [86], which is exact in the active space for isolated 4-electron systems (2 pairs). PQ is thereby a subset of VOO-CCDTQ, where amplitudes and integrals are only retained if they couple ≤ 2 different pair indexes (in contrast to PP which couples only one).

We turn now to the next level of the hierarchy, the PH model, which is designed to be exact for isolated systems of 3 electron pairs (6 electrons), as well as properly extensive. To accomplish this objective, the equations of the PH model are defined to be those of the valence optimized orbital coupled cluster theory truncated at hextuples (ie. VOO-CCDTQ56), where amplitudes and integrals are allowed to possess indices originating in at most three different pairs, as indicated below:

$$\hat{T} = \sum_{n \leq 6} \hat{T}_n \text{ ; with } \hat{T}^{a_1, a_2, \ldots, a_n}_{i_1, i_2, \ldots, i_n} \text{s.t.} \{a_k, i_k\} \subset \{\text{Pair}_1\} \times \{\text{Pair}_2\} \times \{\text{Pair}_3\} \tag{5.9}$$

There are an enormous variety of amplitudes that satisfy the condition of Eq. (5.9), beginning with every PP amplitude and every additional PQ amplitude (since they couple only 1 and 2 pairs respectively).

Additionally the D, T and Q substitutions that couple 3 different pairs are included in PH. For example, if there are 4 pairs in the active space (a CASSCF (8,8) space), PH would include doubles that might be denoted as $\hat{T}^{3\alpha 3\beta}_{1\alpha 2\beta}$, which are omitted in PQ. Their inclusion, without increasing the level of substitution, serves to reduce the truncation error associated with only including amplitudes with 2 distinct pair indexes in the PQ model. There is still a remaining pair truncation error at the D, T, and Q levels of substitution. In the same 4 pair example mentioned above, the double substitution, $\hat{T}^{3\alpha 4\beta}_{1\alpha 2\beta}$, does not satisfy Eq. (5.9) and is omitted from PH. Similar considerations apply to the T and Q substitutions. In addition, pentuple and hextuple substitutions that couple 3 pairs are included in the PH model, which permits exactness for a system of 3 electron pairs. Regardless of the level of substitution, all amplitudes retained in the PH model can be labeled with indices that grow cubically with the size of the system, giving $O(o^3)$ amplitudes.

PH is a much more severe approximation to VOO-CCDTQ56 than PQ is to VOO-CCDTQ. There are $O(o^8)$ quadruples in the nonlocal VOO-CCDTQ model, of which only $O(o^2)$ are retained in PQ – a fraction that goes as $O(o^{-6})$. For hextuples in VOO-CCDTQ56 and PH, these figures change to $O(o^{12})$ and $O(o^3)$, meaning the fraction of amplitudes re-

tained has diminished from $O(o^{-6})$ in PQ to $O(o^{-9})$ in PH. This is quantified in the next section, but it should be clear that simultaneous truncation by maximum substitution level and by number of pairs coupled permits levels of substitution that would otherwise be possible only for toy systems.

Given the definition, Eq. (5.9), of the substitutions to be retained in the PH truncation, the remainder of the model is specified at least implicitly by the VOO-CC machinery summarized in the previous sub-section. There is one retained equation for each retained amplitude, of the form of Eq. (5.5). There is one Lagrange multiplier for each retained equation, enabling the construction of the pseudo-energy, Eq. (5.7), in terms of the retained equations. The pseudo-energy of the PH model must be made stationary with respect to all energetically signficant degrees of freedom, following Eq. (5.8). We replace the occupied-virtual orbital rotation condition with the Bruckner condition to ensure exactness (see below). While this prescription is clear and well-defined in outline, the PH model is none-the-less enormously algebraically complex – even the CCSDTQPH equations do not (to our knowledge) appear explicitly in the literature, let alone the VOO-CCDTQ56 equations, orbital variations and all! Therefore we turn next to the challenge of implementing the PH model for practical calculations.

5.2.3 Single excitations, orbital-optimization and exactness

We claim exactness for 3 electron pairs but as realized by others [232], and mentioned in the PQ paper, OO-CC [233] is not generally exact without singles, except in situations where the odd-particle numbered amplitudes are zero because of a symmetry. Without them the Schrödinger equation has not been projected against singles and expanded completely over the given many-body basis, although there is very little correlation physics in that block of $e^{-T}\hat{H}e^{T}$. The error caused by this approximation is small relative to the local approximation (tens of μE_h), and usually of opposite sign, so we have avoided treating this technicality until now.

To satisfy the SE exactly within the active space, our model must satisfy the SE projected against singles, the Bruckner [234] condition:

$$\langle 0|\mu_1\{\hat{H}e^{\hat{T}}\}_c|0\rangle = 0 \tag{5.10}$$

To ensure this without introducing $\hat{T}_1$, we provide a hybrid variational/Bruckner [234–236] orbital gradient as first suggested by Olsen and Kohn [232]. At each orbital iteration we calculate the few terms of the singles residual which do not depend on $\hat{T}_1$:

$$\hat{F}^{\mathrm{p01}}_{\mathrm{h02}} + \hat{F}^{\mathrm{p3}}_{\mathrm{h1}}\hat{T}^{\mathrm{h02h1}}_{\mathrm{p01p3}} + \frac{1}{2}\hat{T}^{\mathrm{h02h2}}_{\mathrm{p4p5}}\hat{V}^{\mathrm{p4p5}}_{\mathrm{h2p01}} + \frac{1}{2}\hat{T}^{\mathrm{h2h3}}_{\mathrm{p01p4}}\hat{V}^{\mathrm{h02p4}}_{\mathrm{h2h3}} + \frac{1}{4}\hat{T}^{\mathrm{h02h1h2}}_{\mathrm{p01p4p5}}\hat{V}^{\mathrm{p4p5}}_{\mathrm{h1h2}} \rightarrow R_1 \tag{5.11}$$

Model	Variational	Hybrid
VOD	0.000737	0.000740
PQ	0.001467	0.001468
PH	0.000060	0.000060

Table 5.1: Energetic effects of the hybrid gradient with several OO-CC models of water with the 6-31g** basis. Energies reported are Model - CASSCF(8,8) in E_h. R(O-H) = 1Å, and $\angle(HOH) = 103.1°$.

The active-occupied → active-virtual block of this matrix is divided by usual eigenvalue denominator, multiplied by negative one, and inserted into the usual variational expression for $\delta\tilde{E}_c/\delta U_p^q$ as a gradient descent step. $\delta\tilde{E}_c/\delta U_p^q$ is used with the current θ to form $\delta\tilde{E}_c/\delta\theta_p^q$ and extrapolated with DIIS alongside the amplitudes [237]. Given that the cluster model's energy is invariant to the occupied-occupied and virtual-virtual orbital rotations (and PH is invariant to these rotations in the limit of 3 electron pairs) this recipe is now exact.

For our purposes this hybrid doesn't have any computational cost drawbacks and is algebraically much simpler than evaluating the Λ_1 equation. However in any useful application the singles omission error which exaggerates the correlation energy is much less significant than the error induced by three-pair correlation which diminishes E_c. Quantitatively, considering N_2 in a 6-31G basis at 1.1Å, PH is above the CASSCF (10,10) energy by $323\mu E_h$ with the variational gradient; application of the hybrid gradient increases that number to $328\mu E_h$. Likewise for the case of H_2O (Table 5.1) the hybrid gradient slightly increases the energy at a scale which isn't meaningful for our method as it is intended. Our approximate Hessian for Eq. (5.10) is of lower quality than the variational gradient, and so in the results of this paper we always employ the latter although the code exists for the former.

5.2.4 Implementation

Coupled cluster theory becomes algebraically cumbersome as the level of substitution is increased. The level of complexity of the working equations at a given level of substitution is further increased (even as the scaling of the actual computational complexity is reduced!) if a sparsity pattern is imposed, such as neglect of all amplitudes and integrals coupling more than 3 pairs in PH. Never-the-less, the theory which produces these equations is well-known, and the rules that lead to the governing algebra have been automated by several groups [103, 109, 112, 238], to permit high level benchmark CC calculations. We have developed our own automated code generation tools for the purposes of this work, which include the additional feature essential for PH of exploiting an imposed sparsity. It has been discussed in detail in a separate publication [239]. For completeness, the resulting equations specifying our implementation of the PH model (i.e. the particular factorizations

chosen) are given in the supporting information.

It is as important to demonstrate the tractability of the PH model as its accuracy, so we will describe how our approximations lead to affordable scaling. At convergence $\partial \tilde{E}_c / \partial \hat{\Lambda}_6 = 0$, but on intermediate iterations we call this quantity the hextuples residual, R_6. It is composed of 135 separate terms, which are up to quartic in the operator $\hat{T}$. In the non-local VOO-CCDTQ56 model, computation of this quantity dominates the CPU cost in the $\hat{T}$ iterations. Consider one particular term of the hextuples residual for illustration:

$$\hat{T}^{a_1,b_1}_{i_1,i_2}(\hat{T}^{a_2,a_3}_{j_1,i_3}(\hat{T}^{a_4,b_2}_{i_4,i_5}(\hat{T}^{a_5,a_6}_{j_2,i_6}\hat{V}^{b_1,b_2}_{j_1,j_2}))) \rightarrow \hat{R}^{a_1,a_2,a_3,a_4,a_5,a_6}_{i_1,i_2,i_3,i_4,i_5,i_6} \tag{5.12}$$

Summation over repeated indices is implied and the parentheses indicate one choice of factorization. Assuming only a three-pair approximation on the amplitudes and integrals the contraction of the inner-most intermediate with the next amplitude is $O(o^6)$. This is easily seen if the contraction is re-indexed by pair labels, $\{P_n\}$ in one of the many possible ways:

$$\hat{T}^{a_4,b_2}_{i_4,i_5}(\hat{T}^{a_5,a_6}_{j_2,i_6}\hat{V}^{b_1,b_2}_{j_1,j_2}) \rightarrow \hat{T}^{a_4,b_2}_{i_4,i_5}(I^{a_5,a_6,b_1,b_2}_{i_6,j_1}) \rightarrow \hat{T}^{P_1,P_5}_{P_6,P_1}(I^{P_1,P_2,P_4,P_5}_{P_1,P_3}) \tag{5.13}$$

In our code (and most others), the CC equations are represented like the middle expression of Eq. (5.13), where pairwise contractions are replaced by intermediate quantities like $I^{a_5,a_6,b_1,b_2}_{i_6,j_1}$ and performed in series. The 3-pair constraint on the residual dictates that the $\{a_k\}$ and $\{i_k\}$ are cubic and reduces the cost of this term by several orders of magnitude. Pair labels are not used to index the summation loops. Instead the contraction algorithm skips multiplications which would produce an element indexed by more than three pairs.

In the usual CC theory the cost of the nonlinear and linear (CI-like) terms grow with the same power, but the linear terms of this model increase as only $O(o^5)$ while the non-linear ones, such as the example in Eq. (5.13) above, are order $O(o^6)$. Of course, if not just the residuals, but also all *intermediates* could be indexed by three-pairs this property would be restored. The resulting intermediate pair approximation seems ambitious at first blush, but would reduce the cost of the method by one more power of system size. We did not enforce the three-pair constraint on intermediates lightly because *in principle* the accuracy now rests on the choice of factorization, but experiments on every system we've examined (see representative examples in Table (5.2)) suggest that use of the 3-pair approximation for intermediates gives error that is insignificant relative to the accuracy of the PH model. Additionally the intermediate-pair approximation conserves the pair-exact property. It is similar in spirit to the recursive commutator decomposition approximation introduced by Yanai and Chan in their canonical transformation [204] work. In the supplementary information the factorization used in this code is listed, but we emphasize that only Figure 1 would be significantly altered by omission of this approximation. In other words: this computational convenience does not compromise the reproducibility of the model and both cases have an implementation.

Our pilot implementation captures the essential feature of PH: the exponent of cost.

System	Local Intermediates	Exact Intermediates
F_2 (14,14) R=1.4Å	-198.914613	-198.914619
N_2 (10,10) R=1.1Å	-109.119738	-109.119746
Cope R=1.642Å(14,14)	-232.848401	-232.848381
Dioxirane (18,18)	-188.869097	-188.869109

Table 5.2: Typical energetic effects of the intermediate-pair approximation. Frozen PP-orbitals and 6-31g* basis employed in all cases.

To verify this a series of small diamondoid hydrocarbon molecules were chosen for timing benchmarks. Their bonding grows in a three-dimensional way, but they are otherwise random[1]. The wall computation times verify that the algorithm scales more cheaply than the 6^{th} power (Figure 1) of molecular size. The number of parameters in the model grows more slowly than even the lowest order non-local VOO-CCD method, and the pre-factor is such that storage required becomes smaller for more than 31 electron pairs (Figure 2). The accuracy of the CCSDTQ56 implementation was verified by comparison with 6-electron Full-CI results provided by the PSI3 [240] program package.

New algorithmic refinements are required beyond the tools already developed for the PQ model [86] to make PH tractable. If one has two tensors with quadratic numbers of nonzero elements apiece, 4^{th} order scaling can be realized by simply focusing on the nonzero entries in coordinate representation. The same algorithms would scale 9^{th} order for cubic numbers of significant elements. In order to realize the correct scaling for PH one must only contract together blocks of amplitudes whose results will obey the 3-pair constraint and be able to construct those block pairs cheaply. The permutational symmetries of all tensors must also be leveraged.

Spin and pair-blocking were incorporated in such a way that spatial symmetries and further locality could be added as well in future work. The price paid for this generality is that each floating point operation is accompanied by many integer manipulations which vectorize poorly. Floating point effort of contraction is performed as a matrix-matrix or vector-double product in the BLAS package [241]. The algorithm has been described in a separate publication [239].

Optimization of orbitals is a highly nonlinear problem and deserves more attention than it can be given here. Our implementation of PH includes a two-step [48, 242, 243] OO-CC solver. Formally these self-consistent orbitals are desirable, but technically it is challenging to make their CPU costs outweigh their energetic benefits relative to RPP. When the orbital-optimized three-pair doubles model (3p) does not collapse these orbitals are often found to be converged for PH within tolerance. Because of the significant costs of calculating PH's orbitals we observe a simple protocol for choosing orbitals, attempting increasingly

[1] geometries are provided in the supplement

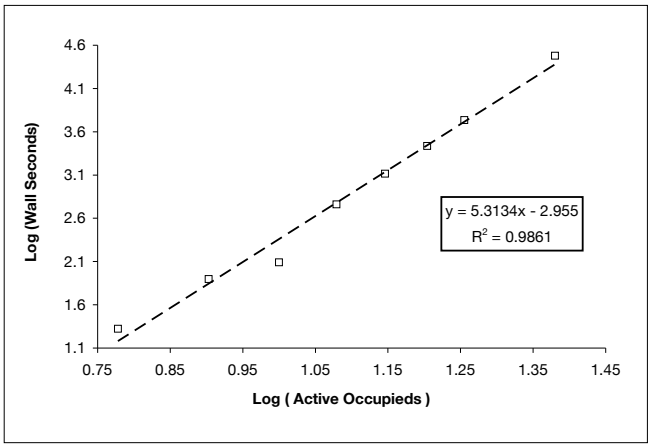


Figure 5.1: Scaling of amplitude iteration wall-time with system size. Calculations performed on one core of an Apple XServe (Fall 08). Largest system: t-butane (24,24) pictured. Parameters of a linear least-squares fit are inset.

