LOW-SCALING CORRELATED METHODS FOR INTERMOLECULAR INTERACTIONS

ROBERT A. SHAW

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

October 3, 2019

The University Of Sheffield.

ABSTRACT

In this work, I develop theories, and their implementations, for the high-accuracy study of large molecular systems. Intermolecular interactions are fundamental in nature, in particular in materials chemistry and biological systems. They are difficult to study theoretically due to the small energy differences and vast numbers of molecules involved. Present high-accuracy methods can typically only deal with at most a few molecules in a single calculation, limiting applications to the gas phase. Using absolutely localised molecular orbitals (ALMOs, [1, 2]), I develop a new correlated method for intermolecular interactions that is linear scaling in the number of molecules, with accuracy similar to coupled-cluster with single and double excitations (CCSD). I give details of an implementation that minimises the memory imprint and processor time, as well as allowing for extensive parallelisation. Results over benchmark databases of non-covalent interactions show consistent agreement within 0.5 kcal/mol of the CCSD result, with timings two orders of magnitude smaller. Subsequently, I derive analytical derivatives for the total energy, allowing for rapid geometry optimisations on large scale systems. Again, geometrical parameters and vibrational frequencies are shown to agree well with CCSD results. Finally, I extend the ALMO approach to multi-configurational systems, and in particular excited states, demonstrating how it can then be used as an embedding-type method, where different subsystems are treated to different levels of accuracy. This culminates in the reproduction and elucidation of experimentally observed shifts in the photoelectron spectrum of phenol when in water, compared to the gas phase. The methods developed herein thus allow for the high-accuracy treatment of much larger condensed-matter systems than has previously been possible.

PREFACE

They say a little knowledge is a dangerous thing, but it's not one half so bad as a lot of ignorance.

— *Terry Pratchett*

First, some housekeeping. Portions of Chapters 2 through 5 have been published in the following paper:

> **R. A. Shaw** and J. G. Hill, "A linear-scaling method for non-covalent interactions", *J. Chem. Theory Comput.*, 2019, https://doi.org/10.1021/acs.jctc.9b00615

However, the contents of the above is presented in this thesis entirely rewritten and greatly expanded. The aforementioned chapters also contain several not-yet-published developments. Secondly, the most recent stable release of my quantum chemistry program, Gamma, in which are the implementations described herein, can be found at:

> https://www.github.com/robashaw/gamma

This PhD has been an eventful one. Completely separate to the work in this thesis, I have published six papers, with several more in the process of being written or submitted. I have helped supervise almost a dozen undergraduates, becoming good friends with many of them. I have attended conferences and given talks, developed tutorials and workshops, and worked on (at great risk to my sanity) multiple open-source software projects. The PhD environment has allowed me to collaborate and work on topics ranging from pure mathematics, to computational spectroscopy, to even the dreaded field of biochemistry. On that last note, I have learned to be very careful about the number of hydrogens in any and all molecules.

It would be easy to leave it at that, reducing myself to the sum of my productivity. But I am not a product. I have also suffered, been to the very bottom of the darkest places, and scraped and screamed until I found my way out again. In the last few years, I have: been isolated from my friends, gotten married, had a mental breakdown, watched my personal relationships break down, almost been financially crippled, come out to my friends and family, suffered the consequences of coming out while also finally learning how to be myself, and be comfortable in who I am. I am now a proud advocate for diversity and better mental health in science; I am more than just a cog in the terrible, ineffable machine we call academia.

I would not be here without the amazing people around me. First and foremost, I must thank my supervisor, Grant, for his constant support. You have become a friend as well as a colleague, and your kindness, care, and sense of humour, have been just some of the things that have kept me going. I would also like to thank Tim Craggs; your boundless enthusiasm is infectious, as are the systems you study. The other PhD students I have worked with and supported, particularly Heather, Joe, and Adam, have made everything more bearable, as have the wonderful students: Abbie, Hannah, Vicky, Tom, Dan, Emma, Henry, Ollie, Aidan, Paddy, I hope you enjoyed sharing the office with me. Finally for the work friends, special thanks must go to Theo and Tristan - I owe so much of my professional and personal development to them.

Most importantly, I must thank Eren. I have seen you grow in ways I could never have imagined. Your love, strength, and resilience, have been truly inspiring. I would not be here without you, and I would not change a minute of the time we have spent together in the last eight years. I love you and I always will.

CONTENTS

ACRONYMS

ABS Auxiliary basis set, used in density fitting.

ALMO Absolutely localised molecular orbitals, molecular orbitals restricted to be expanded in the local AO basis of a fragment.

AO Atomic orbital, the atom-centred basis functions used in SCF calculations.

BCH Baker-Campbell-Hausdorff expansion, used to expand similarity-transformed operators in terms of commutators.

BSEE Basis-set extension effect, the BSIE term in an energy difference that is not attributed to BSSE.

BSIE Basis-set incompleteness error, the error in the total energy due to using a finite basis set.

BSSE Basis-set superposition error, the error in energy differences due to mismatched basis set sizes between terms in the difference.

CBS Complete basis set, an in principle infinite set of functions which can be linearly combined to exactly represent any function in the function space.

CC Coupled cluster, a popular renormalised MBPT.

CCSD CC with the cluster operator truncated at single and double excitations.

CCSD(T) CCSD with non-iterative (perturbative) triples.

CHA Chemical hamiltonian approach, a method where BSSE terms are directly removed from the molecular Hamiltonian.

CI Configuration interaction, a method for expanding the wavefunction in terms of excitations from a reference determinant, so as to calculate the correlation energy.

CP The counterpoise correction method of Boys and Bernardi, used to correct for BSSEs.

CPU Central processing unit, the main processor on most computers.

CT Charge transfer, a non-local form of polarisation of the electron density on a fragment due to the presence of other fragments.

DCBS Dimer-centred basis set, the combined basis set of the dimer, i.e. the union of two MCBSs.

DF Density fitting, an approximation used to expand four-centre ERIs in terms of three-centre ERIs.

DFT Density-functional theory, where the Schrödinger equation is solved in terms of the electron density rather than the wavefunction.

DIIS Direct inversion of the iterative subspace, used to accelerate SCF convergence.

dRPA Direct random phase approximation, RPA with the bare, non-antisymmetrised Coulomb integrals.

dRPAd Direct random phase approximation decomposed into dispersive excitations.

EDA Energy decomposition analysis, a method for decomposing interaction energies into physical components, such as electrostatics and dispersion.

EOM Equation-of-motion, methods used (particularly in tandem with CC) to solve the time-dependent quantum problem starting from a time-independent solution. Useful for excited states.

ERI Electron repulsion integral, the two-electron integral over the Coulomb operator.

FCI Full configuration interaction: CI with all possible excitations included in the expansion.

FLOP Floating-point operation, the low-level commands executed on floating point numbers on a computer processor.

GP-GPU General-purpose GPU, a GPU card typically designed for consumer graphics applications, such as gaming, rather than high-throughput scientific computation.

GPU Graphical processing unit, the processors typically used to accelerate graphical rendering on a computer.

GWA The GW approximation, a method for calculating electron correlation, related to RPA and rCCD.

HF Hartree-Fock, a mean-field theory where the wavefunction is described using a single determinant.

MBE Many-body expansion, the expansion of an energy as a series in the number of bodies (typically atoms or molecules) involved in each energy term.

MBPT Many-body perturbation theory, used to expand the correlation energy as a series in increasing orders of the electron-electron potential.

MCBS Monomer-centred basis set, the basis set located strictly on a monomer.

MCSCF Multi-configurational self-consistent field, a form of SCF where the wavefunction is built from HF expansions of a linear combination of determinants (or configurations).

MD Molecular dynamics, a method for propagating nuclear motion through time using a forcefield.

MO Molecular orbital, the linear combination of atomic orbitals determined in an SCF calculation.

MP2 Second-order Møller-Plesset perturbation theory, a second-order form of MBPT used to calculate correlation energy from an HF reference.

OBS Orbital basis set, the union of all AOs.

PES Photoelectron spectroscopy, a way to experimentally measure electron ionisation potentials by detaching electrons using high-energy radiation.

rCCD Ring coupled-cluster doubles theory, CCD with only ring-type Goldstone diagrams included, closely related to dRPA.

RPA Random phrase approximation, an approximation to the time-dependent expansion of the electron density, used when calculating electron correlation energies.

RPAx Random phase approximation with the full exchange interaction included.

RPAxd Random phase approximation including exchange decomposed into dispersive excitations.

SAPT Symmetry-adapted perturbation theory, a perturbative method for directly calculating supermolecular interaction energies.

SCF Self-consistent field, a fixed-point approach where a many-body quantity is iteratively approximated by an effective one-body field until self-consistency is achieved.

SOSEX Second-order screened exchange correction to the direct random phase approximation.

1 | INTRODUCTION

Chemistry relies on the ability to take results determined in one context and apply them in entirely different situations. At the most fundamental level, the very notion of molecules belies the assumption that the properties of a system are mostly local in nature; there is no physical distinction, for example, between the individual water molecules in ice, but it is helpful, and undeniably successful, to treat them as such. Thus, it is useful to categorise different systems and types of interactions. A very important class is that of noncovalent interactions, generally defined to be the interactions between molecules, where the inherent locality implies that these are weaker than what chemists consider to be intramolecular, covalent bonds. These are vitally and ubiquitously important - in molecular recognition, van der Waals clusters, hydrogen-bonded systems, physisorption, self-assembly, and many more [3–9].

This entails such a broad class of interactions that it is natural to decompose the classification further, based on the dominant physics in each case. This leads to the intuitive ideas of electrostatics, polarisation, exchange, dispersion, and charge transfer [10]. The first two of these describe the largely classical attraction and repulsion between charge distributions, and the anisotropic response of these to each other. Exchange and dispersion, on the other hand, are purely quantum mechanical terms arising from the antisymmetry constraints on the wavefunction and the instantaneous changes in the electron distributions of each fragment in response to the presence of the other fragments, respectively. The final term, charge transfer, is somewhat controversial, suggesting a transferral of electron density between fragments; it has been shown to be important in many intermolecular interactions [11–16], but highlights the main problem with any such energy decomposition: it is largely arbitrary. Again, this is a reflection of the fact that these classifications are entirely imagined, but nonetheless very useful.

Within the context of quantum chemistry, noncovalent interactions present an interesting and difficult problem. Experimentally, they are very hard to study; in solution, it is virtually impossible to isolate specific interactions - although, recent experiments involving molecular balances are attempting to do just that [17] - while in the gas phase very low temperatures are required so that the thermal energy does not overcome the weak binding [18]. As such, theoretical studies are an invaluable tool, but the small energy differences involved mean

that only the highest accuracy methods yield quantitatively meaningful results. This is further compounded by the fact that dispersion often makes up a substantial portion of the interaction energy, and this term is entirely due to electron correlation. As a result, mean-field methods, namely Hartree-Fock and density-functional theory (DFT), are incapable of describing it adequately [19–23]. In fact, even the cheaper correlated methods are unsuccesful: second-order perturbation theory (MP2) overbinds some complexes, especially those with significant delocalisation such as the benzene dimer [24], while underbinding others, such as saturated systems. Coupled cluster with single and double excitations (CCSD) performs better, but still underbinds many complexes [25]. Only CCSD with a perturbative triples correction [CCSD(T)] is sufficiently accurate across the majority of systems [25, 26]. This has led to it being referred to as the "gold standard" method, although lower-order methods (particularly CCSD) have also been categorised into "silver" and "bronze" standards, denoting their relative usefulness when balancing accuracy and efficiency [27]. Moreover, the dispersion portion of the energy converges very slowly with basis set size [28], such that very large bases are needed. The end result is that anything over roughly thirty atoms becomes unachievable without using significant computational resources.

To make the problem worse, in real systems - e.g. biological and supermolecular environments - sterics and competition between multiple interactions mean that complexes are not in their equilibrium geometries. This necessitates the calculation of potential energy surfaces so as to understand the geometry dependence of the interactions, greatly increasing the magnitude of the computations that must be carried out. While there are several examples of such surfaces at the CCSD(T) level [29–33], the unreasonable $\mathcal{O}(N^7)$ scaling, where N is a measure of system size, quickly makes such investigations unfeasible [34]. If we then extend the problem to that of clusters, which are in general far more interesting and relevant than dimers, we must contend with several additional difficulties [35–37]:

1. many-body effects become very important;

2. the number of local minima increases exponentially with system size;

3. the inter- and intra-molecular degrees of freedom are strongly coupled;

4. the gradients of post-HF methods are tainted with basis set superposition errors.

The latter point in particular presents a difficulty, as it requires that either a very large, near-complete basis is used, or several separate

calculations are carried out on each fragment in the cluster basis (the so-called counterpoise correction approach).

In this thesis, a method is presented that attempts to address many of these problems. One of the biggest problems is the basis set superposition error, which will be described in the next section. Methods to correct for this can be *a priori* or *a posteriori*. The most prominent method is the counterpoise correction of Boys and Bernardi [38], which falls into the latter category. It is desirable to avoid this from the beginning, which leads to the absolutely-localised molecular orbital (ALMO) approach [1, 2, 39–42]. The requirement for a high-accuracy description of dispersion but at a low-scaling cost necessitates methods for the reduction of computational cost, and several such approaches will be discussed. Finally, we will see how the energy from such calculations can be decomposed into physically meaningful terms. The culmination of solving these issues is a method that is capable of treating large supermolecular systems in an accurate and highly efficient manner, yielding specific information about the interactions of interest, including the oft-neglected many-body terms.

1.1 BASIS SET SUPERPOSITION ERROR

In the supermolecular approach, the interaction energy is calculated as the difference between the energy of the whole system, E_t, and the sum of the N fragment energies, E_n:

$$\Delta E = E_t - \sum_{n=1}^{N} E_n \qquad (1.1)$$

As the fragments are brought together from infinity, the basis functions on the adjacent fragments become available for each individual fragment to use, improving the description of the overall wavefunction on that fragment. Thus, if the monomer energies were computed only in the monomer basis, the difference in completeness between the cluster and monomer bases leads to an unphysical lowering of the interaction energy, i.e. overbinding. This error is termed the basis set superposition error (BSSE). It is a purely mathematical artefact of using a finite basis, and in the limit of a complete basis would not be present. Most importantly, it has nothing to do with the physics of the system, a point which is sometimes forgotten [43]. In the case of very small bases, the BSSE can be substantial enough to result in a false minimum on a repulsive curve; for instance, in the example of the helium dimer at the HF/4-31G* level [44]. This particular error can be avoided by simply using a sensible basis set, and in general, BSSE could be made negligibly small by using near-saturated sets. As

This assumes the energy is variational, such that a more complete basis implies a lower energy.

was noted above, however, this is impractical, such that corrections - either before or after the fact - are necessary.

If we consider the simplest case of two fragments, A and B, each would have a monomer-centred basis, {A} and {B}. These would span some finite-dimensional spaces $\mathcal{A}$ and $\mathcal{B}$, respectively, such that if $\mathcal{N}$ is the total function space associated with the problem, $\mathcal{A} \cup \mathcal{B} \subset \mathcal{N}$. Defining $\mathcal{K} = \mathcal{N} \setminus (\mathcal{A} \cup \mathcal{B})$ to be the orthogonal complement to the space spanned by the union of the monomer bases, the complete basis set (CBS) limit interaction energy is given by

$$\Delta E_{AB}^{CBS} = E_{AB}[AB\overline{K}] - E_A[A\overline{BK}] - E_B[B\overline{AK}] \tag{1.2}$$

where $E_{A\cdots M}[A\cdots M\overline{N\cdots Z}]$ indicates the energy of the subcluster $A\cdots M$ calculated in the basis formed from the union of the bases in brackets. The overline implies that the functions are not located on the subcluster, so-called 'ghost functions', as they have no particles associated with them. In comparison, the uncorrected energy would be

$$\Delta E_{AB}^{unc.} = E_{AB}[AB] - E_A[A] - E_B[B] \tag{1.3}$$

The difference between these values is then the basis set incompleteness error (BSIE). We can split this into a basis set superposition error, and an intrinsic basis set incompleteness error. The latter reflects the fact that a more complete basis will give a better description of all components, regardless of whether energy differences are being taken. This leads to so-called basis set extension effects (BSEEs), which are not erroneous but rather a necessary component of the energy [35, 43]. Clearly, the BSSE is closely entangled with these BSEEs, making it very hard to determine whether a discrepancy is an artefact of taking differences, or simply due to the incomplete description of the system. Brauer [45] noted that, as these terms are opposite in sign - BSSE leads to overestimation of the interaction energy, intrinsic incompleteness to underestimation - in the case of small basis sets, they can cancel each other out, and in fact the latter term can become dominant. This is why in some cases double zeta quality basis sets can appear to perform better than much larger sets [45].

The situation only deteriorates as more fragments are added. The many-body expansion (MBE), one of the most commonly used methods for considering large systems [35, 46], is entirely based around energy differences, such that the superposition errors compound. If the monomer basis is used, BSSE leads to slow and oscillating convergence of the MBE [47]. If either the full cluster (monomers calculated using all functions) or subcluster basis (Nth monomer calculated using functions of the N-cluster) is used, the MBE converges rapidly, but to different values. The difference in these values, due to the differing quality of the descriptions of the wavefunction, is a BSEE - if

it were purely a superposition error, we would not expect rapid convergence [48]. This demonstrates the complications that arise in the supermolecular scheme.

It should be noted that there may also be a matching error to BSSE present at the level of the n-electron configuration set [49], termed the configuration state superposition error. Unlike BSSE, this will not disappear as the basis tends to completeness. Additionally, some have argued that intramolecular BSSE may also be possible [50, 51]; we stress, however, that BSSE is only relevant when energy differences are being considered - in any other case, the errors are due to the intrinsic basis set incompleteness. While it is possible to define the fragments within a supermolecular scheme as the atoms, there is no need, or indeed use, in most cases to try and determine the pairwise atom interactions, rendering any attempt at correcting for an intramolecular BSSE pointless.

1.2 COUNTERPOISE CORRECTIONS

The most popular method of eliminating BSSE is the counterpoise (CP) correction [38, 43]. In this scheme, the interaction energy is computed as

$$\Delta E_{AB}^{CP} = E_{AB}[AB] - E_A[A\overline{B}] - E_B[B\overline{A}] = \Delta E_{AB}^{unc.} + \delta CP \qquad (1.4)$$

where the CP correction is defined as

$$\delta CP = \left(E_A[A] - E_A[A\overline{B}] \right) + \left(E_B[B] - E_B[B\overline{A}] \right) \qquad (1.5)$$

In this way, all quantities are treated on an equal footing. Note that, if the method used is variational, the more complete basis necessarily gives a lower energy, such that $\delta CP \geqslant 0$.

The CP correction can significantly improve convergence of the interaction energy to the complete basis set limit; for example, in the case of the benzene dimer [52]. However, there has been considerable debate in the literature about whether it is a valid procedure, and whether it overcorrects for the BSSE [53, 54]. The majority of these complaints stem from either a misunderstanding of the fact that the finite basis is a completely mathematical construct or from a conflation of BSSE with BSEEs [43, 45]. One point of contention is that, as has been shown for several hydrogen-bonded [55] and dispersively bound [52] dimers, the CP-corrected potential energy curves converge from above, while the uncorrected curves come from below. This is not indicative of an overcorrection, but rather the correct behaviour: given an incomplete basis and a variational method, we fully anticipate that the curve should converge from above. From a practical

point of view, it does suggest that a half-half method, blending the corrected and uncorrected results, might be a useful ad hoc approach [56, 57].

Possibly the biggest and most often repeated complaint was first put forward by Mayer and coworkers [58], who claimed that there is a hidden non-additivity assumption implicit in the counterpoise method. At infinite separation, the dimer total energy can rigorously be written as the sum of the monomer energies. As they are brought together, mixing of the basis sets is allowed to occur, such that the dimer energy becomes the sum of the monomer energies in the new environment, plus a counterpoise correction. Thus, the total energy is implicitly no longer additive. Of course, there is no reason why it should be additive, as the complex itself is not some noninteracting addition of unperturbed monomers.

Mayer highlighted the particular case of the relaxation energy [44], the energy associated with the change in geometry of the fragments that occurs in going from the noninteracting to the interacting system. Take, for example, the water dimer, and allow one of the intra-monomer OH bonds to relax in the dimer basis. If the intermonomer $O \cdots O$ distance is kept constant, the ghost orbitals do not move, but if the $O \cdots H$ distance is fixed, they do. As such, different relaxation energies are obtained along different paths. This implies that relaxation should be determined in the monomer basis, such that each point on a potential surface would require seven energies: the dimer and relaxed monomers in the dimer basis, along with the relaxed and unrelaxed monomer energies in the monomer basis. Certainly, this means that arguments based around the point that CP is "exact due to the same basis being used throughout" are not valid. However, this ambiguity is not a problem with the counterpoise method, but rather with the fact that basis sets are atom-centred. While this is very convenient for most calculations, it means that different regions of space are treated in unequal ways, such that significant changes in geometry such as those above lead to an uneven description.

In a similar vein, it is often claimed [18, 44] that the counterpoise method allows for the delocalisation of the monomer electrons into orbitals on the partner that would, in the dimer, be occupied and thus excluded from the available space. However, this is wrong in several ways. Firstly, the basis functions are not themselves orbitals: all functions are available to be used in the description of any given orbital, and have absolutely no concept of whether that orbital will be occupied. Expansion of the basis simply improves the description of the function space within which the orbitals reside, and so can only be a good thing. In addition, as was pointed out by van Duijneveldt [43], the exclusion of electrons from monomer A in the region of the occupied orbitals of monomer B, and vice versa, is an important part of

the physics trying to be described - namely, it constitutes the exclusion principle. These regions would categorically not be excluded if the second fragment was not present.

Misunderstandings aside, the above does not mean that the CP approach is without issues. First and foremost is that it only accounts for the error in the energy, and the wavefunction itself remains tainted with BSSE. One particularly striking case was noted by Fowler and Buckingham [59], where the neon atom in the Ne⋯HF complex acquires a dipole moment, which it clearly should not have. Unfortunately, CP corrections for quantities other than the energy are hard to define. Perhaps the biggest problem, however, is with the extension to systems of more than two fragments. There is a great deal of ambiguity in how such a scheme would be defined; for example, when determining the interactions in a chain, is it best to calculate each additional monomer in the whole basis up to that point, or should the basis of the whole chain be used in all instances? Mathematically, the answer is obvious: only using the full cluster basis rigorously removes the BSSE. For instance, in the 'one at a time' scheme in the case of the hydrogen fluoride chain, different energies would be obtained if the next monomer was added at the hydrogen end or the fluorine end, a situation that is clearly unphysical [60].

This means that for a complex of N fragments, $N + 1$ calculations need to be done in the full cluster basis. Clearly, this will rapidly become very computationally expensive. In addition to this, the counterpoise correction at the correlated level can be much larger than at the mean-field level [37]; given that such methods in general scale as at least $\mathcal{O}(N^5)$ [34], the CP scheme is impractical. While methods have been devised to try and alleviate this, in particular by using a many-body expansion type decomposition of the correction terms [46, 60–63], it seems that *a posteriori* schemes such as this are fundamentally limited in their scale. Thus, it is necessary to attempt to eliminate BSSE explicitly from the beginning.

1.3 A *priori* METHODS FOR ELIMINATING BSSE

There is a veritable zoo of methods that attempt to get around the need for a counterpoise correction by removing BSSE from the calculation in some way. These include symmetry-adapted perturbation theory [64, 65], block-localised wavefunctions [66], absolutely/extremely localised molecular orbitals [2, 67–69], local correlation methods [70–75], dual basis set methods [76–78], and the chemical Hamiltonian approach [44, 79, 80], to name but a few. We will briefly discuss and contrast the main facets of several of these in this section, before using

the concept of ALMOs as the basis for our own approach in the rest of this work.

1.3.1 Symmetry–adapted perturbation theory (SAPT)

By definition, noncovalent interactions are relatively weak compared to the intramonomer interactions, making them ideal to be treated in a perturbative manner. The most successful variant of this is SAPT [64, 65, 81, 82]. In SAPT, the dimer Hamiltonian is partitioned into contributions from the Fock operator of each monomer, F_X, the interaction, V, and the fluctuation potential, W:

$$H = F_A + F_B + V + W_A + W_B$$

The fluctuation potential contains the monomer correlation energy terms (the difference between the true monomer Hamiltonian, and the monomer Fockian), while $\hat{V}$ contains all the intermolecular interaction terms. The interaction energy is then a triple perturbation series:

$$E_{\text{int.}} = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left(E_{\text{pol.}}^{(nkl)} + E_{\text{exch.}}^{(nkl)} \right)$$

where n is the order in V, while k, l, are the orders in W_A and W_B, respectively. The polarisation and exchange terms are from the polarisation expansions and repulsive interactions due to the antisymmetry of the wavefunction, respectively. The series is then truncated, with this generally written as $E_{\text{label}}^{(vw)}$, with v being the order of V, and w the order in $W_A + W_B$. The label divides the terms into physical contributions, which is usually the primary reason for using this method. These terms are broadly split into four categories: electrostatics, which is the interaction of the monomers' charge distributions; exchange, due to the Pauli repulsion between electrons; induction, due to the anisotropy of the monomer charge distributions in response to the presence of the other monomers; and dispersion, the purely quantum-mechanical interaction between electronic fluctuations on different centres.

The zeroth-order wavefunction in SAPT is a product of monomer wavefunctions, computed within either a monomer-centred (MCBS) or dimer-centred (DCBS) basis, where the latter includes the basis functions of the other monomer just as in the counterpoise procedure. In practice, only a few parts of the calculations are affected by this, not including the general improvement due to a larger basis [18]. In particular, exchange terms are affected. In the DCBS, charge transfer will be included in induction, whereas this cannot happen in a finite MCBS as there is no possibility of delocalisation between monomers. Therefore charge transfer can be defined as the difference between

the sum of the second-order induction and exchange-induction contributions, as calculated using the DCBS and MCBS [16]. This gives a sensible prescription for a quantity that is otherwise hard to define, removing some of the arbitrariness of the energy decomposition, allowing for detailed studies into the importance of charge transfer in weakly-bound systems [12, 14, 83–85]. It is this arbitrariness that is usually quoted as the main flaw of SAPT, although it does suffer from other problems; in particular, as it is perturbative it struggles to describe strongly correlated systems, especially in the case of anionic complexes [18].

Computationally, the cost of SAPT is no better than traditional correlated methods, and in fact can be prohibitively worse. Recently [86], density fitting has been used to write the three-index integrals in the molecular orbital basis, such that the four-index integrals need never be computed, analogously with many-body perturbation theory methods. This has allowed SAPT to compete, with DF-SAPT0 being feasible for systems with more than 2500 basis functions. In conjuction with natural orbitals - a method of localising the molecular orbitals - higher-order SAPT of near-CCSD(T) quality is possible for about 1200 basis functions on a single computer [18]. Additionally, DFT-SAPT variants exist, but these suffer from the usual problems of inaccuracy inherent to density functionals [22, 87].

Density fitting will be described in the next section.

1.3.2 Absolutely localised molecular orbitals (ALMO)

Empirical evidence suggests that electron distributions in complexes are predominantly localised to each fragment, as is chemically intuitive. The canonical orbitals found through conventional SCF methods are, however, highly delocalised due to the orthonormality constraint. As the SCF energy is invariant to unitary transformations within the occupied and virtual spaces, it is possible to localise the molecular orbitals after the fact [88–91]. This does not adequately get rid of the orthogonalization tails, though, such that truncation is necessary, leading to large errors in total energies on the order of 1.5 eV [92, 93]. The ALMO approach ensures that truly - absolutely - localized orbitals are obtained by constraining the orbital coefficient matrix to be block diagonal, with the blocks defined by the fragments. This is done by expanding each fragment only in its own basis, and then coupling the fragments through the supermolecular Fock matrix. Thus, this is essentially an MCBS level of description. A much more detailed exposition of the ALMO SCF procedure is given in Chapter 2.

Unitary transformations in this context correspond to orbital rotations.

The primary advantage of this formulation is that it necessarily eliminates any possibility of BSSE. It was first suggested by Stoll [1], then further developed by Cullen [39]. It is most often erroneously attributed to Gianinetti *et al.* [40, 94], who rediscovered the approach al-

most two decades later. The method has then been further expounded upon by Iwata and coworkers [41, 95–98], and then later by the Head-Gordon group, who have developed energy decomposition analyses based on correlated extensions of the ALMO approach [2, 42, 99–104]. Several different names have been used for the same idea over the years, most notably SCF-MI (where the MI stands for molecular interactions) [40] and LP-SCF-MI (locally projected SCF-MI) [41]; it is my belief that ALMO much better represents the nature of the method (it is not necessarily constrained to noncovalent interactions) and is truer to both the original nomenclature of Stoll and the most recent developments.

Several studies have demonstrated that ALMO SCF methods greatly underpredict the binding energies in all systems [39–41], doing particularly badly for strongly polar complexes. This is due to a necessary lack of any charge transfer between fragments, as will be discussed in more detail later. For example, Cullen found that in the hydrogen fluoride dimer, charge transfer accounts for -0.9 kcal/mol of the interaction, which is roughly a third of the total, while for the water dimer, it is -0.62 kcal/mol or around a fifth [39]. These deviations are therefore significant unless moderate to large separations are being considered; in that case, as charge transfer falls off exponentially with distance, the localised method performs surprisingly well.

This flaw actually becomes an advantage, however. Nagata [41] suggested a method for perturbatively accounting for the charge transfer, which was then greatly refined by Khaliullin *et al.* [42] While this does reintroduce some BSSE, the perturbation is small enough for this to be insignificant, and the resulting ALMO plus charge transfer energies are found to be almost indistinguishable from counterpoise corrected HF energies [2, 99]. The advantages are then threefold: the need for a counterpoise correction is avoided, representing a substantial computational saving; only the fragment Fock matrices need be diagonalised, never the full Fock matrix, representing an $\mathcal{O}(N^3)$ saving; and the energy is immediately decomposed to give a rigorously defined charge transfer term. The result is a method that is both more efficient, and more readily extendable to traditional electron correlation methods, than SAPT. Moreover, the absolute localisation of the orbitals is beneficial in achieving linear scaling with system size, as will be demonstrated later. It has also been shown that there is very little dependence on basis set size [41], in stark contrast to CP-corrected traditional methods and SAPT. Despite these advantages, however, it is undeniable that the description of the wavefunction is lacking, due to the severe constraint imposed on it.

1.3.3   The chemical Hamiltonian approach (CHA)

An alternative approach is to try to identify the terms in the supermolecular Hamiltonian that introduce BSSE into the interaction energy. In the free monomer case, one deals with intramolecular operators. When the basis is extended by the introduction of a partner molecule, these operators become contaminated with intermolecular terms, which Mayer and coworkers [44, 79, 105] identify as being responsible for superposition errors. They suggested avoiding this problem by projecting out the complementary space to that of the monomer in question. That is, if $\hat{h}_A$ acts on the orbital ψ_i^A on monomer A, then this can be decomposed as

$$\hat{h}_A \ket{\psi_i^A} = \hat{P}_A \hat{h}_A \ket{\psi_i^A} + \left(\hat{1} - \hat{P}_A\right) \hat{h}_A \ket{\psi_i^A} \tag{1.6}$$

where $\hat{P}_A$ is a projection operator onto the space spanned by the monomer A orbital basis, i.e.

$$\hat{P}_A = \ket{\mu} \left(S_A^{-1}\right)_{\mu\nu} \bra{\nu}$$

where $\ket{\mu}$ is an atomic orbital on monomer A, $\mathbf{S}_A$ is the overlap matrix of all such basis functions, and the summation convention is implicit. It was argued that the second term on the right of equation 1.6, in the complementary space, is what contains the BSSE, such that mapping $\hat{h}_A \to \hat{P}_A \hat{h}_A$ should result in a BSSE-free ansätz. This leads to the matrix elements associated with the intramonomer operator, $\mathbf{h}_A$, becoming $\mathbf{SS}_A^{-1}\mathbf{h}_A$, where $\mathbf{S}$ is the full overlap matrix. Clearly, when $\mathbf{S} = \mathbf{S}_A$, nothing has changed, and so the intramonomer terms are preserved. The suggestion is then that, unlike in the ALMO approach, only the spurious intermolecular terms have been removed, not the necessary and physically relevant charge transfer terms.

It is possible to introduce this approximation throughout the standard SCF procedure by replacing the on-fragment Fock operator, $\hat{F}_X$, by its projection $\hat{P}_X \hat{F}_X$, leading to the CHA/F method [44]. This does result in a non-Hermitian Fock matrix, however, although supposedly the only practical consequence of this is the need to orthogonalize the occupied orbitals in each iteration before calculating the density. The energy is calculated using the full Hamiltonian, such that the energy is real - not including the omitted terms above leads to wildly erroneous results [106]. This might seem like a condemnation of the choice to remove these terms in the first instance, but it is not dissimilar to the approach taken in some variants of SAPT [107]. Indeed, when the energy is calculated in this way, the results of CHA/F and counterpoise-corrected SCF are virtually indistinguishable [79].

Mayer then argues [44] that any remaining differences between the two are due to two extra contributions present in the CP scheme: the electron affinity of the receiving molecule increases the delocalisation,

and thus the BSSE; the ghost orbitals allow delocalisation into orbitals that would otherwise be occupied. As has been noted above, both of these are in fact physical and necessary, and it is the CHA method that is approximate. Regardless, both CHA and CP energies converge to essentially the same result far more quickly than the BSSE disappears in the uncorrected case [44], suggesting that such quibbles are immaterial. As with ALMOs, one of the primary advantages of CHA is that it can be readily extended to include electron correlation, unlike SAPT, and can be used to compute the interaction energy directly, unlike supermolecular methods. Results have been reported for full configuration interaction and second-order perturbation theory variants of the chemical Hamiltonian [108, 109], although these have not found widespread use. Applications have been restricted to small systems, as the method offers no computational savings over traditional methods, and actually introduces an overhead [44, 110–112].

The main flaw, however, is the non-Hermiticity of the resulting operators within the approximation. Hamza *et al.* claim [79] that this is a necessity, as the basis set superposition error is not a physical observable. This is quite clearly a logical fallacy. While it is a generally accepted axiom of quantum mechanics that any physical observable can only be represented by a Hermitian operator, the converse is demonstrably not true; that is, not every Hermitian operator represents an observable. Given a quantum system embedded in a Hilbert space, the set of all bounded (not to mention unbounded) Hermitian operators on that space is uncountable. The real underlying structure of such a system is in fact the C^*-algebra formed from all sum and product combinations of operators that do represent observables, of which the Hilbert space is only one possible representation. Moreover, if the BSSE is the result of some non-Hermitian operator $\hat{B}$, one could form an equivalent Hermitian representation by taking the average $(\hat{B} + \hat{B}^\dagger)/2$ as the error is necessarily a real quantity. In fact, the same group over a decade later realised that their method could easily be made to be Hermitian by treating the bras and kets symmetrically in this way, disproving their own argument without realising it [80].

It is possible to formulate quantum mechanics with the assumption of parity-time symmetric operators instead; these are Hermitian, albeit with respect to a different sesquilinear form.