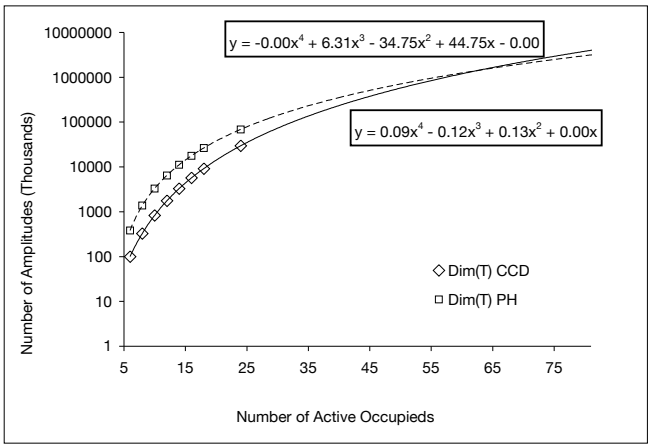


Figure 5.2: Scaling $\hat{T}$ with system size. Same systems represented by each data point as Figure 1, as included in the supplementary information. Least-Squares fits by quartic polynomials are inset.

accurate orbital optimizations (PP, 2p, PQ, PH) until convergence. This compromise is proven practical in the results. The odd-numbered multiplier equations which were also avoided in our previous work have been included in the current implementation, along with the contributions to the gradients which arise from the odd-hole-numbered density:

$$\gamma_{ov} \cdot \partial\langle f_o^v \rangle / \partial U_p^q + \Gamma_{ovvv} \cdot \partial\langle ov||vv\rangle / \partial U_p^q + \Gamma_{ooov} \cdot \partial\langle oo||ov\rangle / \partial U_p^q \rightarrow \frac{\delta \tilde{E}_c}{\delta U_p^q} \tag{5.14}$$

There are several directions where this implementation could be improved to bring it in line with the codes of related methods [244]. The model is trivially parallel up to hundreds of processors because there are literally hundreds of independent terms which involve the contraction of two three-pair objects. Only cubic quantities would need to be communicated to/from cores where 5^{th} order CPU effort would be performed. Lastly we should comment on the possibility of an implementation like those available for the PP and imperfect-pairing [1] models where the pair sparsity is built into loop structure and not recomputed "on-the-fly". We can estimate the speedup which might result by measuring the amount of time this code spends in integer manipulations vs. BLAS. In a typical profiling run of the code roughly .6% of the process' samples find the code in BLAS. The work presented here is essentially a pre-requisite for the pair-indexed algorithm because the number of unique loops will skyrocket.

5.3 Results

The PH model was implemented in a developmental version of the QChem [192] package. CASSCF calculations were performed with the aid of the GAMESS [195] package and/or PSI3 [240]. Unless otherwise noted a model employs its own orbitals which are restricted in all cases.

We have chosen the model systems to probe the approximations of PH, or a failure of it's parent. Benzene addresses spatial symmetry breaking and performance. The Cope rearrangement addresses the neglect of singles in the orbital gradient. F_2 addresses active space locality (in the context of a simple multi-reference situation), and the Bergman reaction tests locality and rank truncation simultaneously. Our figure of merit is the non-parallelity error vs. CASSCF (NPE). In every case we follow a continuous geometrical coordinate. We do this to ensure we follow a single orbital solution, and eliminate solution hopping as a source of error. To follow a continuous curve with most CASSCF implementations requires similar care, although OO-CC is somewhat more sensitive to the guess provided.

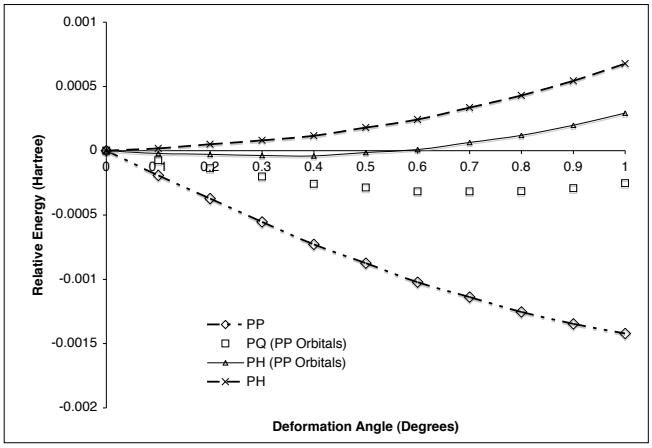


Figure 5.3: $D_{3h} \rightarrow D_{6h}$ deformation of benzene in the 6-31G basis

5.3.1 Benzene

The resonant delocalization of π-electrons in a benzene ring, and the resulting D_{6h} symmetry are undoubtedly elementary to the chemistry of carbon, however local valence correlation models are known to favor bond alternation if the locality demanded of the wavefunction is beyond reality of the electronic structure [125, 126, 190]. A similar effect can be seen with other popular correlation models if an insufficient basis is employed [245]. Given that the 3-pair constraint treats 6 electrons quantitatively, we would expect it to repair this deficiency. Indeed we find this to be the case in both the (6,6) and (30,30) active spaces (Figure 2). Even given the orbitals of the PP model (which are severely symmetry broken) the PH amplitude equations distort benzene by less than $.1mE_h$. An iteration of the CC-6 amplitude equations in the (30,30) space given 3-pair constraint and local intermediate approximation requires roughly 30,000 seconds of wall-time in our most recent implementation on one core of a typical cluster node. No spatial or spin symmetries are exploited. Our sparse storage scheme and factorization of the CC equations require roughly 10GB of disk for applications of this size. The deformation angle is the difference between consecutive C-X-C angles, where X is the center of benzene's mass. R(X-C) = 1.3Å, and R(X-H) = 2.0 Å.

R(Allyl-Allyl) $\mathring{A}$	CASSCF (6,6)	PP	PQ	PH
1.64	-232.98024	0.00778	0.00112	0.00002
1.695	-232.97954	0.01074	0.00174	0.00002
1.75	-232.97820	0.01453	0.00259	0.00002
1.805	-232.97703	0.01892	0.00360	0.00001
1.86	-232.97635	0.02350	0.00466	0.00000
1.915	-232.97615	0.02794	0.00570	-0.00001
1.97	-232.97628	0.03208	0.00666	-0.00001
2.025	-232.97656	0.03589	0.00754	-0.00001
2.08	-232.97685	0.03941	0.00832	-0.00001
2.135	-232.97706	0.04268	——	-0.00001
2.19	-232.97713	0.04577	——	-0.00002
NPE(kcal/mol)	——	23.84	4.52	0.02

Table 5.3: Energies along the D_{2h} coordinate of the Cope rearrangement in the (6,6) space and 6-31g* basis. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH.

5.3.2 Cope Rearrangment

The Cope rearrangement is a classic of organic chemistry and has been carefully examined [211,212] in a quantum chemical context, especially by the group of Davidson [209,210]. An interesting coordinate to examine is the inter-alyllic distance along the D_{2h} slice of the potential surface the reaction coordinate passes through. Two qualitatively different minima are found by CASSCF as this distance increases, one "concerted" and another of biradicaloid character. We will examine several geometries along this curve obtained from CASSCF optimizations with the inter-allylic distace constrained[2].

The most often used CASSCF space for this problem is (6,6), the same number of electron pairs we have constructed PH to treat quantitatively and so we use it as another control experiment to assess the neglect of singles with the variational gradient. Table 1 compares the performance of PH to its predecessor PQ. Distinguishing between the two minima separated by only $3mE_h$ on the CASSCF surface requires resolution beyond the usual $15mE_h$ that can be expected from PQ for a system of this size. The NPE indicates the difference between the maximum and minimum deviation from CASSCF. Neglect of the singles block does not hinder PH's ability to faithfully reproduce CASSCF for this system.

[2]We would like to acknowledge Troy Van Voorhis for generating these geometries during his Ph.D. and Greg Beran for a related conversation. See Supporting info.

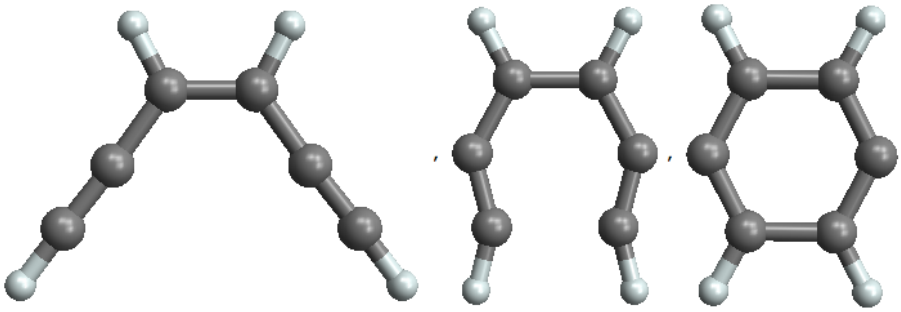


Figure 5.4: Three representative points along the path of the Bergman reaction followed in this study. From left to right, Step 0.0 (eneyne), 0.5 (near transition structure, and 1.0 (*p*-benzyne).

5.3.3 Bergman Reaction

The Bergman reaction is interesting for many reasons, not the least of which is its role as a fuse for a class of cytotoxic natural products with therapeutic potential. The mechanism of the reaction is well-understood theoretically and experimentally, and is known to be a difficult case [246, 247] for correlation models making it an ideal benchmark system. The product: p-benzyne [215, 248], is a singlet diradical. DFT calculations can provide poor results for such systems, even predicting a bond-stretch isomer of m-benzyne which upon further examination appears not to exist [249]. Adopting the eneyne and transition state geometries of Cramer, and the benzyne geometry optimized with Spin-Flip DFT [182], we will make a very coarse study of a reaction coordinate obtained by a quadratic interpolation between these geometries (Figure 2) in Cartesian coordinates[3] and compare against CASSCF(8,8) (the space employed by Mazziotti for benzynes) to assess our model on a practical problem with a strongly delocalized electronic character.

PH's parent model PQ provides a semi-quantitative picture of the reaction coordinate with accuracy in keeping with our previous findings, usually with correlation energies a few mE_h less than CASSCF for only a few seconds of CPU time. Immediately at Cramer's Transition state (step 0.5) where we expect most multi-reference character, we see that PQ over-correlates presumably because of strong non-separable correlations of more than 4 particles whereas PH remains variational. These geometries likely deviate significantly from the lowest energy path between endpoints, but that only adds difficulty to the problem of reproducing this slice of the CASSCF surface. Without the intermediate-pair approximation the cost of the calculation increases significantly, but the result does not change much. For meta-benzyne (step 1) the correlation energy of PH without the intermediate-pair approximation is $0.4mE_h$ larger, and for the ene-yne the same figure is 0.2 mE_h.

[3]The coordinates of each point are available in the supplementary information.

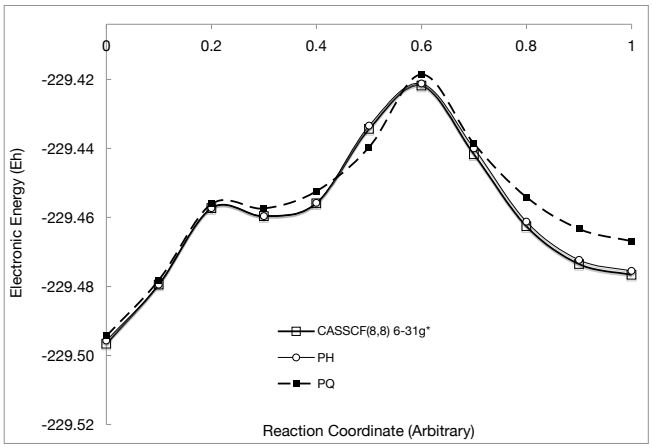


Figure 5.5: Electronic energy along Bergman reaction coordinate. CASSCF(8,8) in the 6-31g* basis.

Step (arb.)	CASSCF(8,8)	PP	PQ	PH
0.0	229.49761	0.02789	0.00227	0.00090
0.1	-229.48047	0.02530	0.00129	-0.00002
0.2	-229.45835	0.02487	0.00124	0.00001
0.3	-229.46065	0.02601	0.00224	0.00003
0.4	-229.45690	0.02885	0.00350	0.00012
0.5	-229.43533	0.03809	-0.00532	0.00092
0.6	-229.42281	0.03541	0.00327	0.00061
0.7	-229.44270	0.03237	0.00301	0.00170
0.8	-229.46354	0.04247	0.00842	0.00129
0.9	-229.47450	0.03843	0.01017	0.00118
1.0	-229.47767	0.03924	0.00977	0.00117
NPE(kcal/mol)	—	11.0	9.7	1.1

Table 5.4: Study of Bergman Reaction in the (8,8) space 6-31g* basis. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH

R(F-F) Å	CASSCF	PP	PQ	PH
1.6	-198.94906	0.09720	0.01590	0.00091
1.9	-198.92881	0.08053	0.01311	0.00095
2.2	-198.91505	0.07110	0.01049	0.00121
NPE(kcal/mol)		16.37	3.40	0.19

Table 5.5: F_2 dissociation in the DZ basis and (14,14) space. Total energies are given for CASSCF in E_h, while deviations from CASSCF in E_h for PP, PQ, PH.

5.3.4 F_2

A pair of fluorine atoms nominally share a single bond, but the extreme reactivity of the gas and poor performance of CCSD [250, 251] demonstrate significant multi-reference character. The full-valence active space (14,14) is also roughly the limit of a routine CASSCF calculation and large enough to seriously assess the energetic impact of the three-pair locality constraint. Table 3 quantitatively compares CASSCF, PP, PQ and PH for this problem with the DZ [252] basis and (14,14) active space. Again PQ semi-quantitatively parallels the CASSCF curve whereas PH maintains accuracy throughout. Without the intermediate-pair approximation the energy does not change significantly even with this very large active space. At $1.5\mathring{A}$ with the approximation PH yields an energy of: -198.943619E_h, and without it: -198.943622E_h. In the absence of symmetry are 11,778,624 determinants in the CI space, and 11,102 amplitudes in PH.

PH can be seen as two approximations to CASSCF: rank truncation and spatial locality. This model system is directed largely at the latter approximation. We only enforce locality *implicitly* through orbital optimization of the correlation energy. The rank-truncation approximation is more explicit, in the sense that we know what determinants are relevant for a given bond dissociation process and we know that our model will be faithful if it contains those determinants. These results are not predictable by construction, and interesting for that reason.

5.4 Discussion and Conclusions

We have presented a tractable approximation to CASSCF asymptotically containing fewer parameters than MP2 with many nice properties. The foremost of these is the systematic improvement of PH over PQ which results from the well-tempered nature of the pair approximation. The pair-approximation based models (PP, PQ, PH) can now be made into something bigger than the sum of their parts because one may calibrate affordable calculations with quantitative calculations; for carbon systems this is now possible at a cost which is tractable in many new and exciting cases of interest on a single processor. This

may be contrasted with the situation at present: if an active space larger than 16 electrons is needed there is no routine approach, but many promising candidates which may find broader application soon [180, 205, 215, 242]. In future work dynamical correlation must be added to the OO-CC references in an efficient way; we have mentioned many possible approaches already and have some results in-hand which will immediately follow this paper.

The success of this model for first-row systems stems from the fact these atoms participate in at most three bonds. Moving down the periodic table to higher nuclear charge more electron pairs are forced into a smaller region of space and correlate with each other. The effectiveness of the pair approximation in transition metal bonding is an open question; however the infrequent appearance of Lewis structures with more than triple bonds in the chemical literature offers a reason for optimism.