1.4 SCALING WITH SYSTEM SIZE

Other than BSSE, the biggest issue with calculations on noncovalent interactions is the prohibitive scaling that accurate *ab initio* methods show with system size. For MP2, CCSD, and CCSD(T), these go as order N^5, N^6, and N^7, respectively [34], implying that just considering a dimer will result in between 32 and 128 times more computational cost compared to the monomer. The 'cost' in this context is both in terms of time and resources, particularly memory. Even the SCF

portion of a calculation scales formally as N^4 due to the need to calculate two-electron, four-centre integrals in the AO basis. In addition, results often converge very slowly with basis size, in particular the correlation contribution, such that very large basis sets are needed for high accuracy results. We consider here the main approaches for dealing with the major bottlenecks in correlated, wavefunction-based methods.

1.4.1 Dual basis sets

In the self-consistent field methods that dominate electronic structure theory, there are two computationally significant steps in each iteration: the building of the Fock matrix, and its subsequent diagonalisation. Low-scaling methods have been developed for the formation of the Coulomb [113–119] and exchange [120–123] contributions of the Fock matrix, such that it is the second step that can become the main bottleneck. Even the most efficient matrix diagonalisation routines necessarily scale as $\mathcal{O}(N^3)$ [124], and while decades of developments in the numerical analysis community have reduced the prefactor significantly, especially for symmetric matrices such as the Fock matrix, the formal cubic scaling is unavoidable. The ALMO method described above sidesteps this problem by only requiring diagonalisation of each fragment Fock matrix, such that effectively linear scaling is achieved; the other approaches described above cannot claim this, even SAPT when the δHF term is included.

An alternative is to use dual basis sets. This came from the observation that the requirements for a complete description of the occupied orbitals are drastically different to those for the virtual orbitals [76]. The root of this is that the former describe static correlation, while the latter describe dynamical correlation. It is this dynamic, dispersive contribution that requires many nodal surfaces in the orbital space, which do not contribute significantly in the occupied space. The consequence is that the occupied orbitals converge fairly rapidly with increasing basis dimension, whereas the electron correlation does not [125, 126]. In fact, the latter goes as ℓ^{-4} in the large-ℓ limit, where ℓ is the angular momentum of the functions used, reflecting their nodal structure [127]. Summing over all such contributions yields L^{-3} convergence, with L being the highest angular momentum included. This is unreasonably slow, especially when one considers that each additional ℓ gives $2\ell + 1$ additional functions, and integrals over these functions get progressively more expensive with increasing ℓ. In addition to this, investigations with correlation consistent basis sets show that the radial nodes should increase proportionally to the angular nodes [128, 129], resulting in the need for around 300 correlating orbitals per atom if 0.1 kcal/mol accuracy is desired.

The fundamental problem is that the strong, short-range correlation requires nodal surfaces that effectively partition the relevant volume of space into a fine grid of small cells, something that is very difficult to achieve with an atom-centred Gaussian basis [130]. The natural solution to this is to use dual basis sets, where a compact basis is used for the initial SCF, avoiding large-N diagonalisation steps, before the Fock matrix is then perturbatively expanded into a much larger, dual basis that can then be used in the correlated portion of the calculation. This was first proposed several decades ago [131, 132], before later being applied to MP2 [133]. In this way, speedups on the order of 12 times were found for the water dimer using a 5Z quality basis, with errors on the micro-Hartree scale compared to the full calculation [76]. In fact, the dual basis calculation is both cheaper and more accurate than the equivalent full QZ calculation, and the accuracy only improves when energy differences are considered, such as in determining interaction energies.

This method has found several uses [78, 134–137], but of particular significance is the complementary auxiliary basis set singles correction used in explicitly correlated methods [138–143]. These account for the poor description of the Coulomb hole by explicitly including terms in the wavefunction that are linear in the interelectron separation, resulting in considerably faster convergence with respect to basis [139, 144]. When combined with the dual basis approach, essentially complete basis set limit HF energies can routinely be achieved [143]. This comes at minimal extra cost, as a large auxiliary basis is necessary for subsequent portions of the F12 calculations. An analogous use is in perturbatively correcting for the lack of charge transfer in the ALMO approach, where the monomer Fock matrices are expanded into the full cluster basis. This will be described in more detail in Chapter 3.

It is also possible to use this approach to determine counterpoise-corrected energies without ever doing the calculation in the full basis [78]. In this context, it is natural to define the small basis as the monomer-centred one, and the dual basis as that of the whole cluster. As the energy is then calculated either as a single sum, iteratively, or as a single diagonalisation, depending on the approximation used, the cost is then significantly less than performing the full calculation. This results in speedups on the order of five to ten times and recovers more than 95% of the counterpoise correction, even for small initial bases [78]. This can also be applied to DFT, where the resulting corrections are larger [145], but the percentage recovered is broadly similar.

1.4.2 Density fitting

After the initial SCF calculation, a set of orbital coefficients is generated, $\{C_\mu^p\}$, describing the component of atomic orbital χ_μ in the molecular orbital ϕ_p. As in general more basis functions are used than there will be occupied orbitals, the molecular orbital space disconnects into occupied and virtual subspaces, usually denoted by the letters i, j, k and a, b, c, respectively. All traditional post-HF methods that try to account for electron correlation require that the two electron integrals, $\langle\mu\nu|r_{12}^{-1}|\lambda\tau\rangle$ be transformed from the AO to the MO basis [34]:

$$(pq|rs) = C_\mu^p C_\lambda^q \langle\mu\nu|r_{12}^{-1}|\lambda\tau\rangle C_\nu^r C_\tau^s \tag{1.7}$$

A naive consideration of the above suggests that this requires eight nested loops to carry out, thus scaling as $\mathcal{O}(N^8)$, where N here is specifically the number of basis functions. In fact, if we instead factorise equation 1.7 as follows

$$(pq|rs) = C_\mu^p \left[C_\lambda^q \left[\left[\langle\mu\nu|r_{12}^{-1}|\lambda\tau\rangle C_\nu^r \right] C_\tau^s \right] \right] \tag{1.8}$$

then the transformation splits into four quarter transforms, each of which is over five indices, reducing the scaling to $\mathcal{O}(N^5)$. In fact, second-order perturbation theory only requires the elements connecting occupied orbitals i, j with virtual orbitals a, b, of the form $(ia|jb)$. If we transform the occupied terms first, then the quarter transforms scale as oN^4, o^2N^3, o^2vN^2, and o^2v^2N, where o and v represent the number of occupied and virtual orbitals, respectively. For any given system, o is independent of basis size, meaning that oN^4 can represent a substantial saving on N^5.

Nevertheless, equation 1.8 represents a fundamental limitation for the scaling of correlated methods. For this reason, the density fitting approximation was introduced [146–152]. This rests on the fact that the two-electron integrals can be rewritten in terms of the one-particle orbital densities, $\rho_{pq}(\mathbf{r})$, as

$$(pq|rs) = \int d\mathbf{r_1} \int d\mathbf{r_2}\ \rho_{pq}(\mathbf{r_1})r_{12}^{-1}\rho_{rs}(\mathbf{r_2})$$

Introducing an auxiliary basis set, $\{\chi_P(\mathbf{r})\}$, of dimension M, the densities can be expanded as

$$\rho_{pq}(\mathbf{r}) \approx \overline{\rho}_p q(\mathbf{r}) = d_P^{pq}\chi_P(\mathbf{r}) \tag{1.9}$$

where the d_P^{pq} are fitting coefficients, hence the name density fitting. As demonstrated by Dunlap $et\ al.$ [147, 148], the difference between the fitted and exact densities in the r_{12}^{-1} weighted norm is minimised when the coefficients are

$$d_P^{pq} = (pq|Q)\left[\mathbf{G}^{-1}\right]_{QP} \tag{1.10}$$

where $G_{QP} = \langle Q|r_{12}^{-1}|P\rangle$ and the three-centre integrals are given by

$$(pq|Q) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_1)r_{12}^{-1}\chi_Q(\mathbf{r}_2) \qquad (1.11)$$

Equation 1.9 means that the integrals in equation 1.7 can be written as

$$(pq|rs) \approx (pq|P)\left[\mathbf{G}^{-1}\right]_{PQ}(Q|rs) = b_{pq}^Q b_{rs}^Q \qquad (1.12)$$

where $b_{pq}^Q = (pq|P)\left[\mathbf{G}^{-1/2}\right]_{PQ}$. This is sometimes conflated with the resolution of the identity approximation, due to the similarity in form of the above to the spectral representation of the identity operator in a complete, nonorthogonal basis:

$$\hat{1} = |p\rangle \, \langle p|q\rangle^{-1} \, \langle q|$$

This approach is used separately in explicitly correlated methods, using a different auxiliary basis, but it is not the same as density fitting [140, 144, 153, 154].

The advantages of density fitting are that the four-index quantities have been reduced to three-index ones, and that the evaluation and construction of the integrals may be carried out in the AO basis. This reduces the steep $\mathcal{O}(N^4)$ storage requirements to $\mathcal{O}(N^2M)$, where M is the size of the auxiliary basis, and the transformation to the MO basis reduces to two $\mathcal{O}(N^2M)$ steps. While the fitting basis is necessarily four or five times as large as the orbital basis, the savings in cost both in terms of speed and memory can be over an order of magnitude, even for small systems [155]. The resulting error is usually on the order of 0.01 kcal/mol in calculations on noncovalent interactions [156, 157]. Similar approaches to density fitting, such as Cholesky decomposition of the integrals into three-index quantities [156, 158], or the pseudospectral approximation [159] could also be used to achieve similar results. Most importantly, all of these approaches are not limited to one specific method. They have been applied successfully in the Fock-build portion of SCF calculations [122, 123], SAPT [86], DFT [160–162], and post-HF methods [152, 163, 164]. The primary disadvantage, however, is that different applications generally require very different auxiliary bases, so that several different sets have to be used within a single calculation. For example, fitting sets used for density fitting of the Coulomb and exchange integrals (so-called 'JK-fit' sets) in HF and DFT calculations do not work well in density-fitted MP2 [151, 162].

1.4.3 Local correlation

Even when the unreasonable scaling of the integral transformation step has been accounted for, correlated methods at higher levels of theory than MP2 still have an inherently unphysical scaling. For example,

the configuration interaction with single and double excitations scales as n^2N^4 [34], where n is the number of orbitals included in the correlating space and N is the total number of orbitals. The n^2 term is unavoidable without neglecting distant pairs of orbitals, but the N^4 portion is unsatisfactory, and entirely due to the use of highly delocalised canonical orbitals [70]. As has been noted above, orbitals should be relatively well-localised, especially in the occupied subspace, and it is only the requirement for orthogonality that gives substantial delocalisation. One way around this is fragmentation methods [165–168], similar to in the ALMO approach, but where subsystems beyond a distance cutoff are treated as multipoles [10]. This can lead to problematic discontinuities on the potential energy surface, however.

Core orbitals, i.e. non-valence orbitals, are often neglected in correlated calculations, as excitations from them will usually have vanishingly small amplitudes.

Alternatively, one could try to find a local representation for the orbitals by optimising some localisation criterion. This was first suggested by Lennard-Jones in 1949 [169], and since then many different criteria have been devised. The most notable of these are those of Boys [170, 171], Edmiston-Ruedenberg [171], and Pipek-Mezey [172]. The former minimizes the sum of the orbitals' second moments (essentially, their variances), while the latter minimizes the number of atomic centres with which an orbital has significant overlap. The Edmiston-Ruedenberg scheme instead maximises the sum of orbital self-repulsion energies, which is somewhat more complicated. All local correlation methods rely heavily on the use of these methods [173–175], or more advanced variants proposed recently [88], to strongly localise the occupied orbitals.

Unfortunately, applying these procedures to truncate the virtual, correlating space is not successful [70], so that truncation needs to occur at the level of the atomic orbitals. Instead, the inherent locality of the atom-centred basis is used. The virtual orbitals were originally chosen by Pulay [70] to be the set of AOs but with the occupied space projected out - so-called 'projected atomic orbitals' - so as to ensure strong orthogonality between the two subspaces. This basis need not be the same as that used in the original SCF, but can be a larger, auxiliary basis, in analogy to the dual basis methods described earlier; in fact, it is not uncommon for a dual basis correction to the SCF energy to be included in local correlation methods. Since then, several different choices of orbitals have been proposed, such as pair natural orbitals [142, 176, 177] and orbital-specific virtuals [178, 179], each with different localisation characteristics [88]. The advantage of forming the orbital spaces in this way are twofold [76]:

1. pair correlation between distant, localized orbitals is small and so can be treated more cheaply, or not at all, i.e. the number of configuration state functions drops from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2)$;

2. the steep and unphysical expansion of the virtual space available to each electron can be eliminated by restricting excitations

to AOs spatially close to the occupied MOs, as the coupling $(ia|jb)$ disappears when the differential overlaps (ia) or (jb) are small.

In this way, essentially linear scaling has been achieved for MP2 and CCSD, and considerable reductions in scaling for their explicitly correlated variants [72, 73, 142, 174, 180, 181]. Such scaling is achieved for surprisingly small systems, for example in the glycine chain $[\text{gly}]_n$, it begins at around $n = 4$ [77].

The primary disadvantages of this approach are the loss of simplicity, and the somewhat arbitrary nature of choosing which terms to neglect. In particular, there is always a danger that a given choice of threshold will apply to one system but not others, and perhaps even the same system but at different geometries. This would result in discontinuous potential surfaces. It should be noted that linear scaling is also possible in MP2 by using Laplace transforms of the energy denominators [133, 182, 183], but this requires an additional numerical integration step and thus has a higher prefactor than local correlation methods. However, the integration only requires very few quadrature points, and it is possible to achieve arbitrary accuracy simply by increasing the number of such points, a quality that local methods lack. In practice, though, the potential loss of accuracy does not seem to be a problem - the effects of using LMP2 as compared to canonical MP2 have been shown to be negligible for structures [72, 184, 185], frequencies [185] and enthalpies [72]. An additional, accidental consequence of the local approach is that the BSSE in the correlation energy can be reduced substantially, as fragments cannot access functions outside their local virtual space [37, 72]. Moreover, the local nature of the excitations allows for a meaningful decomposition of the correlation energy into physically distinct terms, such as dispersion [37].

1.5 ENERGY DECOMPOSITION ANALYSES

This categorising, or energy decomposition analysis (EDA), is one of the most appealing aspects of the non-counterpoise methods to chemists. The first example of such a scheme was proposed by Morokuma [186, 187], and then improved upon by Frey [188], which essentially separated contributions by zeroing different blocks of the Fock matrix. This was unreliable, however, with the energies so obtained often blowing up to infinity due to the loss of well-posedness in the eigenvalue problem [39]. Since then, several other schemes have been suggested, such as the effective fragment potential method [189–191], which is similar to but less computationally expensive than SAPT. Another very popular approach is the natural EDA [192, 193], which utilises natural bonding orbitals. These have been found to be

A problem is well posed if a unique solution exists for any given parameters, and the solution changes continuously with respect to those parameters.

excellent at elucidating qualitative trends, but are not quantitatively valid [194, 195]. In particular, the definition of charge transfer within the natural EDA can lead to very strange results, such as charge transfer contributions far greater than the total interaction energy, even for weakly-bound systems [2]. A recent analysis by Stone [195] has further demonstrated how the definition is completely unphysical.

The natural EDA and SAPT both rely on a perturbative treatment of the interaction energy, which fundamentally limits their applicability to weak interactions. Instead, many methods calculate variationally optimised intermediate wavefunctions, such as the Ziegler-Rauk approach for the Xα method [196]. The Kitaura-Morokuma EDA [187], an extension of that of Morokuma described above but with charge transfer explicitly defined, was the first of these, with the block-localised wavefunction [66], Sadlej [197], and ALMO methods following as natural extensions [2, 42]. Recently, there have been considerable efforts towards extending the latter EDA to correlated methods [100, 101, 198]. Similarly, one could impose a grid-based charge constraint, as proposed by Wu *et al.* [199], or proceed by partitioning the system [200, 201]. The latter in particular allows for the analysis of the movement of charge within a complex. Using the CHA ansätz, Mayer proposed a bond-order index [80, 110, 111], and then later a 'chemical energy component analysis' [80, 112]. These have the advantage of being able to be calculated either directly, or applied after a conventional SCF calculation has been carried out.

The main complaint about all of these analyses is the level of arbitrariness in defining the different components. Common to all of these is usually some form of polarisation and dispersion. However, other possibilities include (but are not limited to) frozen orbital contributions, induction, exchange, and charge transfer. None of these is strictly a physical observable, and so none has an associated operator, which makes them impossible to rigorously define. Perhaps the easiest to separate is the classical electrostatic terms, involving interactions between multipoles; however, even for these, there will necessarily be interactions with, in particular, exchange terms, as the system is fundamentally quantum mechanical. While the bulk of the energy is usually electrostatic in nature [10], arguments that the other terms are unimportant have time and again been shown to be false [11, 12, 14, 15, 84, 99, 202]. Putting aside that some complexes, such as those involving noble gases, are entirely dispersion bound [203], the electrostatics are usually less sensitive to small changes in environment than other terms, such that the differences between two closely related systems can only be described by consideration of these contributions. Perhaps some of the clearest examples of this can be found in halogen-bonded complexes, where unexpected changes in the mode of binding can be almost entirely attributed to charge transfer [12, 14].

As has already been noted, the importance of such dispersive and charge transfer terms necessitates both a high level treatment of electron correlation, and a rigorous way of separating the relevant contributions. There have recently been attempts to do this with local coupled cluster methods [204], but the cost for multi-fragment systems is still prohibitive. It would therefore be very useful to have an *ab initio* approach that is BSSE free, low-order scaling, and capable of a meaningful energy decomposition for a broad variety of chemical systems.