A comprehensive understanding of the efficiency of this model relative to other approaches is also desirable, but challenging to realize. There are two obvious measures which can be tested easily: wall computation time and number of parameters. In a Density Matrix Renormalization Group(DMRG) calculation the dimension of the many-body basis is a free parameter and so there is an opportunity to compare "apples to apples". It is not necessarily enough to judge each active-space method without dynamical correlation because ultimately exact solution within the active space may prove inefficient.

The pair-approximation idea might be fruitfully combined with other formalisms. The renormalized cluster models [116] could be combined with PH, projecting the generalized moments onto the cubic subset of hextuples which we know are relevant and avoiding explicit construction of T_6. We could apply the dynamical correlation theories [225] that were originally conceived for density-matrix based approaches to our own work. Incomplete active spaces are required to address large systems, and these ideas can provide them in an accurate and well-defined way.

Chapter 6

The +SD Correction

6.1 Introduction

The electronic structure of molecules can often be captured by an approximate wavefunction consisting of a single determinant. This is usually the case for molecules near their equilibrium geometries. For these situations the combination of coupled cluster methods with techniques for tackling the basis-set problem [102] can be relied on to reproduce or predict chemical phenomena given some polynomial amount of computational time which is always decreasing. If more than one determinant is required to qualitatively treat a problem, even obtaining a qualitatively correct electronic structure becomes challenging. Efforts in the quantum chemical community have largely adopted a divide-and-conquer approach where the static and dynamic correlation problems are handled separately. The Complete Active Space Self-Consistent Field (CASSCF) wavefunction [46] is used routinely to solve the former, although its cost scales exponentially limiting the method to roughly 16 electrons. Dynamical correlation is most often dealt with by second-order perturbation theory [167]. Correlations near the interface can cause "Intruder-State" problems [253], which must be pushed away by level-shifts.

This artificial separation of correlation problems is not reflected in the coupled cluster theory itself [50, 254]. It is non-perturbative, and only made unsuitable for multireference problems because of the rank-truncation of the correlation problem. Even double excitations have significant flexibility [54], but the Schrödinger equation must be projected against the significant high-rank determinants [116] to maintain pseudo-variational behaviour. Rank truncation is motivated by Möller-Plesset perturbation theory and computational convenience. For situations where perturbation theory fails [255], we can do much better than rank truncation. This paper explores one such choice, based on our local, orbital-optimized coupled-cluster (OO-CC) models [86, 126].

The formalism is nothing more than the standard CC theory applied everywhere in quantum chemistry (Eqns. 3-4), but with an unconventional truncation of $\hat{T}$. Like CASSCF,

OO-CC moves the strong correlations into an active space of a few orbitals. Orbital optimization is also used to localize the active space, reducing the number of amplitudes we introduce by several orders of magnitude [86] especially for the higher ranks. To treat the dynamical correlations we must allow for excitations from these configurations into the external space. The resulting model blurs the line between static and dynamic correlation. A strong analogy can be drawn between this model and MRCI [256–259].

The traditional coupled cluster theory has many nice properties: orbital insensitivity, size-consistency, and formal simplicity. Unfortunately no multireference version has been so uniquely defined, although given an extended normal ordering [133] it's formulation may be possible. Even generalizing without sacrificing vital features required a large body of work [169, 260–263] which is now growing into mature implementations [201, 264–266]. There are two points of departure between these ideas and the model presented here. The MkMRCC, and its relatives are based on a wave-operator formalism [267], which presents the advantage that they are truly free of a 1-determinant reference [268], and several challenges: redundant amplitudes, intruder states, and size-intensivity.

The second difference is the choice of an incomplete reference space. Most all developmental MRCC codes have been based on complete active spaces, which make them effectively exponentially scaling and limited in scope. Of course the wave-operator methods are easy to imagine with a reduced space [269], but this isn't often done. The ideas presented in this paper for limiting the cost of the active-space treatment with a controlled size-consistent approximation are transferrable to other MRCC formalisms. However incomplete active spaces are somewhat at odds with the orbital invariance feature. Indeed the goal of a local model often becomes strong orbital *variance* for the sake of efficiency. Making this compromise we must be careful that the orbitals are well-defined, and well-behaved with respect to symmetry breaking. This paper exists largely to see how well we can tackle those problems.

Shortly after the appearance of the first coupled cluster quadruples (CCSDTQ) implementations [96, 97] the resulting code was adjusted to formulate a coupled cluster model for 2-configuration multi-reference problems [171, 172], and this idea enjoys continued application and development today under many names [270–278]. The advent of general-rank symbolic coupled cluster codes [112, 199, 279] have made these ideas commonly available, and this work owes much to the significant literature in the area. In this paper we use the conventional name SRMRCC (Multi-reference Coupled-Cluster Theory based on a Single Reference formalism) for these models.

Given only a two-electron density matrix there are also techniques which can dynamically correct towards a total electronic energy [204, 226]. These have a clear strength that their cost is insensitive to the size of the reference space, although through extended normal ordering [133] this property could be brought to any model. One of the features of this work relative to the aforementioned models is that with minimal modification there are already (freely available) codes capable of iterating the coupled cluster equations as this paper will describe. Some of these are highly-optimized, well-understood, and are accompanied by

code for their gradients and response properties. Furthermore the orbitals of OO-CC are orthogonally localized [56, 122], and this sparsity can be used to construct affordable local models with techniques already known [57] to the quantum chemistry community. Because the model proposed in this paper can be described with minimal formalism, and coded with minimal effort (beyond a general CC implementation), it offers a simple test-bed for local models of the total energy.

6.2 The Models

We assume the usual spin-orbital basis of orthogonal, but non-canonical spin-orbitals. These are divided into four subspaces: external occupied $\{\mathcal{O}_k\}$, active occupied $\{i_k\}$, active virtual $\{a_k\}$, and external virtual $\{\mathcal{A}_k\}$. The active space is further partitioned into pair-quartets of spin orbitals $\{i_\alpha, a_\alpha, j_\beta, b_\beta\}$ and all these labels are defined by the orbital-optimization of the underlying active-space cluster model. Occupied and virtual orbitals in either space are denoted $\{o_i\}$ and $\{v_k\}$, respectively.

These local approximations to CASSCF are uniquely defined by a number of pairs **n**, and any rank truncation imposed on the OO-CC amplitude. If the rank truncation is implied by the pair constraint, because **n** pairs can make at most 2**n** excitations, we call the model the "perfect"-"2**n**-tuple" by analogy with the perfect pairing model (ie: PP, PQ, PH). These models are exact for non-interacting **n**-pair systems by construction. If the excitation rank is separately limited we just give it a "pair number"-"rank limit" appellation, ie: the three-pair quadruples are abbreviated "3Q". With certain algorithms the cost of 3Q scales like PH, but the former can usefully save on pre-factor. Unlike a string-based MCSCF code OO-CC is naturally extensive in all cases, and often cheap enough that the reliable, uniquely defined valence active space can be chosen.

The basic idea is now quite simple (but simplicity should never be regarded as a vice). We would like to dynamically correlate any determinant within a reference space, and based on our experiences with the traditional cluster theory we say these dynamic correlations lie in the space of doubles above this reference space. This suggests we allow any amplitude of the form:

$$\hat{T}^{a_1,a_2,\ldots,a_n,\mathcal{A}_1\ldots\mathcal{A}_k}_{i_1,i_2,\ldots,i_n,\mathcal{O}_1\ldots\mathcal{O}_l} \; ; \; s.t.\{a_k,i_k\} \subset \{\mathbf{n}\text{-pair model}\}, \text{ and } 0 \leq k,l \leq 2 \tag{6.1}$$

and allow this truncation to replace any concerns of rank in our cluster theory. We only do this after the orbitals have been optimized. This truncation, which will be called *Ansatz (6.1)* in the remainder of the paper, includes the underlying OO-CC model, external and semi-internal doubles, and singles over the whole space space. An n-pair model excites to at most 2n particles, and the model described above excites to 2n+2. If the frozen-core and valence-PP active space are chosen (as we recommend) then this model *doesn't increase*

the maximum excitation rank above PQ/PH because $\{\mathcal{O}_i\}$ is empty. The storage costs and computational scaling grow with reference size and basis-set size like the usual state-specific SRMRCC [280].

We have performed experiments with another choice *Ansatz (6.2)* which relaxes the locality of the active space by including any amplitude which is related to its parent by the replacement of at most two *labels* (pair labels or spin-orbital labels for the external space):

$$\hat{T} = \begin{cases} \hat{T}^{a_1,a_2,\ldots,a_n,v_1,v_2}_{i_1,i_2,\ldots,i_m} + \hat{T}^{a_1,a_2,\ldots,a_n,v_1}_{i_1,i_2,\ldots,i_n,o_1} + \hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_m,o_1,o_2} & \text{s.t.}\{a_k,i_k\} \subset \{\mathbf{n}\text{-pair model}\} \\ \hat{T}^{a_1,a_2,\ldots,a_n,v_1}_{i_1,i_2,\ldots,i_m} + \hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_n,o_1} & \text{s.t.}\{a_k,i_k\} \subset \{\mathbf{n+1}\text{-pair model}\} \\ \hat{T}^{a_1,a_2,\ldots,a_n}_{i_1,i_2,\ldots,i_m} & \text{s.t.}\{a_k,i_k\} \subset \{\mathbf{n+2}\text{-pair model}\} \end{cases} \tag{6.2}$$

$$E = \langle 0|e^{\hat{-T}}\hat{H}e^{\hat{T}}|0\rangle = \langle 0|\{\hat{H}e^{\hat{T}}\}_c|0\rangle \tag{6.3}$$

$$\langle \mu_s|\{\hat{H}e^{\hat{T}}\}_c|0\rangle = 0 \tag{6.4}$$

Note every amplitude of Ansatz (6.1) is also an amplitude of (6.2), along with most of the non-local excitations although the gross storage cost scales similarly in the large-basis limit, $(\text{pairs})v^2 \approx \text{NMO}^3$ for PP+Ansatz 2, and $(\text{pairs})^3v^2 \approx \text{NMO}^5$ for PH+Ansatz 2. This model is more appropriate if the reference is far from exactness, but likewise more costly.

Given a perfect pairing active space of 2n electrons in 2n orbitals and v external virtual orbitals, in the most costly case (PHSD) and assuming a three-pair constraint on amplitudes and intermediates the storage cost of Ansatz (6.1) scales n^3v^2, and the CPU effort scales n^3v^4 These scalings are frankly prohibitive for molecules of appreciable size, but before we can experiment with local dynamical approximations in the external space we must determine the accuracy of this base model. The difference between these models and the SRMRCC [270] is two-fold. The active-space problem is not solved separately in a CI, and the expansion in the active-space is not complete. All amplitudes are on the same footing (most amplitudes are not clearly static or dynamical) and every correlation is made consistent with the others during CC iterations. This is much like what occurs if $\hat{C}$ and $\hat{T}$ are solved in a two-step iterative process [270] in the SRMRCC.

The simplicity of these models is pleasing. Every amplitude possesses a unique nonlinear equation it must satisfy derived from projections of the Schrödinger equation, and these equations are well known [281]. There are no sufficiency conditions [169], or intruder state problems [282, 283], but experience with the usual SRMRCC [280, 284] raises some concerns. Because the coefficient of the OO-CC's reference determinant is set to one by intermediate normalization, in a situation where its relative weight in the wavefunction drops to zero, convergence of the amplitude equations will become difficult. This doesn't differ from the usual CC and isn't much of a practical concern. A denominator shift

gradually relaxed during iterations is usually all which is required. In a similar vein, only the formal determinant experiences the highest rank of excitation and so when the dominant configuration changes abruptly the correlation energy may not behave smoothly. In the literature of SRMRCC this is called a Fermi vacuum or orbital invariance problem [279, 280]. Luckily the CC-theory is strikingly insensitive to choice of reference because of the exponential single excitations, and our model includes these naturally between all spaces. Moreover, orbital-invariance was already sacrificed to introduce locality into the underlying OO-CC model, a sacrifice which must be made to escape exponential scaling. We emphasize that our goal for this work was to demonstrate the strength of the local OO-CC as a reference in a conceptually straightforwards, but quantitative model. Like any formal drawbacks, the final measure should be the results.

6.3 Results

For the purposes of this paper ansatz (6.1) will be called "+SD", and that will be attached to the name of the OO-CC model underneath (ie: PPSD, PQSD, PHSD, etc...). Applications of ansatz (6.2) will be specifically noted. Our figure of merit will be the non-parallelity error (NPE), the difference between the maximum and minimum error relative to the exact result. In future work the external space should also be made local. Until then larger active spaces will imply a more local model and necessarily smaller correlation energies. Still if the pair approximations are good (and they usually are) the NPE should remain good and this will be shown below.

We employ some common benchmark systems to evaluate the performance of our models. The double dissociation of H_2O demonstrates the compactness and chemical accuracy afforded by this truncation. F_2 largely probes the impact of active-space locality. BeH_2 is troublesome from an orbital invariance standpoint, but the PQ model is non-local and rank-complete for the reference problem. H_8 provides an example where orbital invariance, symmetry and locality are all challenging at once.

The "+SD" models have been implemented in a developmental version of the QChem [192] package. Exact results were furnished by the PSI3 [240] program package. We also compare against the previously available dynamical correlation correction to OO-CC [230] developed by our group. The valence PP active space is generally assumed, as is the frozen-core approximation. The basis is generally DZ [285] as obtained from EMSL's Basis Set Exchange [286, 287].

6.3.1 H_2O

We examined the simultaneous dissociation of water with PQSD and the (8,8) active space, the errors relative to complete solution in the DZ basis with the frozen-core approx-

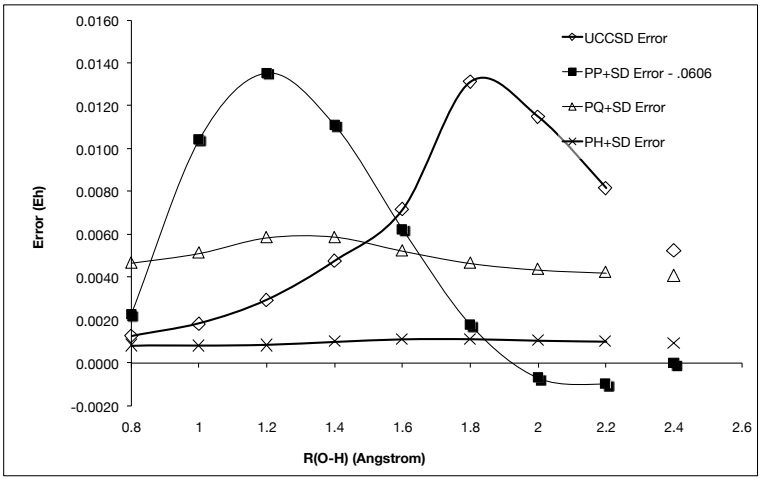


Figure 6.1: Simultaneous dissociation of H_2O in the DZ basis. Orbitals are those of the GVB-RCC model [1].

imation are pictured in Figure 1. In the absence of symmetry the complete expansion is 1,656,369 determinants and there are 2,745 amplitudes in CCD. PQSD is a significant truncation of the correlation problem with only 2,984 amplitudes, but maintains a very modest non-parallelity error (NPE) of 1.07 kcal/mol, re-enforcing the efficiency of the pair-based reference. Even with rather crude PP orbitals PH+SD maintains an NPE of 0.19 kcal/mol.

6.3.2 F_2

Electrons are loosely paired between Fluorine atoms in a multi-referenced bond that is poorly described by low-rank single-reference cluster theory [288, 289]. We also include the best unrestricted curve which can be coaxed from B3LYP [290] for any newcomers to this area. The valence perfect-pairing space, (14,14), is also roughly the limit of complete diagonalization, and so this one of the more stringent tests of the locality approximation we can produce. We've previously demonstrated that the PH model provides chemical accuracy for the valence CASSCF correlation problem. In the same study we saw that two-pair locality can only provide an NPE to CASSCF of roughly 4 kcal/mol.

The truncation introduced in this paper does not relax active-space locality, and so we expect the errors relative to CASSCF to translate almost quantitatively into error relative to the total correlation energy (if we have captured the bulk of the dynamical correlations). Unfortunately we are unable to provide a FCI curve, but for the purposes of this work the complete CCSDTQ model should provide near-exact results and the errors for this example are reported relative to this benchmark. The core orbitals are frozen in all cases. In the

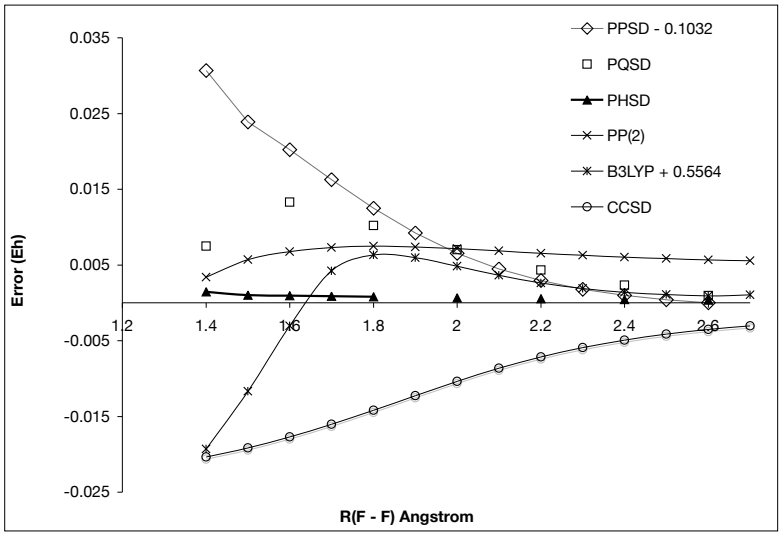


Figure 6.2: Dissociation of F_2 in the DZ basis, Errors relative to CCSDTQPH (frozen-core). PHSD employs the intermediate-pair approximation.

nonlocal model there are 4,597 doubles, 112,236 triples and 1,697,198 quadruples. In the 3-pair model those figures are reduced by factors of roughly $\frac{1}{2}$ and $\frac{1}{10}$ respectively.

The PH paper also introduced an intermediate-pair approximation which offers a palpable reduction in computational complexity of a three-pair model. This approximation only takes effect when the number of active pairs significantly exceeds the locality constraint (as it does in this example). The same idea can be applied to the "SD" model without any modification. The results of Figure 2 are obtained in the presence of that approximation. If this approximation did not carry to the "+SD" models accurately, other avenues would have to be pursued to relax the cost of this method. One direction would be including only rank density-matrix information from the reference as is done in Canonical Transformation [227].

6.3.3 BeH_2

Insertion of Beryllium into a hydrogen molecule is a classic multi-reference problem because of an avoided crossing in the intermediate region. Of the examples recently considered

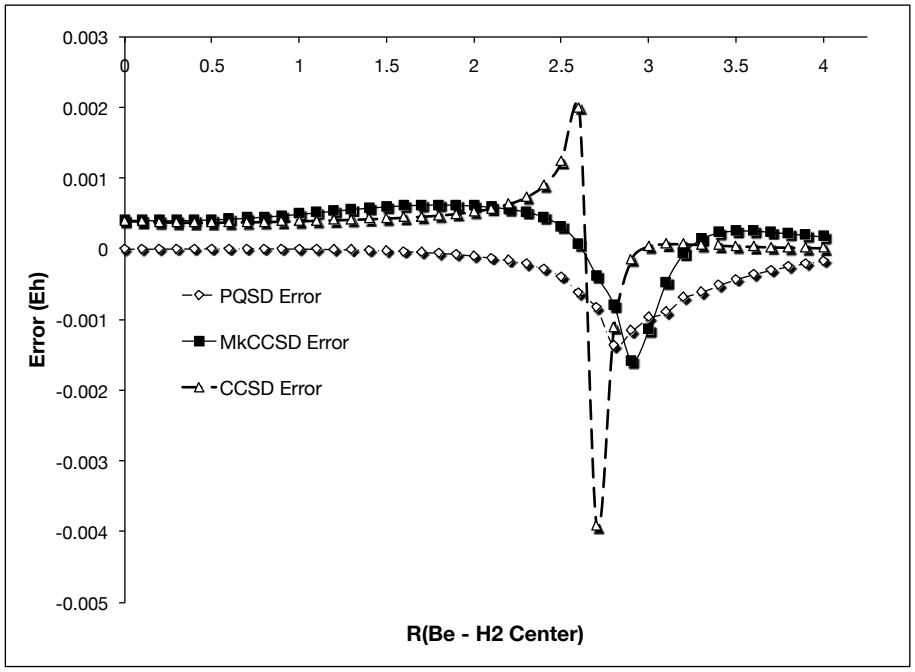


Figure 6.3: Insertion of Be into H_2. PQSD is built on the (4,4) active space. Orbitals are those of PQ.

by Hanrath [280] with several MRCC approaches this system posed the most difficultly for SRMRCC. The dominant reference changes symmetry between endpoints with the highest-occupied molecular orbital shifting($3a_1 \rightarrow 1b_2$). We employ the geometry and basis set of Evangelista [265][1] from whom we also borrow benchmark results. At either endpoint the orbitals are optimized for PQ from a guess of GVB-RCC [68]. All internal data points use an orbital guess provided by the previous geometry.

In Figure X the errors relative to FCI are shown for MkCCSD(2,2), PQSD(4,4) and CCSD. The active spaces employed in MkCCSD (2,2) and PQSD (4,4) *are not the same.* We stress that MkCCSD possesses strong orbital invariance properties which this model does not, and the fact that they perform similarly in different active spaces is a reflection of that strength in MkCCSD. This graph suggests that orbital-invariance isn't a fatal concern for these PQSD. With larger spaces the locality we exploit in PQ or PH is a more significant error.

[1]We would like to thank the authors of this important paper for providing very clear and complete data throughout.

α/π	FCI	PQSD Error (Eh)
0.0	-2.07753	0.00321
0.1	-2.17098	0.00449
0.2	-2.21589	0.00438
0.3	-2.23302	0.00433
0.4	-2.23958	0.00434
0.5	-2.24131	0.00434
NPE (kcal/mol)		0.80109

Table 6.1: The H_4 model system. Zero is a 2-Bohr square of hydrogen atoms, and .5 is the linear configuration. The orbitals are those of PQ. The reference space is (4,4) and basis DZP.

6.3.4 H_4

Four hydrogen atoms are arranged in a 2 Bohr square, and the opposite sides of that square are folded downwards by an angle determined by parameter, α. As a square ($\alpha = 0$) the electronic structure is dominated by two equally important determinants which are separate as the structure is made linear ($\alpha = \frac{1}{2}$). This model was introduced by Jankowski and Paldus, and has been previously examined by several authors [201, 280, 291, 292].

Table 1 summarizes the error of PQSD relative to FCI in the DZP basis. Given that this is a 4 electron system PQ is not a local approximation and the anstaze closely resembles SRMRCC. The orbitals are orbital-optimized doubles-quadruples rather than CASSCF (2,2), and there is no CC/CI separation. There are far fewer amplitudes in PQSD (501) than the complete CCSDTQ model(6529). The performance of PQSD is satisfactory, but predictable. We include this example mostly for the sake of completeness.

6.3.5 H_8

In this standard MRCC test system [201, 291–295] eight hydrogen atoms are arranged in a 2 Bohr octagon. Two opposite faces are pulled away from each other by a displacement α (Figure 4). As in the previous examples two determinants differing by the replacement of an orbital ($b_{1g} \rightarrow a_g$) exchange dominance in the wavefunction becoming quasi-degenerate when the molecule is symmetric. MkCCSD built on CAS(2,2) has an NPE of 2.019 mE_h for this system [201], a PP(2) calculation with the (2,2) space has an NPE of 18 kcal/mol. Table 2 examines the the errors of PPSD(2,2) and PHSD(8,8) (with intermediate pair approximation) relative to complete solution in the DZ basis. The orbitals are optimized for the PH reference. In the (2,2) space there are no symmetry breaking problems for PPSD, and we obtain a curve of quality comparable to MkCCSD. The most inexpensive OO-CC models (PP, IP [1]) over-localize orbitals, and under-correlate weak bonds at the expense

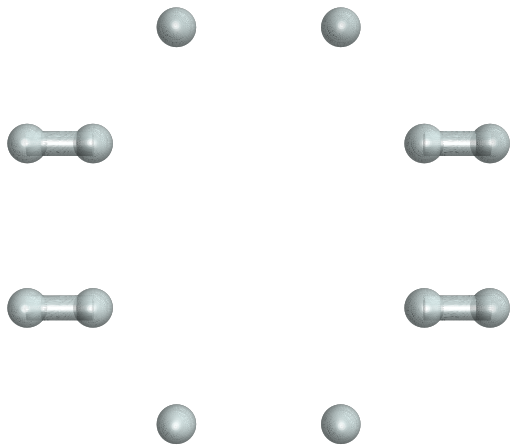


Figure 6.4: Graphical depiction of H_8 model for α = -0.4, 0.4 Bohr

of stronger bonds. Aside from the same multi-reference problems addressed elsewhere in this paper, this symmetry breaking presents a real challenge for a 3-pair local model of the valence electronic structure in H_8.

For the compressed geometries the three-pair approximation and PHSD(8,8) fares well, and the orbitals can be converged easily. When $\alpha > 0$ the 3-pair orbitals become substantially more difficult to converge than the non-local doubles indicating that the local approximation is being forced on a somewhat non-local problem. The errors shown in Table (6.2) for $\alpha > 0$ reflect errors of 3p against CASSCF much more than they reflect any deficiency of "+SD" itself. This motivated us to develop the less-local Ansatz 2.

However Ansatz 2 restores quantitative accuracy, even when built on a two-pair reference as shown in Figure (6.5).

6.4 Discussion and Conclusions

The pair-approximation based models have been developed with computational cost and practicality as the guiding principle. The assumption that an accurate total energy model could be made from PQ or PH given positive comparison to CASSCF has been justified in this paper. Because they make the valence active space affordable, and because the OO-CC orbitals are well-defined Fermi-vaccuum invariance isn't the main challenge for this scheme. The symmetry breaking inherited from GVB, and largely ameliorated with a 3-pair constraint is the dominant source of error.

We do not claim that this work is the grand-unified, black-box correlation model, but the results indicate that the ideas introduced are useful. Combinations of OO-CC with other ideas like MkCC [169], the anti-hermitian contracted Schrödinger equation [226] or

α(Bohr)	FCI (E_h)	PPSD (2,2) Error	3QSD (8,8) Error
-0.4	-4.35706	0.00085	-0.00141
-0.3	-4.35063	0.00101	-0.00178
-0.2	-4.34176	0.00134	-0.00222
-0.15	-4.33701	0.00167	-0.00238
-0.1	-4.33266	0.00223	-0.00232
-0.05	-4.32954	0.00314	-0.00169
0.0	-4.32872	0.00425	-0.00004
0.05	-4.33074	0.00304	0.00929
0.1	-4.33509	0.00209	0.00526
	NPE kcal/mol	2.13	7.32

Table 6.2: Automerization of H_8. The basis is DZ. MRPH employs the intermediate pair approximation

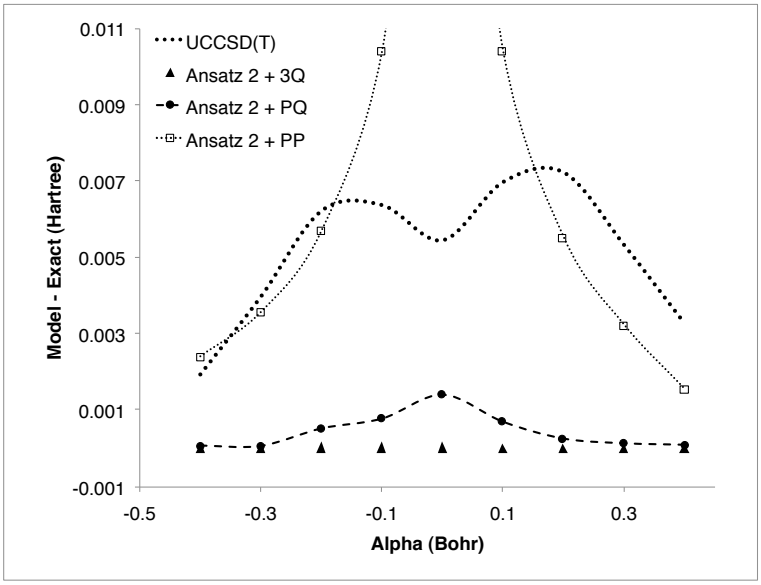


Figure 6.5: Automerization of H_8. The reference space is (8,8) and basis is DZ.

Canonical Transformation [227] would also be interesting.

Given the attractive simplicity of this work, it is interesting to ask how this dynamical correction might be made as efficient as it's static correlation counterpart. We are working off this footing to produce a entirely local correlation model. Along similar lines some recent studies [296,297] made local correlation models from the renormalized cluster theory. Another fascinating direction would replace the Gaussian orbitals of the external space with explicitly correlated geminals. Given sustained growth of computational resources, systematically-improvable approximations like these whose cost grows with some reasonable polynomial are poised to make a large impact.

Chapter 7

A Novel Range Separation of Exchange

7.1 Introduction

Despite rough beginnings [25], the local density approximation (LDA) has been developed through decades of work on the Kohn-Sham (KS) [80] construction into one of the most successful approximations in quantum chemistry and solid state physics. Within this framework, the LDA exchange correlation functional is combined by an adiabatic connection with a non-interacting wavefunction so that an approximate kinetic energy may be extracted and there is no need to develop accurate functionals for the kinetic energy [298] which have proven elusive. Along these lines, Becke [290] realized that the accuracy of Kohn-Sham energy functionals could be improved by the admixture of "exact" exchange coming from the explicit exchange energy of the fictitious Kohn-Sham wavefunction. The resulting hybrid density functionals have been the most commonly applied model chemistry for many years [299] because they have been found to be remarkably accurate with computational costs virtually equivalent to those of the Hartree-Fock (HF) method.