1.6 APPLICATIONS AND SUMMARY

The ability to accurately study intermolecular interactions in larger-scale systems, particularly in the condensed phases, is of wide interest in a number of disciplines. For example: the atomistic study of biomolecular systems [205, 206], which depend heavily on the solvent; solvated reaction mechanisms and computational spectroscopy [207]; and materials and crystal design [208, 209].

Each of these offers unique challenges, which I aim to address through the methods outlined in this thesis. The shared requirements to all applications are that it be efficient and accurate. The exact definitions of either of these terms depend on context, but in general, we wish to be able to routinely run calculations on systems of thousands of molecules at a high level of theory, e.g. equivalent to coupled cluster methods. The accuracy and efficiency should not be strongly dependent on the type of system, other than being specialised to non-covalent interactions in general. Chapters 2 and 3 outline the main theoretical developments of the method, with an emphasis on physical justifications for the choices, so as to maintain accuracy and robustness. Chapter 4 then describes the technical implementation, and the achievement of extreme efficiency, while chapter 5 demonstrates and validates the method and implementation on a wide range of benchmarking systems.

For the study of biomolecular systems, particularly in the context of *ab initio* molecular dynamics, we have more stringent requirements on the efficiency, but equally more leniency with the accuracy. For these, we need to be able to calculate both energies and forces on the order of minutes, as many thousands of timesteps will in general be needed. Accurate forces are, in fact, a wider requirement, as computational spectroscopy and materials design applications rely on being in optimal geometries. In general, the single-point calculations need to be efficient enough to allow for routine geometry optimisations to be carried out. Unlike the molecular dynamics, both these applications depend much more heavily on the accuracy of the method.

Ideally, it should be flexible enough to be able to tune accuracy versus efficiency. Chapter 6 is concerned with the derivation of analytical derivatives for each of the different theory levels of the method, and their demonstration on the optimisation of solvated systems. Chapter 8 gives a case study on the calculation of photoelectron spectra for water-solvated phenol.

Reaction mechanisms offer an additional difficulty, in that bonds are made and broken. The interactions with solvent are generally as spectators, however, so that an ideal extension of any approach would be to have a multiconfigurational subsytem, where the bonds are broken, interacting non-covalently with solvent molecules. Chapter 7 outlines the derivation of such a multiconfigurational variant of the new method, and a prescription for how it can be implemented effectively.

Finally, crystal design extends us from the solvated state, to the solid state. In this regime, periodicity and defects become very important. Of the applications listed in this section, this is the only possibility not directly addressed in this thesis. However, I do outline (in chapter 3) how our choice of correlated wavefunction method would easily allow for future extensions to the solid state.

2 | ABSOLUTELY LOCALISED MOLECULAR ORBITALS

In this chapter, the theory of absolutely localised molecular orbitals is outlined, setting the base for the rest of the thesis. The method was originally developed by Stoll et al. [1], then reformulated by several groups [2, 40–42]. Here, I present it for the first time using the formalism of second quantisation. This facilitates the addition in later chapters of dynamic and static electron correlation. After the theory is outlined, the problems inherent with the approximation are discussed, in particular the lack of charge transfer. Finally, the details of the new implementation are given, as found in my electronic structure package GAMMA.

2.1 BACKGROUND THEORY

The fundamental task in electronic structure theory is to solve the electronic Schrödinger equation. Herein we will focus on the non-relativistic Schrödinger equation for N electrons under the clamped nucleus and Born-Oppenheimer approximations. In the following, lowercase Latinate indices indicate electrons, while uppercase indicates the N_Z fixed nuclei (or, later on, spin orbitals) resulting in the following Hamiltonian:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{N_Z}\frac{Z_A}{R_{iA}} + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{1}{r_{ij}} \tag{2.1}$$

where R_{iA} and r_{ij} are the distances between electron i and nucleus A or electron j, respectively, and Z_A is the charge on nucleus A. We suppress the nuclear-nuclear repulsion term, $Z_A Z_B / R_{AB}$, as it is a constant within the clamped nucleus approximation. Atomic units will be used throughout, such that the electron mass, elementary charge, reduced Planck's constant and Coulomb's constant are all set to unity. For simplicity, from now on summations over i and A will be implied to be over the whole set of electrons or nuclei unless explicitly stated otherwise. The aim is then to solve the time-independent equation $\hat{H}\Psi = E\Psi$ to find both the energy E and electronic wavefunction Ψ.

2.1.1 Fock space and second quantisation

In the usual formulation, the observable - e.g. the energy - is represented by an operator ($\hat{H}$) and the state Ψ is considered to be a function. However, this hides a key requirement on a fermionic state: that of antisymmetry. It is possible to encode this through careful use of determinants, something which is expounded on elsewhere [34, 210], but an alternative is the formalism of second quantisation. In this, the state is represented in terms of creation and annihilation operators acting on a vacuum state. The algebra of these operators then encodes the antisymmetry of the wavefunction, simplifying much of the exposition.

It is conventional to represent a given system using a basis of M orthonormal one-electron spin orbitals $\{\phi_P\}$ in some Hilbert space, H, depending on the nuclear and spin coordinates, $\mathbf{x}_i$ of an electron i. For a system of N electrons, N such orbitals are combined into a Slater determinant:

The Hilbert space is usually taken to be $L^2(\mathcal{X}, d^4x)$, where $\mathcal{X} = \mathbb{R}^3 \times Spin(1/2)$.

$$|\phi_{P_1}\phi_{P_2}\cdots\phi_{P_N}| = \hat{A}\left[\phi_{P_1}(\mathbf{x_1})\phi_{P_2}(\mathbf{x_2})\cdots\phi_{P_N}(\mathbf{x_N})\right] \tag{2.2}$$

where the operator

$$\hat{A} = N^{-1/2}\sum_{\hat{\pi}\in S_N}\hat{\pi}$$

is an antisymmetriser, permuting electron coordinates according to the symmetric group S_N. The total N-electron wavefunction can then in principle be represented exactly by linear combinations of Slater determinants in the limit that the spin orbital basis tends to completion [210]. The space spanned by these Slater determinants is thus an antisymmetrised tensor product of N one-dimensional Hilbert spaces, $H \otimes H \otimes \cdots \otimes H = H^{\otimes N}$, over which the direct sum over all possible N forms an occupation space, known as Fock space:

$$F_- = \bigoplus_{N=0}^{\infty}\hat{A}H^{\otimes N} \tag{2.3}$$

In practice this is restricted to finite M, rather than taken to infinity.

Each point in Fock space is then given by the occupation numbers of the spin orbitals, i.e. by a vector

$$|\mathbf{n}\rangle = |n_1, n_2, \ldots, n_M\rangle, \, n_P = \begin{cases} 1, & \phi_P \text{ occupied,} \\ 0, & \phi_P \text{ unoccupied.} \end{cases} \tag{2.4}$$

In particular, there is a single zero-occupation state, $|0\rangle$, called the vacuum state, from which all other states can be reached by 'creating' particles. To do so, we introduce creation operators, $a_P^\dagger$, which produce an electron in orbital P. The inverse is the annihilation operator,

a_P. These operators satisfy a Lie algebra defined by the following anticommutation relations:

$$\{a_P^\dagger, a_Q^\dagger\} = \{a_P, a_Q\} = 0$$
$$\{a_P^\dagger, a_Q\} = \delta_{PQ}$$

(2.5)

where δ_{PQ} is the Kronecker delta. These ensure the antisymmetry of the state with respect to exchange of electron coordinates - i.e. permuting electron positions in the occupation vector results in a change of sign of the state.

Note that $\{x, y\} = xy + yx$

As such, any point in Fock space can be written as

$$|\mathbf{n}\rangle = \left[\prod_P \left(a_P^\dagger \right)^{n_P} \right] |0\rangle$$

(2.6)

In particular, attempting to add a second particle to an orbital, or to remove a non-existent particle, will annihilate the state, such that $a_P^\dagger a_P |\mathbf{n}\rangle = \delta_{n_P,1} |\mathbf{n}\rangle = n_P |\mathbf{n}\rangle$. This implicitly defines a number operator for each orbital, $\hat{n}_P = a_P^\dagger a_P$. The number operator is a subset of the more general class of elementary excitation operators:

$$X_{PQ} = a_P^\dagger a_Q$$

which can be seen to destroy the electron in orbital Q, 'exciting' it into orbital P. Clearly, if Q is empty or P already filled, this operator annihilates the state. Otherwise, we arrive at a different, 'excited' state in the N-dimensional Hilbert space via a connecting state in the $(N-1)$-dimensional space.

Assuming normal ordering, where the orbitals are in cardinal number order, it follows from the relations in equations 2.5 that both the annihilation and creation operators introduce a phase factor depending on where in the normal ordering the orbital they act on is placed. Specifically, for an operator acting on orbital P, an even number of permutations - i.e. an even number of electrons in spin orbitals $Q < P$ - results in no change, whereas an odd number flips the sign. This phase can be written as

$$\Gamma_P^{\mathbf{n}} = \prod_{Q=1}^{P-1} (-1)^{n_Q}$$

(2.7)

Thus, the action of the excitation operator on a state can be formalised as:

$$X_{PQ} |\mathbf{n}\rangle = \epsilon_{PQ} \Gamma_P^{\mathbf{n}} \Gamma_Q^{\mathbf{n}} (1 - n_P + \delta_{PQ}) n_Q |\mathbf{n} : n_P \to 1, n_Q \to \delta_{PQ}\rangle$$

(2.8)

where $\epsilon_{PQ} = 1$ if $P \leqslant Q$ and -1 otherwise.

2.1.2 The molecular Hamiltonian revisited

An inner product on a vector space is necessarily invariant to the representation of elements of that space. In quantum mechanics, this translates to the expectation value of an operator being invariant. This provides the key for translating operators to a second-quantised representation. In a spin-orbital presentation, these expectation values are given as a sum of matrix elements. The orthogonality of the spin orbitals then means that a k-electron operator only gives non-zero contributions between states that differ in the occupations of no more than k electrons, leading to a natural second-quantised representation in terms of k excitation operators. In particular, the molecular Hamiltonian of equation 2.1 comprises two one-electron operators and a two-electron operator, and so becomes:

$$\hat{H} = \sum_{PQ} \left(t_{PQ} + v_{PQ}\right) X_{PQ} + \frac{1}{2} \sum_{PQRS} g_{PQRS} a_P^\dagger a_R^\dagger a_S a_Q$$

where t, v, and g are the kinetic, nuclear-electronic potential, and electron-electron potential energy elements, respectively. These are evaluated in the spin orbital basis as follows:

$$t_{PQ} = -\frac{1}{2} \int \phi_P^*(\mathbf{x}) \nabla^2 \phi_Q(\mathbf{x}) d\mathbf{x}$$

$$v_{PQ} = -\sum_A Z_A \int \frac{\phi_P^*(\mathbf{x}) \phi_Q(\mathbf{x})}{|\mathbf{x} - \mathbf{R_A}|} d\mathbf{x} \qquad (2.9)$$

$$g_{PQRS} = \int \int \frac{\phi_P^*(\mathbf{x_1}) \phi_R^*(\mathbf{x_2}) \phi_Q(\mathbf{x_1}) \phi_S(\mathbf{x_2})}{|\mathbf{x_1} - \mathbf{x_2}|} d\mathbf{x_1} d\mathbf{x_2}$$

The two one-electron terms are normally combined into one, giving the so-called core Hamiltonian, with matrix elements $h_{PQ} = t_{PQ} + v_{PQ}$. The final molecular Hamiltonian is then

$$\hat{H} = \sum_{PQ} h_{PQ} X_{PQ} + \frac{1}{2} \sum_{PQRS} g_{PQRS} \left(\delta_{RS} X_{PQ} - X_{PS} X_{RQ}\right) \qquad (2.10)$$

where we have expanded the two electron term in terms of one-electron excitation operators using the anticommutation relations of equations 2.5. Note that the Kronecker delta is itself an idempotent one-electron operator, such that the two-electron character is maintained. The formulation in equation 2.10 both makes the following theoretical exposition simpler, and clarifies the meaning of the matrix elements: they are amplitudes connecting the various states in the N-electron segment of the Fock space via excitations. For example, $h_{PP} X_{PP}$, where the excitation operator is just the number operator, gives the on-site one-electron energy, whilst $h_{P \neq Q} X_{PQ}$ gives the one-electron energy difference between states after an excitation from orbital Q to orbital P. Roughly,

this corresponds to a probability weighting of the new state as a component in the superposition of states returned from the action of the Hamiltonian on the original state.

2.1.3 Density matrices and spin

The energy of any given reference state, $|\text{ref}\rangle$, is naturally given as the expectation value of $\hat{H}$ over this state. From equation 2.10 this is given by

$$E_{\text{ref}} = \langle\text{ref}|\hat{H}|\text{ref}\rangle = \sum_{PQ} h_{PQ}\overline{D}_{PQ} + \frac{1}{2}\sum_{PQRS} g_{PQRS}\overline{\Gamma}_{PQRS}$$

where we have defined the new matrix elements

$$\overline{D}_{PQ} = \langle\text{ref}|X_{PQ}|\text{ref}\rangle$$
$$\overline{\Gamma}_{PQRS} = \langle\text{ref}|\delta_{RS}X_{PQ} - X_{PS}X_{RQ}|\text{ref}\rangle \qquad (2.11)$$

For clarity, the overbars indicate these quantities are in the spin orbital basis. More generally, these matrix elements will appear in expectation values of any one- or two-electron operator. They thus encode all information about the state, i.e. the probability density, in the evaluation of the expectation value. As such, $\overline{D}$ and $\overline{\Gamma}$ are termed the one- and two-electron density matrices, respectively.

As the density matrices encode the state, they also contain all information about the electron spin. In our case, it is simpler to deal with this explicitly now. The spin orbitals, ϕ, comprise a spatial part and a spin part. As electrons are spin-1/2 particles, the spin part $\sigma(m_s)$ is either α (spin up) or β (spin down). A general spin orbital is then succinctly written as

In relativistic Hamiltonians, we would need to include both α and β parts, as the Hamiltonian may contain spin operators.

$$\phi_{P=p\sigma}(\mathbf{r}, m_s) = \phi_p(\mathbf{r})\sigma(m_s)$$

where $\alpha(1/2) = \beta(-1/2) = 1$ and $\alpha(-1/2) = \beta(1/2) = 0$. This means that, irrespective of the orthogonality of the spatial orbitals, spin orbitals of opposite spin are necessarily orthogonal. In general, this allows for the same spatial orbital to be used twice, such that we include the spin in a composite index, changing the relations in equations 2.5 to

$$\left\{a^\dagger_{p\sigma}, a_{q\tau}\right\} = \delta_{pq}\delta_{\sigma\tau} \qquad (2.12)$$

All the operators we will consider are spin free - that is, they operate only on the spatial part of the orbital. Therefore we can simplify our problem using the inherent orthogonality of the spin parts. Splitting the summations in equation 2.10 over α and β spins, then re-coupling based on equivalent spatial parts, the one-electron term can easily be

seen to be equivalent to replacing the spin-agnostic excitation opera-tor, $\hat{X}$, with a spin-free, singlet excitation operator:

$$E_{pq} = X_{p\alpha,q\alpha} + X_{p\beta,q\beta} \tag{2.13}$$

The two-electron term is more complicated, as we need to collect spins. We note that the matrix elements g_{PQRS} in equations 2.9 clearly disappear unless the spins are consistent for the two electron coor-dinates. Thus we need to consider terms of the form $a^{\dagger}_{p\sigma}a^{\dagger}_{r\tau}a_{s\tau}a_{q\sigma}$. Repeated applications of the anticommutation relations then yield

$$\sum_{\sigma\tau} a^{\dagger}_{p\sigma}a^{\dagger}_{r\tau}a_{s\tau}a_{q\sigma} = \sum_{\sigma\tau} X_{p\sigma,q\sigma}X_{r\tau,s\tau} - \delta_{rq}\delta_{\sigma\tau}X_{p\sigma,s\tau} = E_{pq}E_{rs} - \delta_{rq}E_{ps}$$

which we define as the two-electron spin-free excitation operator

$$\Sigma_{pqrs} = E_{pq}E_{rs} - \delta_{rq}E_{ps} \tag{2.14}$$

Note that this is similar to, but subtly different from, its spin-agnostic equivalent.

Using equations 2.13 and 2.14, we can find all necessary expectation values in the orbital basis by introducing the spin-free, orbital density matrices via equivalence with equations 2.11:

$$\begin{aligned} D_{pq} &= \langle \mathrm{ref}|E_{pq}|\mathrm{ref}\rangle \\ \Gamma_{pqrs} &= \langle \mathrm{ref}|\Sigma_{pqrs}|\mathrm{ref}\rangle \end{aligned} \tag{2.15}$$

2.1.4 Hartree–Fock theory and orbital rotations

As stated, the set of all N-electron determinants comprises a com-plete basis for the exact wavefunction. If we restrict ourselves to a sin-gle determinant, optimised with respect to the energy expectation in the orbital subspace, we arrive at the Hartree-Fock wavefunction [211–213]. We will not discuss the advantages and disadvantages of such a wavefunction choice, other than to point out that it typically gives electronic energies within a few percent of the exact energy; thus it is often used as the zeroth-order reference for more accurate electronic structure calculations. We restrict our attention to the closed-shell, restricted case for simplicity - the extension to the spin-unrestricted case is simple and given in detail elsewhere [210].