One of the few remaining substantial defects of the Kohn-Sham construction which has attracted theoretical effort is the so-called self-interaction problem [300–306], and it is directly related to the treatment of the exchange energy [83]. In the HF energy expression the Coulomb repulsion of a 1-electron function with itself is cancelled exactly by the corresponding exchange integral. In the KS construction with a pure, local functional the Coulomb energy is non-local, but the exchange energy is not. Considering the 1-particle functions provided by the KS wavefunction we might say that the electron repels itself if the particle is spread over space because the antisymmetric complement of the Coulomb interaction, non-local exchange, is missing. At equilibrium geometries the effect on predicted ground state energies is not severe, but this defect means that dissociation problems may lead to fragments which only possess a fractional number of electrons [307,308], or re-

sponse properties which reflect serious artifacts if charge is significantly redistributed [309]. If globally a fraction of the exchange energy of the KS determinant is mixed with the DFT exchange energy these artifacts are partially remediated. To a stranger unfamiliar with the history of hybrid DFT's development the situation must seem confusing, because it is not obvious why any mixture of "exact" exchange with (semi-)local exchange is advantageous. The answer is that the accuracy of most DFT functionals lies in a cancellation of errors between exchange and correlation functionals. Both are compensating for the single-reference nature of the fictitious KS determinant [310, 311]. From another angle, one might say that these two non-local objects [312] are best considered together because what results is more local.

A way to preserve the local cancellation of errors yet recover correct exchange at long range has emerged in the form of range separated hybrid functionals. The idea which goes back to the pioneering work of Gill and Savin [214, 313, 314], is to divide $1/r$ by multiplying it with a function which varies between 0 and 1, such that both this function and its complement are integrable. The greatest fraction of work in this very active area [315–320] has employed the standard error function to achieve this separation:

$$1/r = \mathrm{erf}(\omega r)/r + \mathrm{erfc}(\omega r)/r \tag{7.1}$$

The LDA exchange functional corresponding to $\mathrm{erfc}(\omega r)/r$, and integral kernel for the exact exchange over $\mathrm{erf}(\omega r)/r$ can then be derived so that locally exchange is provided by the LDA and at a distance exchange is provided by the KS wavefunction. The position where the transition is smoothly made between the two treatments is determined by the adjustable parameter $1/\omega$. The choice of the error function as a Coulomb attenuator is both practical (for most implementations one must be able to perform the integral of Gaussians over the function analytically [321]) and arbitrary because the error function is just one of many which possess this property. In some recent studies promising results have been attributed to more flexible range separation [322].

One can extend the idea of mixing ab-initio and DFT strengths further by imagining range separation of the correlation part of the functional as well. In this scheme (which we will not pursue in this work beyond mention) the ab-initio method is made responsible for static and long range dispersion effects while the LDA correlation functional is adjusted for the modified coulomb interaction to avoid double counting. Savin and coworkers have experimented with the choice of another attenuated Coulomb interaction for these purposes [323] , a linear combination of an error function and a Gaussian which offers a sharper separation:

$$v_{ee,\mathrm{erfgau}} = \mathrm{erf}(\omega r)/r - (2\mu)/\sqrt{\pi} * e^{-(1/3)\mu^2 * r^2} \tag{7.2}$$

More recently this erf-gau LDA functional was combined with a standard GGA in an attempt to surpass the accuracy of exchange-hybrid functionals based on erf [324] with

a GGA correlation treatment. Another 1-parameter attenuated Coulomb interaction, the Yukawa potential, has also been the subject of recent investigations [325, 326].

$$v_{ee,\text{yukawa}} = \exp(-\gamma r)/r \tag{7.3}$$

Improved performance of the resulting functional was attributed to an increased fraction of short-range exchange [327].

Our group has recently published an analytical integral over a more general sort of Coulomb attenuator which allows for separate control of where and how rapidly the shift is made between parts of the Coulomb interaction [328]. It allows continuous variation of sharpness between the limits of erf and the Heavyside function. The function is a linear combination of two error functions (although note that the erf formulas aren't sufficient to describe it) and so we have adopted the name "terf":

$$\text{terf}_{r_0,\omega}(r) = (1/2)(\text{erf}(\omega r - r_0) + \text{erf}(\omega r + r_0)) \tag{7.4}$$

Investigations into the performance of range-separated hybrids [329] have found that existing separations cannot simultaneously describe ground-state and excited state properties with a single choice of ω. Along these lines a common area of intersection has been located amongst many optimized attenuators at roughly .8 Bohr [330] in the $(r, V_{optimal}(r))$ plane . The physical implication is that cancellation of GGA-exchange and GGA-correlation errors in this region is balanced with the error induced by semi-local exchange, and our attenuator should be shaped similarly in this region for thermochemical accuracy. Yet to repair self-interaction error the attenuator should reach its asymptote as rapidly as possible once we leave this region. The terf functional form can pass through this point while still reaching its asymptote more rapidly than erf, and so there is reason to hope that terf may be useful in this respect. Another nice feature of this choice of separation is that it reduces to the erf attenuator if the r_0 parameter is chosen to be zero. The attenuator is plotted for several choices of parameters in Figure 1.

7.2 The terf-attenuated LDA

The exchange energy of a many-fermion system, charge balanced by a structureless positive background is our starting point. This matrix element is [331]. (where θ denotes the Heaviside function)

$$E_x = \frac{-k_f^3}{12\pi^4}\int_0^\infty q^2 v_{ee}(q)(1 - \frac{3}{2}x + \frac{1}{2}x^3)\theta(1-x)dq \text{ ; where: } k_f = (3\pi^2 n)^{1/3},\ x = q/(2k_f) \tag{7.5}$$

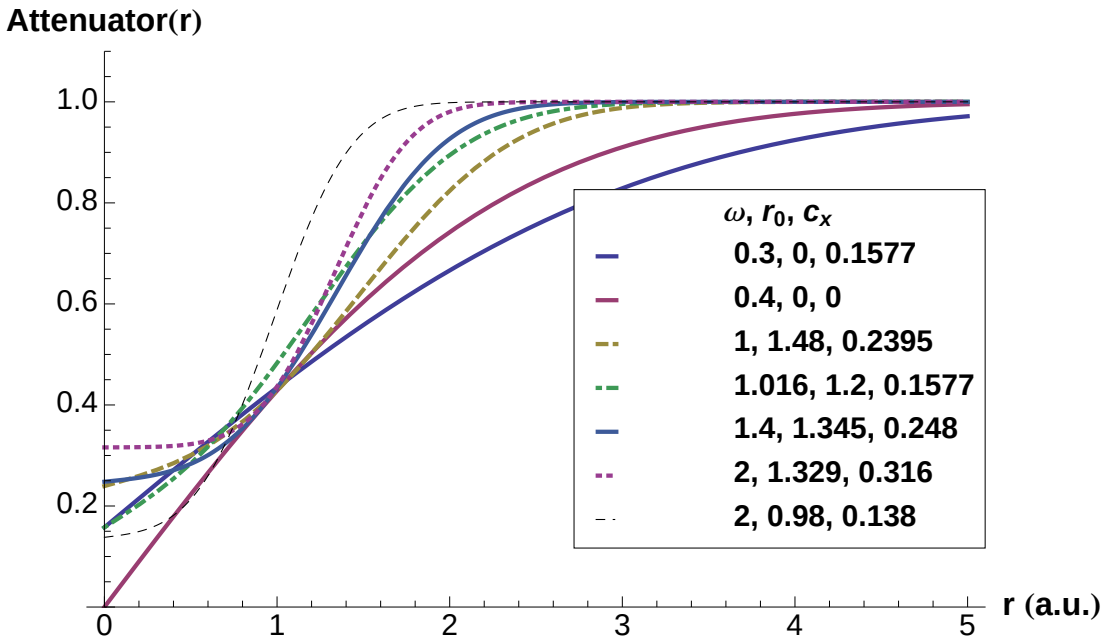


Figure 7.1: Various attenuators plotted for comparison. The first two are equivalent to Erf ($\omega =$ .3, .4).

So we must obtain the Fourier transform of $\text{terfc}_{r_0,\omega}(r)/r$.

$$\mathcal{F}\{\text{terfc}_{r_0,\omega}(r)/r\} = v_{ee}(q) = \frac{4\pi(1 - e^{-\frac{q^2}{4\omega^2}}\cos(\frac{qr_0}{\omega}))}{q^2} \tag{7.6}$$

The integration of this function is algebraically quite tedious, but can be done. Unfortunately the complex error function enters. Note that for $z \in \mathbb{C}, \text{erf}(z^*) = \text{erf}(z)^*$ and for $z \in \mathbb{R}, \text{erfi}(z) \in \mathbb{R}$. We report the exchange energy per particle of the Fermion gas experiencing this interaction, ϵ_{xc} which may be readily implemented in any KS-DFT code.

$$E_x = \int n(R)\epsilon_{xc}(n(R))dR \tag{7.7}$$

The spinless k_f can be easily replaced to obtain the spin-density functional.

$$\epsilon_x^{\omega,r_0}(n) = \frac{\omega^4}{192A^4\pi^3}\Bigg\{ \left(8Ae^{-r_0^2}\sqrt{\pi}\left(Ar_0(3+A^2(4r_0^2-6))(\text{erfi}[r_0]+\text{Im}(\text{erf}(\frac{1}{2A}-ir_0)))+\text{Re}(\text{erf}(\frac{1}{2A}-ir_0)\right)\right) + \left(-(3+24A^2+32A^4(r_0^2-1))+e^{-\frac{1}{4A^2}}16A^2((1+2A^2(r_0^2-1))\cos(\frac{r_0}{A})+Ar_0\sin(\frac{r_0}{A}))\right)\Bigg\}$$

$$\text{where: } A = \frac{\omega}{2k_f} \quad (7.8)$$

One may easily verify that this expression matches the known erf expressions for ϵ_x as $r_0 \to 0$ [313, 332]. For large values of A (>1), a series expansion in powers of $1/A$ is employed up to 10^{th} order in our implementation for purposes of numerical stability.

7.3 Application to Range Separated Hybrids

Without semi-local gradient information the thermochemistry of this functional would be undoubtedly poor and it would be difficult to determine if terf could improve functionals in use today. There are several recipes for combining this LDA exchange functional with a GGA enhancement factor ranging in degrees of technical difficulty and empiricism. Ideally the GGA factor will depend on the attenuating parameters [333], but recent results have shown that superior accuracy [83] can be obtained even if this is only done implicitly through optimized parameters of the GGA. At the end of the day the choice of ω parameter is quite empirical, as will be r_0, even if we introduce them for physical reasons. The final measure of a range-separated hybrid is optimization over a large training set, and evaluation over an independent test set, roughly a year of computer effort. We seek some justification for such an effort and so we combine the terfc-LDA exchange energy with the GGA exchange enhancement factor and correlation functionals of $\omega B97X$ [83] and run some basic tests to establish whether the resulting functional shows promise. To be precise, the resulting functional is obtained directly from ωB97X by replacing the $F(a_\sigma)$ of equation (7) in that paper with the corresponding terfc-LDA $F(a_\sigma)$ obtained from equation (8) above. To begin from a functional as close as possible to the parent (see the previous paper to clarify the notation), we also incorporate a variable fraction of short-range HF exchange in such a way that the UEG limit is respected (Eqns 9, 10).

$$E_x^{SR-DFA} = \sum_\sigma \int e_{x\sigma}^{terfc-LSDA}(\rho_\sigma) g_{x\sigma}^{\omega B97X}(s_\sigma^2) dr \quad (7.9)$$

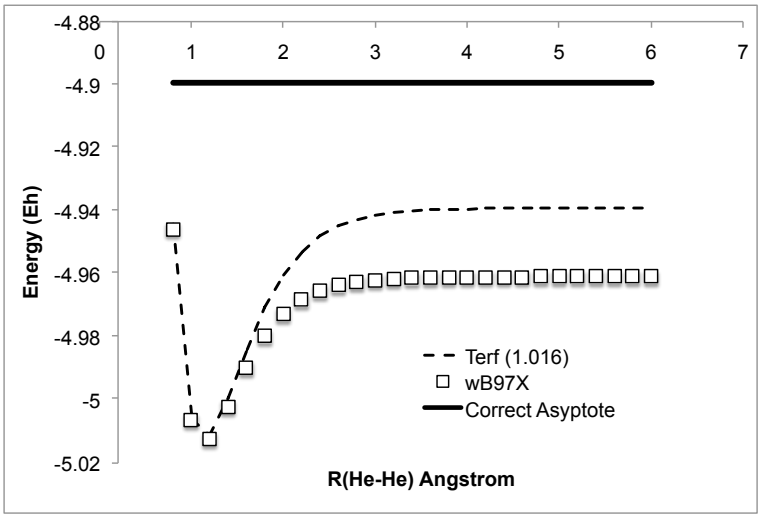


Figure 7.2: Dissociation of He_2^+

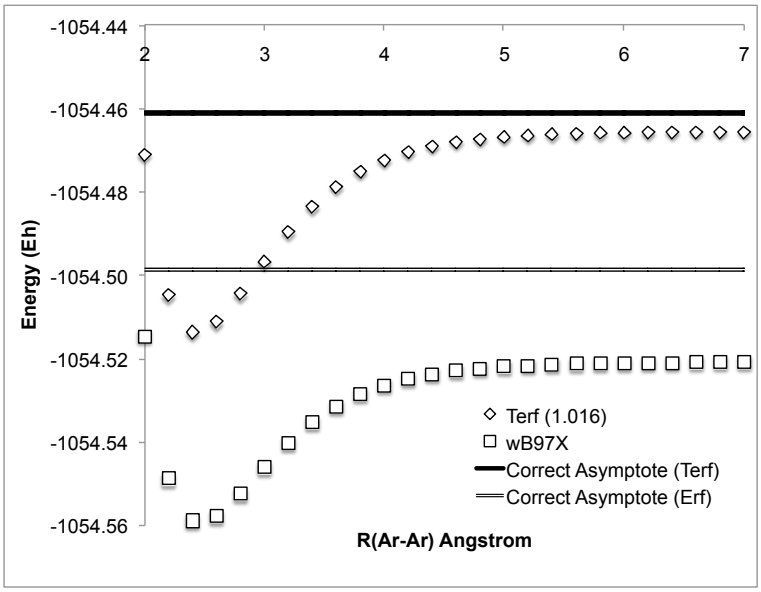


Figure 7.3: Dissociation of Ar_2^+

$$E_{xc} = E_x^{LR-HF} + c_x E_x^{SR-HF} + (1 - c_x) E_x^{SR-DFA} + E_c^{\omega B97X} \tag{7.10}$$

Aside from the many parameters associated with the GGA we must choose reasonable guesses of $\{r_0, \omega, c_x\}$. The physically motivated guess is to reach the asymptote as rapidly as possible while still overlapping significantly with the established attenuators in

ω (a.u.)	r_0 (a.u.)	c_x	MAE	He_2^+ Error	Ne_2^+ Error	Ar_2^+ Error	Kr_2^+ Error
BLYP†	-	-	-	-87.55	-80.37	-48.51	-49.0
0.3*	0	0.1577	2.09	-38.9	-34.4	-14.0	-11.2
0.4**	0	0	2.53	-35.1	-33.3	-9.7	-7.7
1	1.48	0.2395	5.71	-31.7	-29.3	-6.7	-4.2
1.016	1.2	0.1577	4.02	-24.9	-24.7	-3.0	-1.5
1.4	1.345	0.248	5.77	-26.6	-26.2	-2.9	-1.2
2	1.329	0.316	9.16	-23.2	-21.9	-0.4	0.0
2	0.98	0.138	7.38	-13.0	-16.2	-.03	0.0

Table 7.1: Mean absolute error (kcal/mol) of G2 set atomization energies and errors of dimer cation asymptotes for various functionals. *(ωB97X), **(ωB97) †(pure Becke 88 exchange [2] and LYP [3] correlation, errors in this row are upper bounds.)

the previously mentioned critical region [330]. An initial choice of parameters $\{r_0, \omega, c_x\} = \{1.2, 1.016, c_x^{\omega B97X}\}$ was made by this physically motivated criterion, and the usefulness of the resulting functional was assessed on noble gas dimer-cation dissociation (Figures 3,4). Even with only very conservative changes made to the functional form of the attenuator, terf was able to significantly increase the accuracy (relative to its predecessor ωB97X) of the dissociation asymptote associated with the self-interaction problem (SIE) (Figure 4). In the case of He_2^+ the valence density lies so close to the neighboring atom that it seems challenging to reach a compromise between thermochemistry and exact exchange within a transferable range-separated exchange functional but the results for Ar_2^+ are encouraging. The solid thermochemical performance of the parent functional seemed more-or-less conserved, and so we investigated a little further. We make the approximation that the GGA parameters are unchanged between Erf (ωB97X) and Terf attenuated coulomb interactions. Undoubtedly this should be improved upon in future work and the literature already describes many ways this may be done.