Starting with any initial guess determinant, $|\mathbf{n}\rangle$, from a given set of orbitals, we can arrive at any other single determinant state by unitar-ily rotating the orbitals. Any such operation can be represented using the exponentiation of an anti-Hermitian one-electron operator $\hat{\kappa}$; this then preserves orthonormality and spin [34]. Thus, we parametrise the desired determinant as

Effectively, we are trying to locate the optimal point on the unit hypersphere in the N-dimensional region of Fock space.

$$|\mathbf{n}(\hat{\kappa})\rangle = \exp(-\hat{\kappa})|\mathbf{n}\rangle \tag{2.16}$$

with $\hat{\kappa}$ given by

$$\hat{\kappa} = \sum_{p>q} \kappa_{pq} E_{pq}^{-} \tag{2.17}$$

For simplicity, we have defined antisymmetric singlet excitation operators $E_{pq}^{-} = E_{pq} - E_{qp}$. The variational parameters are then the matrix elements κ_{pq}.

Expanding the energy in terms of the rotation, we get

$$E(\kappa) = \langle \mathbf{n}(\kappa)|\hat{H}|\mathbf{n}(\kappa)\rangle = E^{(0)} + \kappa^{T} \mathbf{E}^{(1)} + \frac{1}{2}\kappa^{T} \mathbf{E}^{(2)}\kappa + \cdots \tag{2.18}$$

where $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ are the electronic gradient and Hessian, respectively. By inserting equation 2.16 into the above and using the Baker-Campbell-Hausdorff expansion, we can identify the gradient via

$$E(\kappa) = \langle \mathbf{n}|\exp(\hat{\kappa})\hat{H}\exp(-\hat{\kappa})|\mathbf{n}\rangle = \langle \mathbf{n}|\hat{H}|\mathbf{n}\rangle + \langle \mathbf{n}|\left[\hat{\kappa}, \hat{H}\right]|\mathbf{n}\rangle + \frac{1}{2}\langle \mathbf{n}|\left[\hat{\kappa}, \left[\hat{\kappa}, \hat{H}\right]\right]|\mathbf{n}\rangle \cdots$$

Then, using equation 2.17, we get that the elements of the electronic gradient and Hessian are given by

$$\begin{aligned} E_{pq}^{(1)} &= \langle \mathbf{n}|\left[E_{pq}^{-}, \hat{H}\right]|\mathbf{n}\rangle \\ E_{pqrs}^{(2)} &= (1 + \hat{\pi}_{pq,rs})\langle \mathbf{n}|\left[E_{pq}^{-}, \left[E_{rs}, \hat{H}\right]\right]|\mathbf{n}\rangle \end{aligned} \tag{2.19}$$

where $\hat{\pi}_{pq,rs}$ is the projector swapping indices pq and rs.

The Hartree-Fock wavefunction is obtained by taking the variation of the energy in equation 2.18, with the stationary point defined by $E = \mathcal{O}\left(\kappa^2\right)$. With respect to the parameters, this variational condition is equivalent to

$$\delta E = 0 \implies E_{pq}^{(1)} = 0 \; \forall \; p, q$$

There are many possible solutions to this, however, and not all are necessarily minima. In particular, many of the κ_{pq} parameters are *redundant*, which we here take the convention of defining as any parameter where the corresponding operator, E_{pq}^{-} annihilates the wavefunction; these are then redundant in the energy expression, as their contribution to the expectation values is necessarily zero.

The operator $\hat{\kappa}$ was by construction anti-symmmetric, to maintain the anti-symmetry of the wavefunction; this automatically renders the diagonal elements $\kappa_{pp} = 0$, making them trivially redundant. For all other parameters, we designate them as corresponding to inactive (indices i, j, k), active (indices u, v, w), or virtual (indices a, b, c) orbitals. The latter refer to the unoccupied orbitals not contained in the state $|\mathbf{n}\rangle$, inactive orbitals are doubly occupied, while active orbitals are singly occupied, such that a closed-shell wavefunction only has inactive and virtual orbitals. The excitation operators E_{ij} necessarily annihilate doubly-occupied orbitals unless $i = j$, as they try to create an electron in the orbital, thus inactive-inactive parameters must be

The big-Oh notation, $M = \mathcal{O}(N)$ implies that asymptotically, M is less than some fixed multiple of N; a similar, little-oh notation $M = \wr(N)$ would mean M is asymptotically greater than a fixed multiple of N.

redundant; the same argument applies to virtual-virtual rotations, as we try to destroy a non-existent electron. Consider the inactive-virtual terms, E_{ai}^-: one of the terms destroys an occupied orbital's electron and excites it into a virtual orbital, which will in general be non-zero; the other term destroys an electron in an unoccupied virtual orbital, thus annihilating. As such, the inactive-virtual parameters are in general non-redundant. In a later chapter, we will need to consider rotations involving active orbitals, but for simplicity they can be ignored for now. For a closed-shell state, the only non-redundant parameters are of the form κ_{ai}.

Using this information, we can write the variational conditions on the gradient, using equation 2.19, as

$$E_{ai}^{(1)} = \langle \mathbf{n} | \left[E_{ai} - E_{ia}, \hat{H} \right] | \mathbf{n} \rangle = 2 \langle \mathbf{n} | \left[E_{ai}, \hat{H} \right] | \mathbf{n} \rangle = 0 \qquad (2.20)$$

where we have expanded the commutator, used the Hermiticity of the expectation value, and noted that

$$\langle \mathbf{n} | E_{ai} = E_{ia} | \mathbf{n} \rangle = 0$$

The interpretation of equation 2.20 is that the variational Hartree-Fock wavefunction does not interact (through the Hamiltonian) with any singly-excited states; this is know as the Brillouin theorem. In fact, it is simple to demonstrate this applies to the redundant terms as well, leading to the generalized Brillouin theorem [214], stating that the HF state comprises a "perfect balance" of excitations and de-excitations.

2.1.5 The canonical solution

The construction of the HF wavefunction is through the variational determination of a single Slater determinant. As such, it represents N independent particles, where the only correlation present is that between particles of the same spin, via the antisymmetry enforced on the wavefunction; this correlation is referred to as exchange or Fermi correlation. There is no dynamical correlation, implying that the problem should be identifiable with an effective one-electron problem, described by an effective one-electron Hamiltonian called the Fock operator, or Fockian. As noted above, there are in general many possible solutions; moreover the redundancy inherent in the parametrisation means there are many different orbital representations of the same solution. Our choice of Fock operator then determines the nature of our solution. I will describe in some detail here the construction of the *canonical* Fock operator, so as to clarify and inform the construction of the *absolutely localised* Fockian later.

A one-electron Hamiltonian has some standard requirements: it must be Hermitian and totally symmetric in spin, and must equal the

true Hamiltonian in the non-interacting limit; these requirements are necessary to ensure that the Fock operator connects analytically to the Hamiltonian. These conditions can be ensured by requiring that it take the form

$$\hat{F} = f_{pq} E_{pq}$$

where $f_{pq} = f_{qp}$, and $f_{pq} = h_{pq}$ in the non-interacting limit. The canonical choice is that the variational orbitals diagonalise the Fock operator, i.e. the matrix elements take the form $f_{pq} = \epsilon_p \delta_{pq}$, where ϵ_p is called the orbital energy of orbital p. This requirement needs to be equivalent to the variational conditions of equation 2.20, leading to the natural mapping that

$$f_{ai} = f_{ia} \propto \langle \mathbf{n} | \left[E_{ai}, \hat{H} \right] | \mathbf{n} \rangle$$

However, this identification will not work for the occupied-occupied and virtual-virtual blocks, as they would not tend towards the core hamiltonian in the non-interacting limit. We can make special use of the structure of the wavefunction (that it is closed-shell and real) to rearrange the above into a more suitable form, however:

$$f_{ai} \propto \sum_{\sigma} \langle \mathbf{n} | \left[a_{a\sigma}^{\dagger} a_{i\sigma}, \hat{H} \right] | \mathbf{n} \rangle = - \sum_{\sigma} \langle \mathbf{n} | \left\{ a_{i\sigma}^{\dagger}, \left[a_{a\sigma}, \hat{H} \right] \right\} | \mathbf{n} \rangle$$

This form does not annihilate when generic indices are used, is symmetric, and, as will be shown shortly, reduces to the correct form when two-electron terms vanish, such that we choose to identify the Fock matrix elements as

$$f_{pq} = \frac{1}{2} \sum_{\sigma} \langle \mathbf{n} | \left\{ a_{p\sigma}^{\dagger}, \left[a_{q\sigma}, \hat{H} \right] \right\} | \mathbf{n} \rangle \tag{2.21}$$

with the normalisation factor of $-1/2$ chosen for convenience. Note that this form shows the non-linearity of the HF equations: the Fock elements are constructed from the orbitals, themselves defined as the eigenvectors of the Fock operator.

The matrix element can be expanded by repeated use of the following:

A very useful identity:
$$[ab, c] = a[b, c] + [a, c]b$$
$$= a\{b, c\} - \{a, c\}b.$$

$$[a_{q\sigma}, E_{rs}] = \sum_{\tau} \left[a_{q\sigma}, a_{r\tau}^{\dagger} a_{s\tau} \right]$$

$$= \sum_{\tau} \left(a_{r\tau}^{\dagger} \{ a_{q\sigma}, a_{s\tau} \} - \left\{ a_{q\sigma}, a_{r\tau}^{\dagger} \right\} a_{s\tau} \right)$$

$$= \sum_{\tau} \delta_{qr} \delta_{\sigma\tau} a_{s\tau} = \delta_{qr} a_{s\sigma}$$

so that the anticommutator in equation 2.21 reduces to give:

$$\sum_{\sigma} \left\{ a_{p\sigma}^{\dagger}, \left[a_{q\sigma}, \hat{H} \right] \right\} = 2h_{pq} + \sum_{\sigma\tau} g_{qsrt} \left\{ a_{p\sigma}^{\dagger}, a_{r\tau}^{\dagger} a_{t\tau} a_{s\sigma} \right\}$$

$$= 2h_{pq} + (2g_{pqrs} - g_{psrq}) E_{rs}.$$

Thus, we arrive at the following result for equation 2.21:

$$f_{pq}^{\text{can.}} = h_{pq} + \frac{1}{2}\left(2g_{pqrs} - g_{psrq}\right)D_{rs} = h_{pq} + 2g_{pqii} - g_{piiq} \qquad (2.22)$$

where we have used the fact that the one-electron density for the canonical, closed-shell system is $2\delta_{ij}$ for occupied orbitals, and zero otherwise.

Note that these *canonical* conditions, defined as such, are now over the whole set of orbitals, removing the redundancy found in the *variational* conditions of equation 2.20. This is equivalent to saying the degrees of freedom in the solution have been removed by choosing orbitals that diagonalise the Fockian, thus the choice of Fock operator has defined the solution. Finally, using this and the density-matrix definition of the energy expectation and the above yields

$$
\begin{aligned}
E^{\text{can.}} &= h_{pq}D_{pq} + \frac{1}{2}g_{pqrs}\left(D_{pq}D_{rs} - \frac{1}{2}D_{ps}D_{rq}\right) \\
&= 2h_{ii} + 2g_{iijj} - g_{ijji} = \text{Tr}\{\mathbf{h} + \mathbf{f}^{\text{can.}}\}
\end{aligned}
\qquad (2.23)
$$

2.1.6 The Roothaan–Hall equations

So far, the problem has been treated as a continuous one - only a finite number of molecular orbitals (MOs) are needed, but the form of these orbitals is undefined. However, to solve the problem numerically, we need to discretise it. This is done by expanding the molecular orbitals in a known, finite basis, typically atom-centred and thus known as the atomic orbital (AO) basis. The exact form of this basis only matters insofar as it affects the evaluation of integrals, which does not concern us presently, and so will not be discussed at this point. We will denote the atomic basis as $\{\chi_\mu\}$, using Greek indices to distinguish atomic from molecular orbitals. The expansion is in terms of linear coefficients, $C_{\mu i}$, as follows:

$$\phi_p = C_{p\mu}^\dagger \chi_\mu \qquad (2.24)$$

For a generic one-electron operator $\hat{O}$, we thus have that

$$o_{pq} = \langle \phi_p | \hat{O} | \phi_q \rangle = C_{p\mu}^\dagger \langle \chi_\mu | \hat{O} | \chi_\nu \rangle C_{\nu q} = \left[\mathbf{C}^\dagger \mathbf{O} \mathbf{C}\right]_{pq} \qquad (2.25)$$

Of particular interest are the overlap metric and the one-electron density. In the canonical theory, the former is the identity as the MOs are required to be orthonormal; the AOs, however, are in general nonorthogonal, with overlap matrix $\mathbf{S}$. The orthonormality condition can then be written in the AO basis using equation 2.25 as

$$s_{pq} = C_{p\mu}^\dagger S_{\mu\nu} C_{\nu q} = \delta_{pq} \qquad (2.26)$$

while the spin-free density operator from equation 2.15 yields

$$\rho_{pq} = \begin{cases} 2\delta_{pq}, & \text{if p occupied,} \\ 0, & \text{otherwise.} \end{cases} = C^\dagger_{p\mu} P_{\mu\nu} C_{\nu q}$$

This latter equation implies the transformation matrix is unitary, allowing us to conclude that the AO density matrix is given by

$$P_{\mu\nu} = 2C_{\mu i}\delta_{ij}C^\dagger_{j\nu} = 2C_{\mu i}C^\dagger_{i\nu} \tag{2.27}$$

Using this in equation 2.23 gives the energy as

$$E^{\text{can.}} = \text{Tr}\left\{\mathbf{C}^\dagger\left(\mathbf{H} + \mathbf{F}^{\text{can.}}\right)\mathbf{C}\right\} = \frac{1}{2}\text{Tr}\left\{\left(\mathbf{H} + \mathbf{F}^{\text{can.}}\right)\mathbf{P}\right\} \tag{2.28}$$

where we have used the permutational invariance of the trace.

As we have constructed the Fockian to satisfy the variational conditions in equation 2.20, the numerical problem is to fit the atomic basis to the variational MOs, subject to the condition that the MOs should remain orthonormal (in the canonical case). To do this, we introduce a Lagrangian using equations 2.26 and 2.28:

$$\mathcal{L}^{\text{can.}} = \frac{1}{2}\text{Tr}\left\{[\mathbf{H} + \mathbf{F}^{\text{can.}}(\mathbf{P})]\mathbf{P}\right\} - \lambda_{pq}\left(C^\dagger_{p\mu}S_{\mu\nu}C_{\nu q} - \delta_{pq}\right) \tag{2.29}$$

We minimise this by requiring that it be stationary with respect to the coefficients, $\partial\mathcal{L}/\partial C_{\mu i} = 0$, and the Lagrange multipliers, $\partial\mathcal{L}/\partial\lambda_{pq} = 0$, with the latter reducing to equation 2.26.

To differentiate with respect to the coefficients, we substitute the expansion in equation 2.24 into the integrals from equation 2.9, before using these in to equation 2.22 to get the AO-basis canonical Fock matrix:

$$F^{\text{can.}}_{\mu\nu} = H_{\mu\nu} + \left(2G_{\mu\nu\lambda\tau} - G_{\mu\tau\lambda\nu}\right)P_{\tau\lambda} \tag{2.30}$$

With everything written in terms of the density, we simply note from equation 2.27, assuming the coefficients are real, that

$$\partial_{\gamma k}P_{\mu\nu} = \partial_{\gamma k}C_{\mu i}C_{\nu i} = \delta_{\gamma\mu}C_{\mu k} + \delta_{\gamma\nu}C_{\nu k} \tag{2.31}$$

such that, assuming the Lagrange multipliers are real and symmetric,

$$\begin{aligned}
\partial_{\gamma k}\mathcal{L}^{\text{can.}} &= \frac{\partial E^{\text{can.}}}{\partial P_{\mu\nu}}\partial_{\gamma k}P_{\mu\nu} - \lambda_{kq}S_{\gamma\nu}C_{\nu q} - \lambda_{pk}C^\dagger_{p\mu}S_{\mu\gamma} \\
&= \frac{1}{2}\left\{H_{\mu\nu} + F^{\text{can.}}_{\mu\nu} + \frac{\partial F^{\text{can.}}_{\eta\tau}}{\partial P_{\mu\nu}}P_{\tau\eta}\right\}\partial_{\gamma k}P_{\mu\nu} - 2S_{\gamma\mu}C_{\mu p}\lambda_{pk} \\
&= \left(H_{\gamma\nu} + F^{\text{can.}}_{\gamma\nu}\right)C_{\nu k} + \left(2G_{\eta\tau\nu\gamma} - G_{\eta\gamma\nu\tau}\right)P_{\tau\eta}C_{\nu k} - 2S_{\gamma\mu}C_{\mu p}\lambda_{pk} \\
&= 2\left[\mathbf{F}^{\text{can.}}\mathbf{C} - \mathbf{S}\mathbf{C}\lambda\right]_{\gamma k}
\end{aligned}$$

Then setting this equal to zero to find the stationary point of the Lagrangian yields the Roothaan-Hall equations:

$$\mathbf{F}^{\text{can.}}\mathbf{C} = \mathbf{SC}\lambda \tag{2.32}$$

This is a generalised eigenvalue equation, easily solved by diagonalisation of MO-basis Fock matrix; however, the Fockian itself depends on the coefficients, such that the new eigenvectors must be substituted back into the equation, continuing until self-consistency is reached.

2.2 ABSOLUTELY LOCALISED APPROXIMATION

As noted earlier, the canonical solution above is one of many possible forms of solution. In many ways it is the simplest in terms of formalism, which is why I have described it fully. However, the requirement for orthonormality of the molecular orbitals results in MOs delocalised over the entire system; this leads to unfavourable scaling of post-HF methods [70], and in the introduction of basis set superposition errors in supermolecular calculations. The HF energy is invariant to rotations within the occupied and virtual subspaces, but localisation of the orbitals in this manner is difficult and arbitrary [88]. Instead, it would be useful to localise the orbitals from the beginning. In this section, we will follow the exposition of the previous section, but relaxing the requirement for orthonormality, and instead requiring *absolutely localised* molecular orbitals (ALMOs), where each orbital is constrained to a single molecule in the supersystem [1, 42]. In principle, there is no need to restrict our choice of fragments to molecules, and certainly coarser definitions could be used with multiple molecules, but defining the fragments across covalent bonds would result in unacceptable errors (see section 2.3). This approach yields a succinct development of ALMOs in a formalism that for the first time will allow the development of sophisticated post-HF treatments of the correlation energy, both dynamical and static, as will be described in Chapters 3 and 7, respectively.

The first step is to remove the need for orthonormality - that is, the off-diagonal elements of the overlap metric, s_{pq}, $p \neq q$, are not necessarily zero. This requires new creation and annihilation operators, $b_{p\sigma}^{\dagger}$ and $b_{q\tau}$, which follow the same rules as those in equation 2.5, except the mixed anticommutation relation now reads

$$\left\{b_{p\sigma}^{\dagger}, b_{q\tau}\right\} = s_{pq}\delta_{\sigma\tau} \tag{2.33}$$

As discussed in section 2.1.4, however, we only require orthonormality within the occupied subspace, and between the virtual and occupied

spaces. As such, we can immediately use the above formalism by introducing a transformed set of operators. We let s_{OO} be the occupied-occupied block of the overlap metric, then define our creation and annihilation operators such that

$$a_{i\sigma}^{\dagger} = \left[s_{OO}^{-1/2}\right]_{ij} b_{j\sigma}^{\dagger} \tag{2.34}$$

The existence of the inverse square-root of s_{OO} is implied by the fact the overlap must be positive-definite.

Equation 2.34 clearly implies orthonormality within the occupied subspace, as

$$\langle |a_i a_j^{\dagger}| \rangle = \langle |b_k \left[s_{OO}^{-1/2}\right]_{ki} \left[s_{OO}^{-1/2}\right]_{jl} b_l^{\dagger}| \rangle = \left[s_{OO}^{-1/2}\right]_{jk} [s_{OO}]_{kl} \left[s_{OO}^{-1/2}\right]_{li} = \delta_{ij}$$

From this, we ascertain via equation 2.15 that the density operator, $\hat{\rho}$, is simply given by

$$\hat{\rho} = \sum_{\sigma} a_{i\sigma}^{\dagger} a_{i\sigma} = \sum_{\sigma} b_j^{\dagger} \left[s_{OO}^{-1/2}\right]_{ji} \left[s_{OO}^{-1/2}\right]_{ik} b_k = \sum_{\sigma} b_j^{\dagger} \left[s_{OO}^{-1}\right]_{jk} b_k \tag{2.35}$$

We can then ensure separation of the occupied-virtual subspaces by projecting the virtual subspace out:

$$a_{a\sigma}^{\dagger} = \left(\hat{1} - \hat{\rho}\right) b_{a\sigma}^{\dagger} = \hat{q} b_{a\sigma}^{\dagger} \tag{2.36}$$

where we have defined the complementary projector, $\hat{q} = \hat{1} - \hat{\rho}$.

2.2.1 Redundancy and the non-orthogonal solution

By introducing non-orthogonality, we have essentially mixed the orbital rotation parameters of equation 2.17; by projecting out the virtual subspace, we have returned to the instance where the occupied-virtual parameters are non-redundant. More consideration should be given to the virtual-virtual parameters, as these orbitals are not orthonormal any more, but these do not affect the HF solution, so will be considered in the next chapter.

We thus can directly use the variational conditions from equation 2.20, and in turn construct the Fockian in the same way. By the same logic, the redundancy considerations on the transformation parameters are the same. However, in general, both conditions in the non-orthogonal basis are much more complicated, such that the Generalized Brillouin Theorem is not satisfied if any approximations are introduced in this basis. This is a key consideration when we introduce the absolutely localised formalism in the AO basis shortly. We can see this by transforming equation 2.20 back to the non-orthogonal basis; this is more obvious if we cast it in terms of the individual MOs. For this purpose,

we denote the untransformed orbitals as ϕ_p and the transformed ones as $\tilde{\phi}_p$.

$$
\begin{aligned}
E_{ai}^{(1)} &= 2 \langle \tilde{\phi}_a | \sum_\sigma \left[a_{a\sigma}^\dagger a_{i\sigma}, \hat{H} \right] | \tilde{\phi}_i \rangle \\
&= 4 \langle \hat{q}\phi_a | \left[\hat{q} b_a^\dagger b_i, \hat{H} \right] | \phi_k \rangle \left[s_{OO}^{-1/2} \right]_{kj} \left[s_{OO}^{-1/2} \right]_{ji} \\
&= 4 \langle \phi_a | \hat{q} \hat{f} | \phi_k \rangle \left[s_{OO}^{-1} \right]_{ki} = 0
\end{aligned}
\tag{2.37}
$$

where in the last line we have used the idempotency of the projector $\hat{q}$, and the definitions of the Fock operator. Similarly, we could convert equations 2.22 and 2.23 into the non-orthogonal basis. The latter can then be written in the particularly useful form

$$
E^{\text{non-orthog.}} = \langle \phi_i | \hat{h} + \hat{f} | \phi_j \rangle \left[s_{OO}^{-1} \right]_{ji}
\tag{2.38}
$$

2.2.2 Localisation via the atomic basis

The non-orthogonal treatment is only different from the canonical one insofar as it affects the form that the molecular orbitals take in the final HF wavefunction. As discussed, however, the HF energy is invariant to such orbital rotations, such that physically nothing is changing, only the abstract mathematical representation of the problem. The absolutely localised approximation is introduced by enforcing these MOs to be formed exclusively from the AO basis on each molecule. As these bases are strictly atom centred, the resulting MOs are themselves absolutely localised.

Consider a system of F fragments. Each fragment, X, has an atomic orbital basis, $\{\chi_{X\mu}\}$, associated with it, containing n_X functions. From these, o_X occupied and v_X virtual molecular orbitals, ϕ_{Xp}, will be formed. The total supermolecular basis then comprises the union of all such fragment bases, with $N = \sum_X n_X$ AOs, $O = \sum_X o_X$ occupied orbitals, and $V = \sum_X v_X$ virtual orbitals. Throughout the following, we define

$$
n/o/v = \max_{X=1,2,\dots,F} \{n_X/o_X/v_X\}
\tag{2.39}
$$

Tensor notation is used throughout, as is the Einstein summation convention with the exception of sums over fragments, which will always be explicitly shown for clarity. As is usual, lowered (subscripted) indices indicate covariant quantities, while raised (superscripted) indices represent contravariant quantities. Dots are used as placeholder indices given the need to use composite indices to denote the fragmentation; for example, Xi is a single index. A primer on tensors can be found in Appendix A.

Following the approach in equation 2.24, the covariant MOs are expanded in the basis of AOs, but in this case with these being fragment-localised:

$$|\phi_{Xi}\rangle = |\chi_{X\mu}\rangle T^{X\mu\cdot}_{\cdot Xi}$$
$$|\phi_{Xa}\rangle = |\chi_{X\mu}\rangle V^{X\mu\cdot}_{\cdot Xa}$$

(2.40)

where the coefficient tensors, T and V, are constrained such that

$$T^{X\mu\cdot}_{\cdot Yi} = V^{Y\mu\cdot}_{\cdot Xi} = 0 \text{ if } X \neq Y$$

(2.41)

The overall coefficient matrix, $\mathbf{C}$, is formed from the concatenation of $\mathbf{T}$ and $\mathbf{V}$. The above constraint can thus be seen to remove the redundancy in the occupied-occupied (OO) and virtual-virtual (VV) rotational parameters, but only *between* fragments. The on-fragment OO and VV rotations are still redundant, by the same arguments as in the previous section, thus allowing us to enforce orthonormality within each fragment, i.e. $\langle \phi_{Xi}|\phi_{Xj}\rangle = \delta_{ij}$.

In the language of tensors, we then recognise that the overlap, s_{OO}, acts as the metric for transformations between co- and contra-variant tensors in the ALMO basis. We can also simplify notation surrounding the inverse transformation by whether the indices are raised or lowered. For clarity, we write

$$\sigma_{Yj,Xi} = [s_{OO}]_{ij} = \langle \phi_{Yj}|\phi_{Xi}\rangle = T^{\cdot Y\mu}_{Yj\cdot} S_{\mu\nu} T^{X\nu\cdot}_{\cdot Xi}$$
$$\sigma^{Yj,Xi} = \left[s_{OO}^{-1}\right]_{ij}$$

(2.42)

where $\mathbf{S}$ is the AO overlap matrix. As a result, the contravariant MOs are given by $|\phi^{Xi}\rangle = \sum_Y |\phi_{Yj}\rangle \sigma^{Yj,Xi}$, such that

$$\langle \phi_{Xi}|\phi^{Yj}\rangle = \sum_Z \langle \phi_{Xi}|\phi_{Zk}\rangle \sigma^{Zk,Yj} = \sum_Z \sigma_{Xi,Zk}\sigma^{Zk,Yj} = \delta^{Yj}_{Xi}$$

where δ^{Yj}_{Xi} is the Kronecker delta. This then allows us to rewrite the variational condition in equation 2.37 and energy expression in equation 2.38 as

$$E^{(1)}_{ai} = 4\langle \phi_a|\hat{q}\hat{f}|\phi^i\rangle = 0$$

(2.43)

$$E^{ALMO} = \sum_X \langle \phi_{Xi}|\hat{h} + \hat{f}|\phi^{Xi}\rangle$$

(2.44)

Finally, following the development of the canonical case, we can cast this energy in the same form as equation 2.28 by first finding the one-particle reduced density matrix, $\mathbf{P}$, in the AO basis using equation 2.35:

$$\sum_W |\phi^{Wi}\rangle \langle \phi_{Wi}| = \sum_{W,Z} |\chi_{Z\tau}\rangle T^{Z\tau\cdot}_{\cdot Zj} \sigma^{Zj,Wi} T^{\cdot W\eta}_{Wi\cdot} \langle \chi_{W\eta}|$$

whence using the locality condition from equation 2.41 and the orthonormality of orbitals within a fragment, we conclude that

$$P^{X\mu,Y\nu} = 2T^{X\mu\cdot}_{\cdot Xi} \sigma^{Xi,Yj} T^{Y\nu}_{Yj\cdot}$$

(2.45)

2.2.3 The modified Roothaan–Hall equations

We can now set up a Lagrangian equivalent to that of equation 2.29. There are now two auxiliary conditions, however, requiring two sets of Lagrange multipliers: we will use $\lambda^{Xp,Xq}$ for the on-fragment orthonormality condition, and $\gamma^{Yp\cdot}_{\cdot X\mu}$ for the localisation condition in equation 2.41. The Lagrangian is thus:

$$\mathcal{L}^{\text{ALMO}} = \frac{1}{2}\text{Tr}\left\{[\mathbf{H} + \mathbf{F}(\mathbf{P})]\,\mathbf{P}\right\} - \sum_X\sum_{Y\neq X}\gamma^{Yp\cdot}_{\cdot X\mu}C^{X\mu\cdot}_{\cdot Yp}$$
$$- 2\sum_X\lambda^{Xp,Xq}\left(C^{\cdot X\mu}_{Xp\cdot}S_{X\mu,X\nu}C^{X\nu\cdot}_{\cdot Xq} - \delta_{pq}\right) \tag{2.46}$$

The factor of two in the final term is for convenience later on, and could be absorbed into the Lagrange multipliers. Notice that when optimising for the occupied orbitals, which are fragment localised, the second term is not relevant, and so appears redundant. However, if we were to allow a full variational optimisation of all the orbitals (arriving back at full Hartree-Fock), they would be applicable as we would take the variation with respect to cross-fragment coefficients. This highlights the missing part of the ALMO method compared to canonical HF - that is, the lack of charge transfer.

Taking the variation with respect to a coefficient, $T^{X\mu\cdot}_{\cdot Xk}$, can be done in much the same way as in the lead-up to equation 2.32. However, there is the added complication that the density, equation 2.45, contains the inverse metric, which itself depends upon the coefficients. Therefore the variation of the density changes from equation 2.31, to

$$\frac{1}{2}\partial_{X\mu,Xk}P^{Y\kappa,Z\tau} = \delta_{Y\kappa,X\mu}\sigma^{Xk,Zj}T^{\cdot Z\tau}_{Zj\cdot} + \delta_{Z\tau,X\mu}T^{Y\kappa\cdot}_{\cdot Yi}\sigma^{Yi,Xk}$$
$$+ T^{Y\kappa\cdot}_{\cdot Yi}\partial_{X\mu,Xk}\sigma^{Yi,Zj}T^{\cdot Z\tau}_{Zj\cdot}$$

To calculate the variation in the inverse metric requires use of the fact that, for any positive-definite operator S, $\delta S^{-1} = -S^{-1}(\delta S)S^{-1}$, which can be proved using the Spectral Theorem [215]. Thus

$$\delta\sigma^{-1} = -\sigma^{-1}(\delta\sigma)\sigma^{-1} = -\sigma^{-1}\left[\delta\mathbf{T}^\dagger\mathbf{S}\mathbf{T} + \mathbf{T}^\dagger\mathbf{S}\delta\mathbf{T}\right]\sigma^{-1}$$

Inserting this into the density variation and simplifying then yields

$$\frac{1}{2}\partial_{X\mu,Xk}P^{Y\kappa,Z\tau} = \delta_{Y\kappa,X\mu}\sigma^{Xk,Zj}T^{\cdot Z\tau}_{Zj\cdot} - T^{Y\kappa\cdot}_{\cdot Yi}\sigma^{Yi,Xk}S_{X\mu,W\eta}T^{W\eta\cdot}_{\cdot Wm}\sigma^{Wm,Zj}T^{\cdot Z\tau}_{Zj\cdot}$$
$$+ \delta_{Z\tau,X\mu}T^{Y\kappa\cdot}_{\cdot Yi}\sigma^{Yi,Xk} - T^{Y\kappa\cdot}_{\cdot Yi}\sigma^{Yi,Wm}T^{\cdot W\eta}_{Wm\cdot}S_{W\eta,X\mu}\sigma^{Xk,Zj}T^{\cdot Z\tau}_{Zj\cdot} \tag{2.47}$$

Upon taking the trace with any Hermitian matrix, $\mathbf{M}$, this simplifies considerably:

$$\frac{1}{2}\sum_{Y,Z}M_{Z\tau,Y\kappa}\partial_{X\mu,Xk}P^{Y\kappa,Z\tau} = 2\left[(\mathbf{I} - \mathbf{SP})\mathbf{MT}\sigma^{-1}\right]_{X\mu,Xk}$$

From this we get, in exactly the same way we reached equation 2.32, that

$$\partial_{X\mu,Xk}\mathcal{L}^{\text{ALMO}} = 4\left[(\mathbf{I}-\mathbf{SP})\mathbf{FT}\sigma^{-1}-\mathbf{ST}\lambda\right]_{X\mu,Xk} = 0 \qquad (2.48)$$

We could then right-multiply by $\sigma_{Xk,Xl}$ and sum over fragments, to arrive at a generalised eigenproblem. However, the solution of this would then scale exactly the same as canonical HF, and would provide no computational advantages. Instead, it is more useful to return to equation 2.43 and make use of the fragment localisation. Equation 2.48 will still prove useful, however, in later chapters.

2.2.4 Fragment Fock operators

Returning to equation 2.35 in the ALMO notation, we get

$$\hat{\rho} = \sum_X |\phi^{Xi}\rangle\langle\phi_{Xi}| = \sum_X \hat{\rho}_X \qquad (2.49)$$

implicitly defining fragment-localised density operators, $\hat{\rho}_X$. Consider the action of this on any given co- or contra-variant occupied orbital:

$$\hat{\rho}_X|\phi_{Yk}\rangle = |\phi^{Xi}\rangle\langle\phi_{Xi}|\phi_{Yk}\rangle = |\phi^{Xi}\rangle\sigma_{Xi,Yk}$$
$$\hat{\rho}_X|\phi^{Yk}\rangle = |\phi^{Xi}\rangle\langle\phi_{Xi}|\phi^{Yk}\rangle = \delta_{XY}|\phi^{Xk}\rangle$$

This means that we can use it to localise the variational conditions to each fragment. If we also note that for any occupied orbital, $\hat{q}$ annihilates it, equation 2.43 can be rewritten as

$$E^{(1)}_{Ya,Xi} = 4\langle\phi_{Ya}|\hat{q}\hat{f}(\hat{q}+\hat{\rho}_X)|\phi^{Xi}\rangle = 0$$

which we can make hold for all Ya by requiring that

$$\hat{q}\hat{f}(\hat{q}+\hat{\rho}_X)|\phi^{Xi}\rangle = 0$$

for all Xi. Noting that $\hat{q}$ is Hermitian, we can symmetrise this to give

$$\hat{f}^X|\phi^{Xi}\rangle = (\hat{q}+\hat{\rho}_X)^\dagger\hat{f}(\hat{q}+\hat{\rho}_X)|\phi^{Xi}\rangle = \hat{\rho}_X^\dagger\hat{f}\hat{\rho}_X|\phi^{Xi}\rangle \qquad (2.50)$$

which defines fragment Fock operators, $\hat{f}^X$.