A few more sets of attenuator parameters were obtained by maximizing the least-squared overlap of a terf attenuator with those of ωB97 and ωB97X varying r_0, c_x for a given ω. Noble gas dissociation curves and atomization energies were calculated for the standard G2 [334] thermochemical test set in the 6-311++G(3df,3pd) basis with a saturated quadrature grid. The purpose was not to search for parameters which would surpass ωB97X, because gradient corrections are vital to thermochemical accuracy, but rather to document the balance between thermochemistry and correction of the self-interaction error (Table 1). As expected the results were quite sensitive to the steepness of the attenuator and the amount of middle-range exchange, but note that the further this attenuator departs from Erf the more severe becomes the GGA approximation. An accurate asymptote cannot be obtained by simply increasing the amount of exact exchange in the attenuator

(because eventually the correlation part of the problem is disturbed), but by maximizing overlap with the established ones in the critical region we do obtain reduction of SIE with increasing "exact" exchange. At the moment where the singly-occupied MO's bond is breaking if this MO's density around atom 2 is far enough from the bulk of this MO's density around atom 1 so that it experiences Hartree exchange the density will localize on an atom (as it should physically). Smaller atoms cause greater difficulty in general depending specifically on shell structure. The results suggest that if the GGA enhancement factor of the functional were polished it would possess thermochemistry much like ωB97X with a significantly larger amount of exact exchange, and even in its current incarnation is accurate enough to be used in lieu of others for problems where "exact" exchange might be important.

7.4 Discussion and Conclusions

Owing largely to the popularity of hybrid functionals range separation of exchange has become an intense area of research, and more flexible range separation may prove desirable [311]. Indeed this has already been done with the erf-gau type attenuator [324], although in this case this was done at the expense of abandoning a physically motivated choice of a parameter. An analytical formula [328] is available for the exact exchange energy with the terf attenuation, and this expression has already been efficiently implemented in the publicly available release of the Q-Chem package [192]. This paper provides the other building block, the short-range LDA exchange energy and a proof-of-concept GGA functional. Preliminary results with unoptimized parameters indicate that the new functional may be a useful improvement and development should continue in the area of more general exchange attenuators. Further improvement over the functional developed here might realized through complete reoptimization [83]. Alternatively one could derive the corresponding PBE type GGA-functional [333,335,336]. In either case, the path is clear and only limited by one's curiosity. It will be especially interesting to see if the flexibility of the new attenuator can simultaneously describe ground state electronic structure and excited states. Special attention should be paid to the size of the chromophore relative to the scale of the attenuator, and the distance over which electron density is redistributed. This direction is currently being pursued in our group.

Chapter 8

Conclusions & Outlook

8.1 Summary

In this work we have presented a convergent hierarchy of approximations to the CASSCF method for strong valence correlations and the extension of those methods to a models of the total electronic energy. The cost of these methods grows with accuracy and system size in a tractable, polynomial fashion and systems much larger than those which can be tackled with CASSCF can be addressed already with the implementations developed here. The most accurate choices in this menu afford chemical accuracy for the test cases we have examined. We have also presented a density functional with reduced self-interaction error.

If a chemist is interested in a problem of multiple broken bonds or otherwise uncertain about how to assign a certain number of formal bonds a typical approach with existing technology may have been difficult. One of these local, high-rank cluster approximations could provide significant accuracy and insight. Because they are well-defined for a given chemical situation it is obvious how one should proceed. The user can begin with the most affordable option, choose the well-defined and reliable valence active space, and proceed until convergence or exhaustion of computational resources.

As said on the first page of this thesis, the purpose of this work, and the substance of the insights are to express the wave-function with less and less complexity. In that respect we have indeed made a little progress. With a 5^{th} order number of variables we can offer a very accurate and reliable model for chemistry. A companion problem is to efficiently calculate that model on computers which we have in hand. We have demonstrated that it is possible to do so, but the computer science which would make it as efficient as implementations of DFT is an open question.

8.2 Future Research Directions

The models we have presented are so complex that they could not be perfected in the duration of a single Ph.D. and there are several points of entry where they could be improved significantly and rapidly. The directions of improvement are both physical and algorithmic. In both cases new and interesting systems could be calculated with new levels of accuracy and it seems likely that these directions will be pursued soon.

Excited States

One clear physical shortcoming is that these models have been developed with only the electronic ground state in mind, but excited states need even more strong correlation and are poorly understood. A starting point would be to exploit the CC underpinnings of these models and take an Equation-of-Motion approach, diagonalizing the similarity-transformed Hamiltonian for excited states. Such an approach would have limited accuracy because the localized orbitals (which have been optimized to fit the ground state) are not equally appropriate for the target excited states. In order to achieve an even-tempered treatment, the response of the orbitals to excitation must be included. The formalism for these responses could follow the example of Linear-Response MCSCF [337].

Efficient Dynamical Correlation

+SD was presented as an iterative correction to an active space correlation model, however it's likely that a perturbative correction would provide intermediate accuracy at a significantly smaller cost, and with more easily optimized code. Just as MP2 arises naturally as the first iteration of CCSD, the +SD correction may be calculated perturbatively as the first iteration of it's iterative counterpart. In most applications that correlation energy lies between the active-space reference and the converged energy. The success of a perturbative model relies on a cancellation of errors. Before a specialized code is developed for such a model, this cancellation of errors should be carefully evaluated over several model systems.

Density Functionals

Range-separated density functionals have emerged as a near systematic improvement over global hybrids like B3LYP because they repair a physical deficiency of the base functional, but the more general *terf*-separated version would likely offer even better performance. To develop such a hybrid with the code currently available one should begin with the GGA optimized for ωB97X, and develop a test set which includes several difficult exchange problems then perform a simultaneous optimization of both GGA and attenuator parameters. Isomerizations of branched, saturated alkanes are emerging [338] as one such

class of difficult problems, which might otherwise seem perfectly well-suited for DFT.

Algorithmic Improvments

The code itself limits the scope of possible applications and so I'll outline the places where it can be most rapidly improved. It is difficult to appreciate how significantly the performance of a program can vary based on memory access patterns. To deliver codes which scale properly with system size we introduce significant logical overhead to avoid calculating the parts of the correlation problem we truncate. In the codes developed for this work that logical overhead always takes the shape of just a few operations: lexical sorting of an array of integers, ripping stripes of integers from arrays, and collecting those integers into tagged blocks. Although some attention has been paid to each operation, none of the relevant routines even approach the optimal operation count of a modern processor because a significant amount of time is spent waiting for information from distant regions of memory. Careful computer science on each operation would recover significant performance boosts.

A spin-orbital formalism has been employed, although the target systems usually contain an equal number of α and β electrons. The symmetry between these two sorts of spin-orbitals is not exploited in the code (ie: each $\alpha\alpha$ contribution is calculated separately from $\beta\beta$ even though these are the same). In general each amplitude has an identical spin-flip partner which is redundantly calculated, but should be omitted. If it were, storage costs would immediately be halved.

The electron correlation models we have presented are composed of literally hundreds of nonlinear terms, however there is a significant degree of repetitious effort which can be avoided by proper factorization. For example each term in the coupled-cluster multiplier residual originates in a term of the amplitude equations, and if that amplitude residual term is stored, we can avoid repeating that calculation, algebraically:

$$\hat{V}\hat{T}^n = \hat{V}\hat{T}^{n-1}\hat{T} \rightarrow \Lambda\hat{V}\hat{T}^{n-1} \tag{8.1}$$

$$\text{Save and reuse: } \hat{V}\hat{T}^{n-1} \tag{8.2}$$

This algorithmic improvement would reduce the cost of the gradient by roughly a factor of two.

Self consistent orbitals require repeated integral transformation which costs 5^{th} order, the same exponent governing the cost of the correlation calculation itself. However by employing a tensor decomposition approximation like the resolution of the identity [127, 222] the pre-factor on the transformation operation can be reduced by a factor of more than a hundred. Algebraically implementing this change in the code would be accomplished by replacing the integral in each term with the contraction of two tensors over an auxiliary basis.

Bibliography

[1] T. Van Voorhis and M. Head-Gordon, The Journal of Chemical Physics **117**, 9190 (2002).

[2] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[3] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[4] M. Planck, Archives Neerlandaises des Sciences Exactes et Naturelles , 164 (1900).

[5] M. Planck, Archives Neerlandaises des Sciences Exactes et Naturelles , 55 (1901).

[6] A. Einstein, Annalen Der Physik **1**, 180 (1906).

[7] M. Planck, Annalen Der Physik **6**, 818 (1906).

[8] L. De Broglie, *Research on Quantum Theory*, 1924.

[9] A. Kramers and W. Heisienberg, Zeitschrift Fur Physik **31**, 681 (1925).

[10] W. Heisienberg, Zeitschrift Fur Physik **31**, 617 (1925).

[11] W. Pauli, Zeitschrift Fur Physik **31**, 765 (1925).

[12] E. Schrödinger, Annalen Der Physik **77**, 43 (1925).

[13] W. Heisienberg, Zeitschrift Fur Physik **35**, 557 (1926).

[14] E. Schrödinger, Annalen Der Physik **79**, 361 (1926).

[15] E. Schrödinger, Annalen Der Physik **79**, 489 (1926).

[16] E. Schrödinger, Annalen Der Physik **79**, 734 (1926).

[17] M. Born and R. Oppenheimer, Annalen Der Physik **84**, 457 (1927).

[18] P. Dirac, Proc. Royal Soc. London , 621 (1927).

[19] D. Hartree, Proc. Cambridge Phil. Soc. **24**, 89 (1928).

[20] D. Hartree, Proc. Cambridge Phil. Soc. **24**, 111 (1928).

[21] D. Hartree, Proc. Cambridge Phil. Soc. **24**, 426 (1928).

[22] V. Fock, Comp. Rend. de l'Ac. Sci. de l'URSS **8**, 295 (1935).

[23] J. Slater, Phys. Rev. **32**, 339 (1928).

[24] J. Slater, Phys. Rev. **35**, 210 (1928).

[25] J. C. Slater, Phys. Rev. **81**, 385 (1951).

[26] R. P. Feynman, Phys. Rev. **76**, 769 (1949).

[27] G. C. Wick, Phys. Rev. **80**, 268 (1950).

[28] J. A. Pople and R. K. Nesbet, J. Chem. Phys. **22**, 571 (1954).

[29] J. Pople and R. Nesbet, J. Chem. Phys. **22**, 571 (1954).

[30] J. A. Pople and R. K. Nesbet, J. Chem. Phys. **22**, 571 (1954).

[31] C. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[32] A. T. Amos and G. G. Hall, Proc. Royal Soc. London , 483 (1961).

[33] P. Lö wdin, Adv.Chem. Phys. **2**, 207 (1959).

[34] R. Harrison and N. Handy, Chem. Phys. Lett. **95**, 386 (1983).

[35] Y. Hatano and K. Hirao, Chem. Phys. Lett. **100**, 519 (1983).

[36] P. Knowles and N. Handy, Chem. Phys. Lett. **111**, 315 (1984).

[37] P. Knowles and N. Handy, Chem. Phys. Lett. **111**, 315 (1984).

[38] C. Bauschlicher and P. Taylor, J. Chem. Phys. **85**, 2779 (1986).

[39] C. Bauschlicher and P. Taylor, Theo. Chem. Act. **71**, 263 (1987).

[40] P. Knowles and N. Handy, Comp. Phys. Commun. **54**, 75 (1989).

[41] C. Lanczös, J. Res. Natl. Bur. Stand. **45** (1950).

[42] E. R. Davidson, Journal of Computational Physics **17**, 87 (1975).

[43] L. M. Falicov and R. A. Harris, The Journal of Chemical Physics **51**, 3153 (1969).

[44] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[45] P.-O. Lowdin, Journal of Mathematical Physics **3**, 969 (1962).

[46] B. Roos, Chemical Physics **48**, 157 (1980).

[47] B. O. Roos, P. Linse, P. E. M. Siegbahn, and M. R. A. Bloomberg, Chem. Phys. **66**, 197 (1981).

[48] H.-J. Werner and W. Meyer, The Journal of Chemical Physics **74**, 5794 (1981).

[49] J. Čí žek, J. Chem. Phys. **45**, 4256 (1966).

[50] J. Cizek, The Journal of Chemical Physics **45**, 4256 (1966).

[51] R. J. Bartlett and G. D. Purvis, Int. J. Quantum Chem. **14**, 561 (1978).

[52] R. J. Bartlett and M. Musial, Reviews of Modern Physics **79**, 291 (2007).

[53] B. D. Day and R. B. Wiringa, Phys. Rev. C **32**, 1057 (1985).

[54] T. V. Voorhis and M. Head-Gordon, The Journal of Chemical Physics **113**, 8873 (2000).

[55] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chemical Physics Letters **157**, 479 (1989).

[56] C. Hampel and H.-J. Werner, The Journal of Chemical Physics **104**, 6286 (1996).

[57] J. E. Subotnik, A. Sodt, and M. Head-Gordon, The Journal of Chemical Physics **125**, 074116 (2006).

[58] J. F. Stanton, The Journal of Chemical Physics **99**, 8840 (1993).

[59] S. Hirata, The Journal of Chemical Physics **121**, 51 (2004).

[60] A. C. Hurley, J. Lennard-Jones, and J. A. Pople, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **220**, 446 (1953).

[61] W. A. Goddard and L. A. Harding, Annu. Rev. Phys. Chem. **29**, 363 (1978).

[62] A. F. Voter and W. A. Goddard, J. Chem. Phys. **75**, 3638 (1981).

[63] J. Cullen, Chemical Physics **202**, 217 (1996).

[64] J. Cullen, Journal of Computational Chemistry **20**, 999 (1999).

[65] T. V. Voorhis and M. Head-Gordon, Chemical Physics Letters **317**, 575 (2000).

[66] T. V. Voorhis and M. Head-Gordon, J. Chem. Phys. **117**, 9190 (2002).

[67] T. V. Voorhis and M. Head-Gordon, The Journal of Chemical Physics **117**, 9190 (2002).

[68] T. V. Voorhis and M. Head-Gordon, Journal of Chemical Physics **115**, 7814 (2001).

[69] K. V. Lawler, G. J. O. Beran, and M. Head-Gordon, J. Chem. Phys. **128**, 024107 (2008).

[70] K. V. Lawler, J. A. Parkhill, and M. Head-Gordon, Mol. Phys. **106**, 2309 (2008).