Casting equation 2.50 into the AO basis on that fragment then gives:

$$\langle\chi_{X\mu}|\hat{f}^X|\phi^{Xi}\rangle = \left[\mathbf{F}^X\mathbf{T}\sigma^{-1}\right]_{X\mu,Xi}$$
$$\langle\chi_{X\mu}|\hat{\rho}_X^\dagger\hat{f}\hat{\rho}_X|\phi^{Xi}\rangle = \left[\mathbf{S}_{XX}\mathbf{F}_{XX}\mathbf{T}\sigma^{-1}\right]_{X\mu,Xi}$$

The inverse metric can be easily removed by right-multiplication with σ, leaving something resembling an eigenvalue problem. Remember,

however, that we require that the XX-block of the total Fockian is diagonal, as per in the Lagrangian. That is to say, defining $[\mathbf{F}]_{Xi,Xj} = \epsilon_{Xi}\delta_{ij}$, we rewrite the variational optimisation as F fragmented Roothaan-Hall equations:

$$\mathbf{F}^X\mathbf{C}^X_X = \mathbf{S}_{XX}\mathbf{C}^X_X\epsilon_X \tag{2.51}$$

By construction, this solution then satisfies both equations 2.43 and 2.48. The solution method is not unique however, and we have built $\hat{f}_X$ specifically so as to be Hermitian here. Other choices are possible, which while they will converge to the same solution in principle, may do so at different rates. Previous studies have suggested our choice to be the most robust [1, 2, 42].

2.3 PROBLEMS

By construction, there can be no BSSE present in the solution to the ALMO problem. Moreover, given that the diagonalisation of an $M \times M$ matrix is an $\mathcal{O}(M^3)$ process, the cost for diagonalising F fragment Fock matrices is $\mathcal{O}(Fn^3)$, as opposed to $\mathcal{O}(N^3) = \mathcal{O}(F^3n^3)$ for the full matrix. Thus, we expect savings in this step on the order of F^2, but more importantly, this means that the overall step is linear scaling in the system size. This comes at the expense of a more complicated Fock build, and the need to compute the inverse metric. However, as will be shown later, it is possible to make the former also linear scaling, and the latter can be done very rapidly using Cholesky decomposition [124]. In addition, information about all fragments is only needed when forming the density and calculating the energy from it, such that the rest of the procedure is readily parallellisable.

However, as was noted earlier, by restricting the orbitals in the above way, mixing between fragments is disallowed. This removes the basis set superposition error, but also any semblance of charge transfer. This can be demonstrated explicitly by considering a Mulliken population analysis. The charge on fragment X, q_X, according to a Mulliken analysis is given by [216]:

$$q_X = \text{Tr}_X\{\mathbf{PS}\} \tag{2.52}$$

where $\mathbf{P}$ is either the closed-shell density, or the total density (i.e. the sum of the α- and β-densities), and the trace is over only the X block

of the term in brackets. Expanding this out using equations 2.45 and 2.42, this becomes

$$q_X = \sum_Y T^{X\mu\cdot}_{\cdot Xi} \sigma^{Xi,Yj} T^{\cdot Y\nu}_{Yj\cdot} S_{Y\nu,X\mu}$$
$$= \sum_Y \sigma^{Xi,Yj} \left(T^{\cdot Y\nu}_{Yj\cdot} \langle \nu^Y | \right) \left(| \mu^X \rangle T^{X\mu\cdot}_{\cdot Xi} \right)$$
$$= \sum_Y \sigma^{Xi,Yj} \langle \phi_{Yj} | \phi_{Xi} \rangle = \mathrm{Tr}_X \{ \mathbf{I} \}$$

That is, the charge on X is exactly equal to the number of electrons on X, such that no charge has been transferred.

This is only one choice of population analysis, however, so that while it does suggest something fundamentally wrong with the method on a qualitative level, it should not be taken to be quantitatively exact. A different way to define the partial charges is a Löwdin analysis [217]. In this approach, the density is projected by the symmetric orthogonaliser $\mathbf{S}^{-1/2}$. This still gives the correct total electron density, by the invariance of traces to permutations:

$$O = \mathrm{Tr} \{ \mathbf{PS} \} = \mathrm{Tr} \left\{ \mathbf{S}^{1/2} \mathbf{PS}^{1/2} \right\} = \mathrm{Tr} \{ \mathbf{X} \}$$

where we have defined $\mathbf{X} = \mathbf{MPM}$, with $\mathbf{M} = \mathbf{S}^{1/2}$. If we consider a system of just two fragments, X and Y, we can expand this block by block as

$$O = \mathrm{Tr} \{ M_{XX} P_{XX} M_{XX} + M_{XX} P_{XY} M_{YX} + M_{XY} P_{YX} M_{XX} + M_{XY} P_{YY} M_{YX} \}$$
$$+ \mathrm{Tr} \{ M_{YX} P_{XX} M_{XY} + M_{YX} P_{XY} M_{YY} + M_{YY} P_{YX} M_{XY} + M_{YY} P_{YY} M_{YY} \}$$
$$= q_X^L + q_Y^L$$

Given that $\mathbf{S} = \mathbf{M}^2$, $S_{XX} = M_{XX} M_{XX} + M_{XY} M_{YX}$, which implies that, after some permuting within the trace:

$$q_X^L = \mathrm{Tr} \{ P_{XX} S_{XX} \}$$
$$+ \mathrm{Tr} \{ M_{YX} M_{XX} P_{XY} + M_{XX} M_{XY} P_{YX} + M_{YX} M_{XY} P_{YY} - M_{XY} M_{YX} P_{XX} \}$$

However, we have that

$$\mathrm{Tr} \{ P_{XX} S_{XX} \} = \mathrm{Tr} \left\{ S_{XX} T^X_X \sigma^{-1}_{XX} T^X_X \right\} = \mathrm{Tr} \left\{ \sigma_{XX} \sigma^{-1}_{XX} \right\} = o_X - \mathrm{Tr} \left\{ \sigma_{XY} \sigma^{-1}_{YX} \right\}$$
$$= o_X - \mathrm{Tr} \{ S_{XY} P_{YX} \} = o_X - \mathrm{Tr} \{ M_{XX} M_{XY} P_{YX} + M_{XY} M_{YY} P_{YX} \}$$

Using this, the Löwdin population on X is given by

$$q_X^L = o_X + \mathrm{Tr} \{ M_{YX} M_{XX} P_{XY} + M_{YX} M_{XY} P_{YY} \}$$
$$- \mathrm{Tr} \{ M_{XY} M_{YY} P_{YX} + M_{XY} M_{YX} P_{XX} \} \qquad (2.53)$$

The remaining traces cannot be simplified, and do not necessarily cancel. That is to say in general $q_X^L \neq o_X$, meaning that some fraction of

charge will have been transferred according to the Löwdin analysis. The significance of this will become apparent later, when constructing domains for the Fock build. Essentially, by using the Löwdin populations, it is possible to estimate the importance of exchange between two fragments, whereas the Mulliken populations would give no useful information; indeed, as shown above, they would result in no exchange between fragments at all.

2.4 IMPLEMENTATION

There are a number of differences between the traditional HF SCF procedure and that with ALMO constraints. Firstly, the initial guess at the molecular orbitals, from which the initial density is constructed, needs to satisfy the block-diagonal constraints in equation 2.41. If a fragment-basis full SCF calculation has been performed on each fragment beforehand, this can easily be achieved by using the converged coefficient tensors from these calculations and assigning them to be the diagonal blocks for the ALMO procedure. This has two advantages: monomer calculations will generally need to be performed anyway, as the desired output is a BSSE-free interaction energy; and the guess so generated is of a much higher quality than would typically be obtained from, for example, a superposition of atomic densities, as it is the result of a self-consistent calculation.

2.4.1 The SCF routine

This assumes that there are no linear dependencies in the basis.

The other major differences lie in the need to invert the overlap metric in each iteration when forming the density via equation 2.45, and that the diagonalisation of the full Fock matrix is replaced by the construction and diagonalisation of F locally projected Fock matrices. As the overlap metric is necessarily positive definite, the former can be rapidly achieved using Cholesky decomposition. This is formally $\mathcal{O}(O^3)$ [124], but the number of occupied orbitals is usually only a small fraction of N, and the prefactor for the decomposition is very small, meaning this is not expected to become a bottleneck. The fragment-localised diagonalisations are inherently linear with respect to system size, but each Fock matrix needs to be constructed first. From equation 2.50, we need the complementary projector $\hat{Q} = \hat{I} - \hat{P}$, with AO representation $\mathbf{I} - \mathbf{SP}$, which immediately gives the first term,

$[\mathbf{QFQ}]_{XX}$. This projector can then be used to find the other contributions:

$$\langle\phi_{X\mu}|\hat{P}_X^\dagger\hat{F}\hat{Q}|\phi_{X\nu}\rangle = \langle\phi_{X\mu}|\phi_{Xi}\rangle\langle\phi^{Xi}|\hat{F}\hat{Q}|\phi_{X\nu}\rangle = \mathbf{S}_{X\mu,X\lambda}\,[\mathbf{PFQ}]_{X\lambda,X\nu}$$

$$\langle\phi_{X\mu}|\hat{P}_X^\dagger\hat{F}\hat{P}|\phi_{X\nu}\rangle = \langle\phi_{X\mu}|\phi_{Xi}\rangle\langle\phi^{Xi}|\hat{F}|\phi^{Xj}\rangle\langle\phi_{Xj}|\phi_{X\nu}\rangle$$
$$= \mathbf{S}_{X\mu,X\lambda}\,[\mathbf{PFP}]_{X\lambda,X\tau}\,\mathbf{S}_{X\tau,X\nu}$$

Combined, these give the fragment Fock matrices as

$$\mathbf{F}_X = [\mathbf{QFQ}]_{XX} + [\mathbf{QFP}]_{XX}\,\mathbf{S}_{XX} + \mathbf{S}_{XX}\,[\mathbf{PFQ}]_{XX} + \mathbf{S}_{XX}\,[\mathbf{PFP}]_{XX}\,\mathbf{S}_{XX} \quad (2.54)$$

This therefore only requires the matrix-matrix multiplications $\mathbf{QFQ}$, $\mathbf{QFP}$, and $\mathbf{PFP}$ in the full AO basis. These can then be scattered out to each fragment.

Algorithm 2.1 The ALMO SCF routine

1: $\epsilon \leftarrow$ convergence threshold
2: **for** each fragment X **do**
3:      Perform HF SCF calculation, store energy and MO coefficients
4: **end for**
5: **while** not converged **do**
6:      **for** each fragment pair XY **do**
7:          Form σ_{YX} using equation 2.42
8:      **end for**
9:      Solve $\sigma\sigma^{-1} = \mathbf{I}$ by Cholesky decomposition
10:      **for** each fragment pair XY **do**
11:          Form $\mathbf{P}_{XY}$ using equation 2.45
12:      **end for**
13:      **if** first iteration **then**
14:          Compute ERIs and initial $\mathbf{P}$
15:      **end if**
16:      Build Fock matrix, $\mathbf{F}$
17:      Calculate energy, E, using equation 2.44
18:      **if** $\Delta E, \|\Delta \mathbf{P}\| < \epsilon$ **then**
19:          Converged
20:      **else**
21:          Form $\mathbf{QFQ}$, $\mathbf{QFP}$, and $\mathbf{PFP}$
22:          **for** each fragment X **do**
23:              Form $\mathbf{F}_X$ using equation 2.54
24:              Diagonalise $\mathbf{F}_X$ to get new MO coefficients
25:          **end for**
26:      **end if**
27: **end while**

The overall ALMO SCF algorithm is summarised in Algorithm 2.1. The most expensive step remaining is line 16, the formation of the
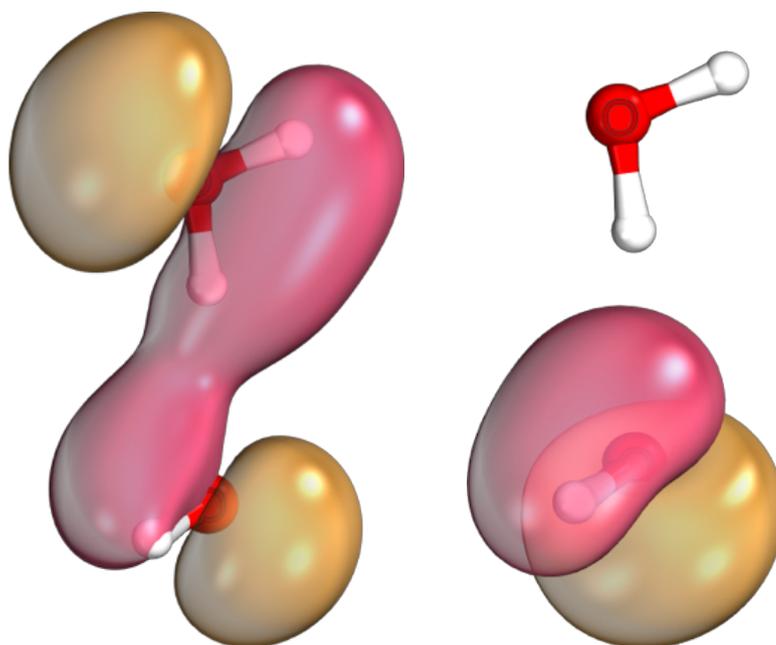
Figure 2.1: Equivalent energy-ordered (fourth occupied) molecular orbitals for the water dimer at a fixed geometry, from canonical HF SCF (left) versus ALMO SCF as implemented here. This shows clearly how the former are highly delocalised, while the latter is absolutely localised, as expected. These orbital depictions were created using IBOView [218].

Fock matrix, which will be discussed later. Given the insignificant cost of the Cholesky decomposition noted earlier, the next time-consuming step is the matrix multiplication in line 21. As the standard procedures for such multiplications are highly optimised, however, this is not expected to become a bottleneck. Figure 2.1 compares a molecular orbital given by the algorithm presented, as compared to the equivalent canonical orbital; this succinctly demonstrates both the efficacy of the implementation, and the main point of using ALMOs.

2.4.2 Convergence acceleration

The SCF equations are simply nonlinear, fixed-point iterations of a functional (the Fock operator) of the density. As such, the problem can be viewed as a constrained Newton-Raphson minimisation [34], for which convergence is only guaranteed if the functional is well-behaved and the starting guess is in a convex region around a minimum. This is not often the case, however, such that a rudimentary algorithm will not converge in most instances [219, 220]. The constraints inherent to the ALMO approach can make this problem worse, as noted in the original formulation by Stoll [1], leading him to suggest that the variational approach be eschewed in favour of a direct,

A region $\mathcal{R}$ is convex if for all $x, y \in \mathcal{R}$ and $t \in [0, 1]$, $(1 - t)x + ty \in \mathcal{R}$.

quadratic optimisation. Since then, however, methods for accelerating convergence of the Roothaan-Hall equations have been developed. The most popular approach in conventional calculations is the direct inversion of the iterative subspace (DIIS) method of Pulay [221].

Consider a set of trial orbital densities, $\{\mathbf{p}^i\}$, generated after each iteration, i. Upon convergence, a final set of densities, $\{\mathbf{p}^f\}$ will have been found. Defining the residual at each step as

$$\Delta\mathbf{p}^i = \mathbf{p}^{i+1} - \mathbf{p}^i$$

It should then be possible to approximate the true solution as a linear combination of the previous densities

$$\mathbf{p}^{m+1} = \sum_{i=0}^{m} c_i \mathbf{p}^i$$

where the c_i are real coefficients and m is the number of steps. This implies that

$$\mathbf{p}^{m+1} = \sum_{i=0}^{m} c_i \left(\mathbf{p}^{i-1} + \Delta\mathbf{p}^{i-1} \right) = \sum_{i=0}^{m} c_i \mathbf{p}^{i-1} + \Delta\mathbf{p}$$

Then, $\mathbf{p}^{m+1} - \mathbf{p}^m$ vanishes if the total residual

$$\Delta\mathbf{p} = \sum_{i=0}^{m} c_i \Delta\mathbf{p}^i = \mathbf{0}$$

Similarly, writing each vector as $\mathbf{p}^i = \mathbf{p}^f + \mathbf{e}^i$, where $\mathbf{e}^i$ is the error from the true solution, we see that

$$\mathbf{p}^{m+1} = \mathbf{p}^f \sum_{i=0}^{m} c_i + \sum_{i=0}^{m} c_i \mathbf{e}^i$$

Thus the true solution is found by minimising the second term above, subject to the constraint that $\sum_i c_i = 1$ so that the first term equates to $\mathbf{p}^f$.

The DIIS method approximates this by equating the minimisation of the second term with the solution of $\Delta\mathbf{p} = \mathbf{0}$. This can be achieved by minimising the residual norm, equivalent to finding the stationary points of the Lagrangian

$$\mathcal{L} = \mathbf{c}^\dagger \mathbf{B}\mathbf{c} - \lambda \left(1 - \sum_i c_i \right)$$

where λ is a Lagrange multiplier and $B_{jk} = \langle \Delta\mathbf{p}^j | \Delta\mathbf{p}^k \rangle$. Differentiation leads to the set of m linear equations

$$\frac{\partial\mathcal{L}}{\partial c_j} = 2B_{ji}c_i - \lambda = 0 \tag{2.55}$$

which along with the constraint completely determine the system. These can be solved by any number of standard algorithms, but it should be noted that the problem is not positive definite, and the linear system may become rank degenerate due to numerical roundoff errors when the residuals become very small. Therefore in the current work an LU decomposition with a rank-preserving purification procedure is employed [124]. In practice, only the last few residuals are stored and used to determine the update vector. Either the density itself can be extrapolated or the Fock matrix, with the latter often being preferred. One important subtlety is that only the raw errors, never the extrapolated errors, should be stored and reused at each step.