[71] T. V. Voorhis and M. Head-Gordon, Chemical Physics Letters **330**, 585 (2000).

[72] E. Byrd, T. Van Voorhis, and M. Head-Gordon, Journal of Physical Chemistry B **106**, 8070 (2002).

[73] S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon, and A. I. Krylov, The Journal of Chemical Physics **113**, 3548 (2000).

[74] S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon, and A. I. Krylov, The Journal of Chemical Physics **113**, 3548 (2000).

[75] S. R. Gwaltney and M. Head-Gordon, The Journal of Chemical Physics **115**, 2014 (2001).

[76] S. R. Gwaltney, E. F. C. Byrd, T. Van Voorhis, and M. Head-Gordon, Chem. Phys. Lett. **353**, 359 (2002).

[77] G. J. O. Beran, S. R. Gwaltney, and M. Head-Gordon, Physical Chemistry Chemical Physics **5**, 2488 (2003).

[78] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, B864 (1964).

[79] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).

[80] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[81] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).

[82] A. D. Becke, The Journal of Chemical Physics **107**, 8554 (1997).

[83] J.-D. Chai and M. Head-Gordon, The Journal of Chemical Physics **128**, 084106 (2008).

[84] J. A. Parkhill and M. Head-Gordon, Molecular Physics **108**, 513 (2010).

[85] K. V. Lawler, J. A. Parkhill, and M. Head-Gordon, The Journal of Chemical Physics **130**, 184113 (2009).

[86] J. A. Parkhill, K. Lawler, and M. Head-Gordon, The Journal of Chemical Physics **130**, 084101 (2009).

[87] J. A. Parkhill, J.-D. Chai, and M. Head-Gordon, Chemical Physics Letters **478**, 283 (2009).

[88] C. L. Janssen and H. F. Schaefer, Theoretical Chemistry Accounts **79**, 1 (1990).

[89] P. J. Knowles and N. C. Handy, Chemical Physics Letters **111**, 315 (1984).

[90] J. Olsen, B. Roos, P. Jørgensen, and H. Jensen, J. Chem. Phys. **89**, 2185 (1988).

[91] P. Knowles, K. Somasundram, N. Handy, and K. Hirao, Chemical Physics Letters **113**, 8 (1985).

[92] S. Hirata and R. J. Bartlett, Chemical Physics Letters **321**, 216 (2000).

[93] J. Olsen, The Journal of Chemical Physics **113**, 7140 (2000).

[94] M. Kallay and P. R. Surjan, The Journal of Chemical Physics **113**, 1359 (2000).

[95] R. J. Bartlett, Annual Review of Physical Chemistry **32**, 359 (1981).

[96] N. Oliphant and L. Adamowicz, The Journal of Chemical Physics **95**, 6645 (1991).

[97] S. A. Kucharski and R. J. Bartlett, The Journal of Chemical Physics **97**, 4282 (1992).

[98] M. Kallay, P. Szalay, and P. Surjan, J. Chem. Phys. **117**, 980 (2002).

[99] G. Baumgartner et al., A high-level approach to synthesis of high-performance codes for quantum chemistry, in *Supercomputing '02: Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pages 1–10, Los Alamitos, CA, USA, 2002, IEEE Computer Society Press.

[100] D. I. Lyakh, V. V. Ivanov, and L. Adamowicz, The Journal of Chemical Physics **122**, 024108 (2005).

[101] J. A. Parkhill, K. Lawler, and M. Head-Gordon, J. Chem. Phys. **130**, 084101 (2009).

[102] T. Shiozaki, M. Kamiya, S. Hirata, and E. Valeev, Physical Chemistry Chemical Physics **10**, 3358 (2008).

[103] S. Hirata, P.-D. Fan, A. A. Auer, M. Nooijen, and P. Piecuch, The Journal of Chemical Physics **121**, 12197 (2004).

[104] E. Neuscamman, T. Yanai, and G. K.-L. Chan, The Journal of Chemical Physics **130**, 124102 (2009).

[105] C. Bauer and H. S. Do, Comput. Phys. Commun. **144** (2002).

[106] J. Noga and R. J. Bartlett, J. Chem. Phys. **86**, 7041 (1987).

[107] D. Cociorva et al., *Towards Automatic Synthesis of High-Performance Codes for Electronic Structure Calculations: Data Locality Optimization*, Number 237-248 in High Performance Computing — HiPC 2001, Springer Berlin / Heidelberg, 2001.

[108] S. Hirata, The Journal of Physical Chemistry A **107**, 9887 (2003).

[109] A. D. Bochevarov and C. D. Sherrill, The Journal of Chemical Physics **121**, 3374 (2004).

[110] D. Crawford and H. F. S. III, Reviews in Computational Chemistry , 33 (2007).

[111] R. D. Mattuck, *A guide to Feynman diagrams in the many-body problem*, Dover, 1992.

[112] M. Kallay and P. R. Surjan, The Journal of Chemical Physics **115**, 2945 (2001).

[113] J. Gauss, J. F. Stanton, and R. J. Bartlett, The Journal of Chemical Physics **95**, 2623 (1991).

[114] J. Gauss and J. F. Stanton, The Journal of Chemical Physics **116**, 1773 (2002).

[115] T. V. Voorhis and M. Head-Gordon, The Journal of Chemical Physics **113**, 8873 (2000).

[116] K. Kowalski and P. Piecuch, The Journal of Chemical Physics **113**, 18 (2000).

[117] Z. Rolik, Á. Szabados, D. Kohalmi, and P. Surján, Journal of Molecular Structure: THEOCHEM **768**, 17 (2006), Coupled-cluster Methods: Theory and Applications. A Collection of Invited Papers in Honor of Debashis Mukherjee on the Occasion of his 60th Birthday.

[118] M. Kallay and J. Gauss, The Journal of Chemical Physics **120**, 6841 (2004).

[119] J. Gauss and J. F. Stanton, The Journal of Chemical Physics **116**, 1773 (2002).

[120] A. A. Auer et al., Molecular Physics **104**, 211 (2006).

[121] B. W. Bader and T. G. Kolda, SIAM Journal on Scientific Computing **30**, 205 (2007).

[122] S. Saebo and P. Pulay, Annual Review of Physical Chemistry **44**, 213 (1993).

[123] J. E. Subotnik, A. Sodt, and M. Head-Gordon, The Journal of Chemical Physics **128**, 034103 (2008).

[124] J. Cullen, The Journal of Chemical Physics **202**, 217 (1996).

[125] G. J. O. Beran and M. Head-Gordon, Molecular Physics **104**, 1191 (2006).

[126] K. V. Lawler, J. A. Parkhill, and M. Head-Gordon, Molecular Physics **106**, 2309 (2008).

[127] R. DiStasio, Y. Jung, and M. Head-Gordon, Journal of Chemical Theory and Computation **1**, 862 (2005).

[128] B. Gedik, R. R. Bordawekar, and P. S. Yu, Cellsort: high performance sorting on the cell processor, in *VLDB '07: Proceedings of the 33rd international conference on Very large data bases*, pages 1286–1297, VLDB Endowment, 2007.

[129] B. He, N. K. Govindaraju, Q. Luo, and B. Smith, Efficient gather and scatter operations on graphics processors, in *SC '07: Proceedings of the 2007 ACM/IEEE conference on Supercomputing*, pages 1–12, New York, NY, USA, 2007, ACM.

[130] H.-J. Werner, The Journal of Chemical Physics **129**, 101103 (2008).

[131] J. Kussmann and C. Ochsenfeld, The Journal of Chemical Physics **128**, 134104 (2008).

[132] D. Manocha, Computer **38**, 85 (2005).

[133] W. Kutzelnigg and D. Mukherjee, The Journal of Chemical Physics **107**, 432 (1997).

[134] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1253 (2000).

[135] T. G. Kolda, SIAM Journal on Matrix Analysis and Applications **23**, 243 (2001).

[136] M. A. O. Vasilescu and D. Terzopoulos, Tensortextures: multilinear image-based rendering, in *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 336–342, New York, NY, USA, 2004, ACM.

[137] G. J. O. Beran, B. Austin, A. Sodt, and M. Head-Gordon, J. Phys. Chem. A **109**, 9183 (2005).

[138] A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon, J. Chem. Phys. **109**, 10669 (1998).

[139] T. Živković and H. Monkhorst, J. Math Phys. **19**, 1007 (1978).

[140] K. Jankowski and J. Paldus, Int. J. Quantum Chem. **18**, 1243 (1980).

[141] P. Piecuch, S. Zarrabian, J. Paldus, and J. Čížek, Phys. Rev. B **42**, 3351 (1990).

[142] J. Paldus, P. Piecuch, L. Pylypow, and B. Jeziorski, Phys. Rev. A **47**, 2738 (1993).

[143] K. Kowalski and K. Jankowski, Phy. Rev. Lett. **81**, 1195 (1998).

[144] K. Jankowski and K. Kowalski, J. Chem. Phys. **111**, 2940 (1999).

[145] K. Jankowski and K. Kowalski, J. Chem. Phys. **111**, 2952 (1999).

[146] K. Kowalski and P. Piecuch, Phys. Rev. A **61**, 052506 (2000).

[147] P. Szakács and P. R. Surján, Int. J. Quantum Chem. **108**, 2043 (2008).

[148] T. P. Hamilton and P. Pulay, J. Chem. Phys. **84**, 5728 (1986).

[149] T. Korona and H.-J. Werner, J. Chem. Phys. **118**, 3006 (2003).

[150] J. D. Watts and R. J. Bartlett, J. Chem. Phys. **101**, 3073 (1994).

[151] H. K. O. Christiansen and P. Jorgensen, Chem. Phys. Lett. **243**, 409 (1995).

[152] M. Kallay and J. Gauss, J. Chem. Phys. **121**, 9257 (2004).

[153] S. Hirata, J. Chem. Phys. **121**, 51 (2004).

[154] J. M. Langlois et al., J. Chem. Phys. **92**, 7488 (1990).

[155] A. Sodt, G. J. O. Beran, Y. S. Jung, B. Austin, and M. Head-Gordon, J. Chem. Theory Comput. **2**, 300 (2006).

[156] K. Kowalski and P.-D. Fan, J. Chem. Phys. **130**, 084112 (2009).

[157] A. P. Morgan, A. J. Sommese, and C. W. Wampler, Numer. Math. **58**, 669 (1991).

[158] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[159] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).

[160] X. Li and J. Paldus, J. Chem. Phys. **130**, 084110 (2009).

[161] I. Mayer, Int. J. Quantum Chem. **63**, 31 (1997).

[162] H.-J. Werner, F. R. Manby, and P. J. Knowles, The Journal of Chemical Physics **118**, 8149 (2003).

[163] T. D. Crawford and R. A. King, Chemical Physics Letters **366**, 611 (2002).

[164] M. Schutz, Physical Chemistry Chemical Physics **4**, 3941 (2002).

[165] H. D. and R. M. A, in *Molecular Physics*, volume 38, page 1795, 1979.

[166] H. Nakatsuji, The Journal of Chemical Physics **113**, 2949 (2000).

[167] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, Journal of Physical Chemistry **94**, 5483 (1990).

[168] J. Paldus, P. Piecuch, L. Pylypow, and B. Jeziorski, Phys. Rev. A **47**, 2738 (1993).

[169] U. S. Mahapatra, B. Datta, and D. Mukherjee, The Journal of Chemical Physics **110**, 6171 (1999).

[170] I. Hubac and S. Wilson, Journal of Physics B: Atomic, Molecular and Optical Physics **34**, 4259 (2001).

[171] P. Piecuch, N. Oliphant, and L. Adamowicz, The Journal of Chemical Physics **99**, 1875 (1993).

[172] P. Piecuch and L. Adamowicz, The Journal of Chemical Physics **100**, 5792 (1994).

[173] G. Chan and M. Head-Gordon, Journal of Chemical Physics **116**, 4462 (2002).

[174] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[175] A. E. DePrince, III, E. Kamarchik, and D. A. Mazziotti, Journal of Chemical Physics **128**, 234103 (2008).

[176] J. H. Sebold and J. K. Percus, The Journal of Chemical Physics **104**, 6606 (1996).

[177] D. A. Mazziotti, The Journal of Chemical Physics **115**, 8305 (2001).

[178] F. Faglioni and W. Goddard, International Journal of Quantum Chemistry **73**, 1 (1999).

[179] W. J. Hunt, P. J. Hay, and W. A. Goddard, The Journal of Chemical Physics **57**, 738 (1972).

[180] V. A. Rassolov and F. Xu, The Journal of Chemical Physics **127**, 044104 (2007).

[181] A. I. Krylov, Chemical Physics Letters **350**, 522 (2001).

[182] L. V. Slipchenko and A. I. Krylov, The Journal of Chemical Physics **117**, 4694 (2002).

[183] D. Casanova and M. Head-Gordon, The Journal of Chemical Physics **129**, 064104 (2008).

[184] P. Pulay and S. Saebø, Theoretica Chimica Acta **69**, 357 (1986).

[185] W. Forner, J. Ladik, P. Otto, and J. Cizek, Chemical Physics **97**, 251 (1985).

[186] D. C. Chiles R.A., Journal of Chemical Physics **74**, 4544 (1981).

[187] H. Nakatsuji, T. Miyahara, and R. Fukuda, The Journal of Chemical Physics **126**, 084104 (2007).

[188] A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon, Journal of Chemical Physics **109**, 10669 (1998).

[189] G. Beran, B. Austin, A. Sodt, and M. Head-Gordon, Journal of Physical Chemistry A. **109**, 9183 (2005).

[190] K. V. Lawler, G. J. O. Beran, and M. Head-Gordon, Journal of Chemical Physics **128** (2008).

[191] T. V. Voorhis and M. Head-Gordon, The Journal of Chemical Physics **117**, 9190 (2002).

[192] Y. Shao et al., Phys. Chem. Chem. Phys. **8**, 3172 (2006).

[193] P. Pulay, Journal of Computational Chemistry **3**, 556 (1982).

[194] F. Aquilante et al., The Journal of Chemical Physics **129**, 024113 (2008).

[195] M.W.Schmidt et al., Journal of Computational Chemistry **14**, 1347 (1993).

[196] K. Jankowski and J. Paldus, International Journal of Quantum Chemistry **18**, 1243 (1980).

[197] M. Schutz and H.-J. Werner, Chemical Physics Letters **318**, 370 (2000).

[198] P. E. Maslen, M. S. Lee, and M. Head-Gordon, Chemical Physics Letters **319**, 205 (2000).

[199] G. K.-L. Chan, M. Kallay, and J. Gauss, The Journal of Chemical Physics **121**, 6110 (2004).

[200] X. Li and J. Paldus, Chemical Physics Letters **286**, 145 (1998).

[201] F. A. Evangelista et al., The Journal of Chemical Physics **128**, 124104 (2008).

[202] G. Gidofalvi and D. A. Mazziotti, The Journal of Chemical Physics **129**, 134108 (2008).

[203] U. Schollwock, Reviews of Modern Physics **77**, 259 (2005).

[204] T. Yanai and G. K.-L. Chan, The Journal of Chemical Physics **127**, 104107 (2007).

[205] D. Small and M. Head-Gordon, Journal of Chemical Physics ((submitted 2008)).

[206] D. Casanova, L. V. Slipchenko, A. I. Krylov, and M. Head-Gordon, The Journal of Chemical Physics **130**, 044103 (2009).

[207] D. Hegarty and M. A. Robb, Molecular Physics **38**, 1795 (1979).

[208] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, Chemical Physics **71**, 41 (1982).

[209] M. Dupuis, C. Murray, and E. R. Davidson, Journal of the American Chemical Society **113**, 9756 (1991).

[210] V. N. Staroverov and E. R. Davidson, Journal of Molecular Structure: THEOCHEM **573**, 81 (2001).

[211] M. J. McGuire and P. Piecuch, Journal of the American Chemical Society **127**, 2608 (2005).

[212] J. Huang and M. Kertesz, Journal of the American Chemical Society **128**, 7277 (2006).

[213] M. Nakata et al., The Journal of Chemical Physics **114**, 8282 (2001).

[214] R. Pollet, A. Savin, T. Leininger, and H. Stoll, The Journal of Chemical Physics **116**, 1250 (2002).

[215] L. Greenman and D. A. Mazziotti, The Journal of Chemical Physics **130**, 184101 (2009).

[216] D. Casanova, L. V. Slipchenko, A. I. Krylov, and M. Head-Gordon, The Journal of Chemical Physics **130**, 044103 (2009).

[217] X. Li and J. Paldus, The Journal of Chemical Physics **107**, 6257 (1997).

[218] D. W. Small and M. Head-Gordon, The Journal of Chemical Physics **130**, 084103 (2009).

[219] D. A. Mazziotti, Phys. Rev. Lett. **93**, 213001 (2004).

[220] T. A. Ruden, T. Helgaker, P. Jørgensen, and J. Olsen, Chemical Physics Letters **371**, 62 (2003).

[221] T. Shiozaki, K. Hirao, and S. Hirata, The Journal of Chemical Physics **126**, 244106 (2007).

[222] A. Sodt, G. J. O. Beran, Y. Jung, B. Austin, and M. Head-Gordon, Journal of Chemical Theory and Computation **2**, 300 (2006).

[223] H. A. Witek, H. Nakano, and K. Hirao, The Journal of Chemical Physics **118**, 8197 (2003).

[224] B. Dunietz, R. R. B. Murphy, and R. A. Friesner, Journal of Chemical Physics **110**, 1921 (1999).

[225] A. E. DePrince and D. A. Mazziotti, The Journal of Chemical Physics **127**, 104104 (2007).

[226] D. A. Mazziotti, The Journal of Physical Chemistry A **112**, 13684 (2008).

[227] E. Neuscamman, T. Yanai, and G. K.-L. Chan, The Journal of Chemical Physics **130**, 124102 (2009).

[228] D. A. Mazziotti, Physical Review Letters **97**, 143002 (2006).

[229] D. A. Mazziotti, Physical Review A (Atomic, Molecular, and Optical Physics) **76**, 052502 (2007).

[230] G. J. O. Beran, M. Head-Gordon, and S. R. Gwaltney, The Journal of Chemical Physics **124**, 114107 (2006).

[231] T. B. Pedersen, H. Koch, and C. Hättig, The Journal of Chemical Physics **110**, 8318 (1999).

[232] A. Kohn and J. Olsen, Journal of Chemical Physics **122**, 084116 (2005).

[233] G. E. Scuseria and H. F. Schaefer, Chemical Physics Letters **142**, 354 (1987).

[234] K. A. Brueckner, Phys. Rev. **96**, 508 (1954).

[235] M. Nooijen and V. Lotrich, The Journal of Chemical Physics **113**, 4549 (2000).

[236] T. D. Crawford, T. J. Lee, N. C. Handy, and H. F. S. III, The Journal of Chemical Physics **107**, 9980 (1997).

[237] C. Hampel, K. Peterson, and H. Werner, Chemical Physics Letters **190**, 1 (1992).

[238] J. Olsen, The Journal of Chemical Physics **113**, 7140 (2000).

[239] J. A. Parkhill and M. Head-Gordon, Molecular Physics **108**, 513 (2010).

[240] T. D. Crawford et al., J. Comp. Chem. **28**, 1610 (2007).

[241] C. L. Lawson, R. J. Hanson, D. Kincaid, and F. T. Krogh, ACM Trans. Math. Soft. **5**, 308 (1979).

[242] T. Yanai, Y. Kurashige, D. Ghosh, and G. K.-L. Chan, International Journal of Quantum Chemistry **109**, 2178 (2009).

[243] H.-J. Werner and P. J. Knowles, The Journal of Chemical Physics **82**, 5053 (1985).

[244] Y. Kurashige and T. Yanai, The Journal of Chemical Physics **130**, 234114 (2009).

[245] D. Moran et al., Journal of the American Chemical Society **128**, 9342 (2006).

[246] O. Demel, K. R. Shamasundar, L. Kong, and M. Nooijen, J. Phys. Chem. A **112**, 11895 (2008).

[247] T. D. Crawford, E. Kraka, J. F. Stanton, and D. Cremer, The Journal of Chemical Physics **114**, 10638 (2001).

[248] R. Jones and B. R., J. Am. Chem. Soc. **94**, 660 (1972).

[249] E. Kraka, J. Anglada, A. Hjerpe, M. Filatov, and D. Cremer, Chemical Physics Letters **348**, 115 (2001).

[250] F. A. Evangelista, W. D. Allen, H. F. Schaefer, and III, The Journal of Chemical Physics **127**, 024102 (2007).

[251] K. Kowalski and P. Piecuch, Chemical Physics Letters **344**, 165 (2001).

[252] A. Schafer, H. Horn, and R. Ahlrichs, The Journal of Chemical Physics **97**, 2571 (1992).

[253] J. P. Finley, R. K. Chaudhuri, and K. F. Freed, The Journal of Chemical Physics **103**, 4990 (1995).

[254] J. Paldus, J. Čížek, and I. Shavitt, Phys. Rev. A **5**, 50 (1972).

[255] P. Knowles, K. Somasundram, N. Handy, and K. Hirao, Chemical Physics Letters **113**, 8 (1985).

[256] C. Nelin, B. O. Roos, A. J. Sadlej, and P. E. M. Siegbahn, The Journal of Chemical Physics **77**, 3607 (1982).

[257] B. Jeziorski and J. Paldus, The Journal of Chemical Physics **90**, 2714 (1989).

[258] P. G. Szalay and R. J. Bartlett, The Journal of Chemical Physics **103**, 3600 (1995).

[259] L. Fusti-Molnar and P. G. Szalay, The Journal of Physical Chemistry **100**, 6288 (1996).

[260] X. Li and J. Paldus, The Journal of Chemical Physics **128**, 144118 (2008).

[261] X. Li and J. Paldus, The Journal of Chemical Physics **128**, 144119 (2008).

[262] N. B. Amor, D. Maynau, J.-P. Malrieu, and A. Monari, The Journal of Chemical Physics **129**, 064112 (2008).

[263] U. S. Mahapatra, B. Datta, and D. Mukherjee, Chemical Physics Letters **299**, 42 (1999).

[264] E. Prochnow et al., The Journal of Chemical Physics **131**, 064109 (2009).

[265] F. A. Evangelista, W. D. Allen, H. F. Schaefer, and III, The Journal of Chemical Physics **125**, 154113 (2006).

[266] S. Das, D. Mukherjee, and M. Kállay, The Journal of Chemical Physics **132**, 074103 (2010).

[267] B. Jeziorski and H. J. Monkhorst, Phys. Rev. A **24**, 1668 (1981).

[268] X. Li and J. Paldus, The Journal of Chemical Physics **131**, 114103 (2009).

[269] P. Celani and H.-J. Werner, The Journal of Chemical Physics **112**, 5546 (2000).

[270] V. V. Ivanov and L. Adamowicz, The Journal of Chemical Physics **113**, 8503 (2000).

[271] O. Hino, T. Kinoshita, G. K.-L. Chan, and R. J. Bartlett, The Journal of Chemical Physics **124**, 114311 (2006).

[272] D. I. Lyakh, V. V. Ivanov, and L. Adamowicz, The Journal of Chemical Physics **128**, 074101 (2008).

[273] D. I. Lyakh, V. V. Ivanov, and L. Adamowicz, The Journal of Chemical Physics **122**, 024108 (2005).

[274] T. Fang, J. Shen, and S. Li, The Journal of Chemical Physics **128**, 224107 (2008).

[275] T. Fleig, L. K. Sørensen, and J. Olsen, Theoretical Chemistry Accounts **118**, 347 (2007).

[276] L. Adamowicz, J.-P. Malrieu, and V. V. Ivanov, The Journal of Chemical Physics **112**, 10075 (2000).

[277] M. Hanrath, The Journal of Chemical Physics **128**, 154118 (2008).

[278] J. Shen and S. Li, The Journal of Chemical Physics **131**, 174101 (2009).

[279] M. Kállay, P. G. Szalay, and P. R. Surján, The Journal of Chemical Physics **117**, 980 (2002).

[280] M. Hanrath, The Journal of Chemical Physics **128**, 154118 (2008).

[281] J. Olsen, The Journal of Chemical Physics **113**, 7140 (2000).

[282] S. Chattopadhyay, A. Mitra, and D. Sinha, The Journal of Chemical Physics **125**, 244111 (2006).

[283] J. P. Finley, The Journal of Chemical Physics **109**, 7725 (1998).

[284] M. Hanrath, The Journal of Chemical Physics **123**, 084102 (2005).

[285] T. H. Dunning, J. Chem. Phys. **53**, 2823 (1970).

[286] D. Feller, J. Comp. Chem. **17**, 1571 (1996).

[287] K. Schuchardt et al., J. Chem. Inf. Model. **47**, 1045 (2007).

[288] F. A. Evangelista, W. D. Allen, H. F. Schaefer, and III, The Journal of Chemical Physics **127**, 024102 (2007).

[289] K. Kowalski and P. Piecuch, Chemical Physics Letters **344**, 165 (2001).

[290] A. D. Becke, The Journal of Chemical Physics **98**, 1372 (1993).

[291] K. Jankowski, L. Meissner, and J. Wasilewski, International Journal of Quantum Chemistry **28**, 931 (1985).

[292] P. Piecuch and J. Paldus, The Journal of Physical Chemistry **99**, 15354 (1995).

[293] P. Piecuch and L. Adamowicz, Chemical Physics Letters **221**, 121 (1994).

[294] I. Hubac, P. Mach, and S. Wilson, International Journal of Quantum Chemistry **104**, 387 (2005).

[295] S. Chattopadhyay, D. Pahari, D. Mukherjee, and U. S. Mahapatra, The Journal of Chemical Physics **120**, 5968 (2004).

[296] M. Kobayashi and H. Nakai, The Journal of Chemical Physics **131**, 114108 (2009).

[297] W. Li, P. Piecuch, J. R. Gour, and S. Li, The Journal of Chemical Physics **131**, 114109 (2009).

[298] P. W. Ayers, Journal of Mathematical Physics **46**, 062107 (2005).

[299] S. Yanagisawa, T. Tsuneda, and K. Hirao, The Journal of Chemical Physics **112**, 545 (2000).

[300] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[301] Y. Zhang and W. Yang, The Journal of Chemical Physics **109**, 2604 (1998).

[302] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

[303] O. A. Vydrov, G. E. Scuseria, J. P. Perdew, A. Ruzsinszky, and G. I. Csonka, The Journal of Chemical Physics **124**, 094108 (2006).

[304] R. O. Jones and O. Gunnarsson, Phys. Rev. Lett. **55**, 107 (1985).

[305] M. R. Pederson, R. A. Heaton, and C. C. Lin, The Journal of Chemical Physics **80**, 1972 (1984).

[306] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, The Journal of Chemical Physics **126**, 104102 (2007).

[307] A. D. Dutoi and M. Head-Gordon, Chemical Physics Letters **422**, 230 (2006).

[308] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, The Journal of Chemical Physics **126**, 154109 (2007).

[309] A. Dreuw and M. Head-Gordon, Chemical Reviews **105**, 4009 (2005).

[310] J. P. Perdew, V. N. Staroverov, J. Tao, and G. E. Scuseria, Physical Review A (Atomic, Molecular, and Optical Physics) **78**, 052513 (2008).

[311] A. V. Krukau, G. E. Scuseria, J. P. Perdew, and A. Savin, The Journal of Chemical Physics **129**, 124103 (2008).

[312] A. D. Becke, The Journal of Chemical Physics **112**, 4020 (2000).

[313] P. M. W. Gill, R. D. Adamson, and J. A. Pople, Molecular Physics **88**, 1005 (1996).

[314] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, Chemical Physics Letters **275**, 151 (1997).

[315] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, The Journal of Chemical Physics **115**, 3540 (2001).

[316] J.-W. Song, T. Hirosawa, T. Tsuneda, and K. Hirao, The Journal of Chemical Physics **126**, 154105 (2007).

[317] D. Jacquemin, E. A. Perpète, O. A. Vydrov, G. E. Scuseria, and C. Adamo, The Journal of Chemical Physics **127**, 094102 (2007).

[318] J.-W. Song, M. A. Watson, A. Nakata, and K. Hirao, The Journal of Chemical Physics **129**, 184113 (2008).

[319] B. M. Wong and J. G. Cordaro, The Journal of Chemical Physics **129**, 214703 (2008).

[320] B. G. Janesko, T. M. Henderson, and G. E. Scuseria, Physical Chemistry Chemical Physics **11**, 443 (2009).

[321] P. M. W. Gill and R. D. Adamson, Chemical Physics Letters **261**, 105 (1996).

[322] T. M. Henderson, A. F. Izmaylov, G. E. Scuseria, and A. Savin, The Journal of Chemical Physics **127**, 221103 (2007).

[323] J. Toulouse, F. m. c. Colonna, and A. Savin, Phys. Rev. A **70**, 062505 (2004).

[324] J.-W. Song, S. Tokura, T. Sato, M. A. Watson, and K. Hirao, The Journal of Chemical Physics **127**, 154109 (2007).

[325] A. Savin and H.-J. Flad, International Journal of Quantum Chemistry **56**, 327 (1995).

[326] R. Baer, E. Livshits, and D. Neuhauser, Chemical Physics **329**, 266 (2006).

[327] Y. Akinaga and S. Ten-no, Chemical Physics Letters **462**, 348 (2008).

[328] A. D. Dutoi and M. Head-Gordon, The Journal of Physical Chemistry A **112**, 2110 (2008).

[329] M. A. Rohrdanz and J. M. Herbert, The Journal of Chemical Physics **129**, 034107 (2008).

[330] J.-D. Chai and M. Head-Gordon, Chemical Physics Letters **467**, 176 (2008).

[331] A. Fetter and D. Walecka, *Quantum Theory of Many-Particle Systems*, Dover, 2003.

[332] J. Toulouse, A. Savin, and H. Flad, International Journal of Quantum Chemistry **100**, 1047 (2004).

[333] J. Toulouse, F. Colonna, and A. Savin, The Journal of Chemical Physics **122**, 014110 (2005).

[334] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, The Journal of Chemical Physics **106**, 1063 (1997).

[335] J. Heyd, G. E. Scuseria, and M. Ernzerhof, The Journal of Chemical Physics **118**, 8207 (2003).

[336] T. M. Henderson, B. G. Janesko, and G. E. Scuseria, The Journal of Chemical Physics **128**, 194105 (2008).

[337] J. Olsen and P. J. rgensen, The Journal of Chemical Physics **82**, 3235 (1985).

[338] D. Brittain et al., Physical Chemistry Chemical Physics **11**, 1138 (2009).