Algorithm 2.2 DIIS-accelerated ALMO procedure

> **for** each fragment X **do**
> > Compute $\mathbf{F}_X$ and $\mathbf{e}_X$ using equations 2.54 and 2.56
> **end for**
> Solve the linear system, equation 2.55, for $\{c_i\}$, using $\mathbf{e} = \cup \mathbf{e}_X$
> **for** each fragment X **do**
> > Extrapolate $\mathbf{F}_X^{m+1} \leftarrow \sum_i^m c_i \mathbf{F}_X$
> > Diagonalise $\mathbf{F}_X$ to get new MO coefficients
> **end for**

The choice of residual vector to use is not uniquely defined, as the true error cannot be known *a priori*. One approach is to choose the energy gradient with respect to rotations among the orbitals [42], such that convergence in the error implies convergence to the energy minimum. The orbitals are rotated by applying a unitary operator $\hat{U}$, which is usually parametrised as

$$\hat{U} = \exp(-\hat{\Delta})$$

where the $\hat{\Delta}$ is some anti-Hermitian (necessary to ensure the unitarity of $\hat{U}$) operator. Note that it must be constrained to only rotate orbitals within each fragment so as to maintain localisation. The same approach can then be applied as was used to find the gradient with respect to the coefficients, equation 2.19. This is somewhat involved, and the full derivation is given in Appendix B. The resulting gradient is given by

$$\frac{\partial E}{\partial (\Delta_\alpha)^{X\zeta,X\tau}} = S_{X\tau,X\eta} \left[\mathbf{PF}_\alpha\mathbf{Q}\right]^{X\eta\cdot}_{\cdot X\zeta} - \left[\mathbf{QF}_\alpha\mathbf{P}\right]^{\cdot X\pi}_{X\tau\cdot} S_{X\pi,X\zeta} \tag{2.56}$$

such that the total residual vector is the concatenation of F local fragment residuals. We note that the residual terms are identical to the second and third terms of equation 2.54, albeit with different signs, so that the error vector can be calculated at the same time as the fragment Fock matrices are built, at no additional cost. Incorporation of
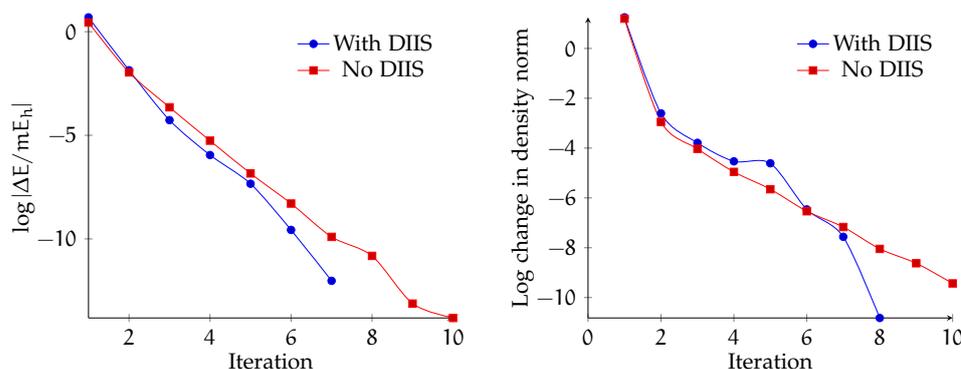
Figure 2.2: Accelerated convergence in the energy and density using Algo-
rithm 2.2 with DIIS, compared to the original Algorithm 2.1 with-
out DIIS. The data is from a calculation on the water dimer using
the cc-pVDZ basis [128]; with the aug-cc-pVDZ basis [128], the
algorithm without DIIS does not converge.

the DIIS procedure in Algorithm 2.1 thus only requires that lines 20
to 23 be replaced with Algorithm 2.2.

As usually at most eight error vectors are stored, solving the linear
system incurs negligible computational effort, and the calculation and
extrapolation steps remain linear in the number of fragments. It has
been shown repeatedly that DIIS substantially reduces the number
of iterations required to achieve convergence [42, 219, 221], and is
considerably more robust than the simple fixed-point procedure. This
is demonstrated in Figure 2.2, where it can be seen that the ALMO
SCF calculation on the water dimer in a small basis converges both
more rapidly and with greater stability. In fact, increasing the basis to
the augmented version (equivalent to adding diffuse basis functions),
the DIIS-based algorithm converges in eight steps, while the basic
algorithm without DIIS does not converge at all.

3 | ENERGY DECOMPOSITION

In this chapter, I describe the new method based on ALMOs in the context of the different terms in the interaction energy. The electrostatic and polarisation terms are based on the work of Head-Gordon and coworkers [2], after which I develop corrections to their scheme so that it includes charge transfer, and in particular, accurate dispersion and exchange-dispersion. The latter terms will be treated with the random phase approximation, the choice of which will be justified, and I will show how it can easily be systematically extended to include more accurate correlation energies. A main focus here is setting up the terms in such a way that they can be implemented to be linear-scaling with the number of molecules, the details of which will be described in the next chapter.

The absolutely localised molecular orbitals of the previous chapter form the base for the new method, from which we can decompose intermolecular interaction energies in terms of physically distinct quantities. This partitioning of the energy, and the motivation behind it, can broadly be broken into four main components:

1. 'frozen' energy, describing the electrostatic interaction between the unrelaxed, frozen monomer orbitals in their distorted geometries;

2. polarisation energy, resulting from the relaxation of the orbitals within the absolutely localised constraint;

3. charge transfer, resulting from delocalisation between fragments;

4. dispersion, due to the correlated excitations between electrons on separate fragments.

The last term can further be split into pure dispersion and exchange dispersion terms, reflecting whether the excitations are on-fragment or between two fragments. If the undistorted monomer energy is available, for example in calculations where the supermolecular geometry is being optimised, then a further, 'relaxation' energy term can be defined as the difference in energy between the distorted and undistorted monomers. This therefore represents the steric penalty associated with bringing the fragments into proximity with one another.

3.1 ELECTROSTATICS AND POLARISATION

After full monomer SCF calculations have been performed on the monomers in the geometries they assume in the complex, each fragment has a set of unperturbed occupied orbitals described by coefficients, $\mathbf{T}^\gamma_{X,\text{frz.}}$. By forming the block-diagonal concatenation of all such coefficients, we can define a supermolecular frozen density matrix for each electron spin, $\mathbf{P}^\gamma_{\text{frz.}}$, as

$$\left(P^\gamma_{\text{frz.}}\right)^{X\mu, Y\nu} = \left(T^\gamma_{\text{frz.}}\right)^{X\mu\cdot}_{\cdot Xi} \left(\sigma^\gamma\right)^{Xi, Yj} \left(T^\gamma_{\text{frz.}}\right)^{\cdot Y\nu}_{Yj\cdot} \tag{3.1}$$

which is analogous to equation 2.45 in the ALMO procedure. The inclusion of the inverse of the metric, formed exactly as in equation 2.42 but with the frozen coefficients, means that this density is not simply the sum of non-interacting fragment densities (assuming that the metric is not simply the identity). That is, in general the fragment-localised orbitals will not be mutually orthogonal. Thus the frozen density represents the interactions between the unperturbed fragment electron densities, represented by the occupied orbitals. This corresponds to multipole-multipole interactions [10], i.e. electrostatics, which will be most favourable when the multipole moments of each fragment align. The presence of the metric, however, ensures that as fragments are brought closer together, these interactions include Pauli repulsion. In the long-range regime, the former will dominate, whereas the latter will become dominant at small separations.

The frozen energy is thus defined to be the difference between the energy as determined using the frozen density, and the sum of the fragment energies in their distorted geometries but at infinite separation, E_X. That is, from equation 2.44:

$$\Delta E^\gamma_{\text{frz.}} = \frac{1}{2}\text{Tr}\left\{\left[\mathbf{H} + \mathbf{F}(\mathbf{P}^\gamma_{\text{frz.}})\right]\mathbf{P}^\gamma_{\text{frz.}}\right\} - \sum_X E_X \tag{3.2}$$

While the term 'frozen' energy makes sense in the context of the original development, it is more convenient to refer to this as the electrostatic term with exchange included, as these are more standard descriptions. The ALMO-EDA is unique in the sense that the exchange is *a priori* included, whereas other methods like SAPT attempt to separate it out; this is due to antisymmetry being enforced on the wavefunction throughout, rather than being included in the perturbative expansion.

During the ALMO SCF procedure, the frozen orbitals above are allowed to relax. This therefore corresponds to the distortion of electron density on each fragment due to the presence of the other fragments, i.e. the polarisation of charge. If the ALMO densities are $\mathbf{P}^\gamma_0$, then

defining $\Delta \mathbf{P}^\gamma_{\text{pol.}} = \mathbf{P}^\gamma_0 - \mathbf{P}^\gamma_{\text{frz.}}$, the polarisation contribution to the interaction energy is given by

$$\Delta E^\gamma_{\text{pol.}} = \frac{1}{2} \text{Tr} \left\{ [\mathbf{H} + \mathbf{F}(\mathbf{P}^\gamma)] \, \Delta \mathbf{P}^\gamma_{\text{pol.}} \right\} \tag{3.3}$$

The polarisation here is more rightly termed *local* polarisation, as it is constrained to the individual molecules. The lack of extended polarisation - or charge transfer - is how it differs from a full Hartree-Fock solution.

3.2   CHARGE TRANSFER

The ALMO SCF procedure leads to an energy for the complex that is free of superposition error, but is necessarily higher than the true solution without the localisation constraint. The only consequence of the restriction, as demonstrated above, is the inability for fragments to share electron density - rather, they are optimised in the effective field of all the other fragments. The difference between the full solution and the ALMO solution can thus be chosen as the definition of charge transfer. This could be calculated exactly by performing the full complex calculation and observing the energy difference, but that would negate any computational savings. As such, it is desirable to be able to estimate what the energy would have been if each fragment had access to the entire supermolecular basis. In doing so we necessarily reintroduce some BSSE, but this is the only term in which this is true.

3.2.1   Perturbative correction

This expansion from a smaller to a larger basis can be achieved through a perturbative correction, directly analogous to that employed in the dual basis methods described earlier [42]. Following the usual Rayleigh-Schrödinger perturbation theory [222], we take the zeroth-order Hamiltonian to be

$$2\hat{H}_0 = \hat{h} + \hat{\rho}_0 \hat{f}(\hat{\rho}_0) \hat{\rho}_0 + \hat{q}_0 \hat{f}(\hat{\rho}_0) \hat{q}_0 \tag{3.4}$$

where $\hat{\rho}_0$ is the ALMO density operator, the discrete representation of which is given in equation 2.45, $\hat{q}_0 = \hat{I} - \hat{\rho}_0$, and $\hat{f}$ is the associated Fock operator. This Hamiltonian contains the projections of the Fock operator onto the occupied space and its orthogonal complement, with no mixing terms. Taking the full Hamiltonian to be the converged ALMO Fock operator, as we have no way of knowing what the full density would be, the perturbation is thus

$$\hat{V} = \frac{1}{2} \left\{ \hat{h} + \hat{f}(\hat{\rho}_0) \right\} - \hat{H}_0 = \hat{\rho}_0 \hat{f}(\hat{\rho}_0) \hat{q}_0 \tag{3.5}$$

It is possible to proceed from this point without enforcing strong orthogonality between the occupied and virtual subspaces. Regardless of the approach, mixing the fragment subspaces will necessarily reintroduce some of the BSSE that we are trying to avoid. However, as the zeroth-order wavefunction is untainted by BSSE, and the perturbation is necessarily small, the error is expected to be very small. Moreover, this term is available as a well-separated contribution, such that the error can be clearly identified. As a result, there is no disadvantage in simplifying matters by first enforcing strong orthogonality, then allowing the occupied and virtual orbitals to mix within their respective subspaces, i.e. by finding orbital rotations such that the occupied-occupied and virtual-virtual blocks of the Fock matrix are (separately) diagonalised. The former is achieved by projecting the virtual coefficients in the AO basis out of the occupied subspace, and will be considered later. For the diagonalisation, we apply a unitary transformation $\hat{u}$ to the orbitals:

$$\langle \phi_i | \hat{U}^\dagger \hat{f} \hat{U} | \phi_j \rangle = \overline{\epsilon}_i \delta_{ij} \quad \text{and} \quad \langle \phi_a | \hat{U}^\dagger \hat{f} \hat{U} | \phi_b \rangle = \overline{\epsilon}_a \delta_{ab}$$

where we use a bar to denote a transformed quantity. In this way, the resolvent operator is given by

$$\hat{R}_0 = \frac{\hat{U}^\dagger \hat{q}_0 \hat{U}}{E_0 - \hat{U}^\dagger \hat{H}_0 \hat{U}}$$

where E_0 is the energy before transformation, i.e. the energy of the zeroth-order, ALMO SCF wavefunction Φ_0. The infinite-order wavefunction Φ and energy correction ΔE are thus as follows:

$$\begin{aligned}
\Phi &= \sum_{m=0}^{\infty} \left[\hat{R}_0 (\hat{V} - \Delta E) \right]^m \Phi_0 \\
\Delta E &= \sum_{m=0}^{\infty} \langle \Phi_0 | \hat{V} \left[\hat{R}_0 (\hat{V} - \Delta E) \right]^m | \Phi_0 \rangle
\end{aligned} \tag{3.6}$$

Writing $\hat{\rho} = \hat{U}^\dagger \hat{\rho}_0 \hat{U}$, and similarly for $\hat{q}$ with $\hat{q}_0$, the transformation implies that $\hat{\rho}$ projects orbitals onto the occupied space, and $\hat{q}$ projects them onto the virtual space. As such, the matrix elements of the perturbation are

$$\begin{aligned}
\overline{V}_{ij} &= \langle \overline{\phi}_i | \hat{\rho} \hat{f} \hat{q} | \overline{\phi}_j \rangle = \langle \overline{\phi}_i | \hat{f} 0 | \overline{\phi}_j \rangle = 0 \\
\overline{V}_{ab} &= \langle \overline{\phi}_a | \hat{\rho} \hat{f} \hat{q} | \overline{\phi}_b \rangle = \langle \overline{\phi}_a | 0 \hat{f} | \overline{\phi}_b \rangle = 0 \\
\overline{V}_{ia} &= \langle \overline{\phi}_i | \hat{\rho} \hat{f} \hat{q} | \overline{\phi}_a \rangle = \langle \overline{\phi}_i | \hat{f} | \overline{\phi}_a \rangle = \overline{f}_{ia}
\end{aligned} \tag{3.7}$$

Using equation 3.6, we thus get that the first- and second-order energy corrections are

$$\Delta E^{(1)} = \langle \Phi_0 | \hat{V} | \Phi_0 \rangle = \overline{V}_{ii} = 0$$

$$\Delta E^{(2)} = \langle \Phi_0 | \hat{V} \hat{R}_0 \hat{V} | \Phi_0 \rangle = \frac{\langle \overline{\phi}_i | \hat{V} | \overline{\phi}_a \rangle \langle \overline{\phi}_a | \hat{V} | \overline{\phi}_i \rangle}{\overline{\epsilon}_i - \overline{\epsilon}_a} = \frac{|\overline{f}_{ia}|^2}{\overline{\epsilon}_i - \overline{\epsilon}_a} \tag{3.8}$$

It is possible, but tedious, to continue in this way to arbitrary order, in particular finding that $\Delta E^{(3)} = 0$ and

$$\Delta E^{(4)} = \sum_{i,j \neq i} \sum_{a,b} \frac{\bar{f}_{ia}\bar{f}_{aj}\bar{f}_{jb}\bar{f}_{bi}}{(\bar{\epsilon}_i - \bar{\epsilon}_a)(\bar{\epsilon}_i - \bar{\epsilon}_j)(\bar{\epsilon}_i - \bar{\epsilon}_b)}$$
$$- \sum_i \sum_{ab} \frac{\left|\bar{f}_{ia}\right|^2 \left|\bar{f}_{ib}\right|^2}{(\bar{\epsilon}_i - \bar{\epsilon}_a)(\bar{\epsilon}_i - \bar{\epsilon}_b)^2} \tag{3.9}$$

where we have explicitly shown the summations for clarity.

The advantages of the above approach are that it is very simple to implement to second- and fourth-order, and these usually give an excellent estimate to the full correction, unless the system is very strongly bound. Alternatively, a full diagonalisation of the untransformed Fock matrix can be performed - equivalent to taking a single, unconstrained Roothaan step - to determine the density $\hat{\rho}_\infty$, and the infinite-order correction calculated as

$$\Delta E = \text{Tr}\left\{ \hat{f}(\hat{\rho}_0)\left[\hat{\rho}_\infty - \hat{\rho}_0\right] \right\} \tag{3.10}$$

It should be noted at this point that in the spin-unrestricted case, all of the above should be applied separately to each of $\hat{f}^\alpha$ and $\hat{f}^\beta$ and the results summed, while in the restricted case, each energy correction gains a factor of two due to double occupation. The primary disadvantage of this approach is that it requires an essentially $\mathcal{O}(N^3)$ diagonalisation step, even if only the second-order energy is required (as $V \sim N$ asymptotically).

The need for diagonalisation can be circumvented by considering directly the orbital rotation, $\hat{U}$. The energy lowering associated with such a transformation is, using the invariance of the core Hamiltonian and the unitarity of the transformation:

$$\Delta E = \text{Tr}\left\{ \hat{U}\hat{\rho}_0\hat{U}^\dagger \hat{f}\hat{U}\hat{\rho}_0\hat{U}^\dagger - \hat{\rho}_0\hat{f}\hat{\rho}_0 \right\} = \text{Tr}\left\{ \hat{\rho}_0\left[\hat{\bar{f}} - \hat{f}\right]\hat{\rho}_0 \right\} \tag{3.11}$$

where the transformed Fock operator is $\hat{\bar{f}} = \hat{U}^\dagger \hat{f}\hat{U}$. Applying $\hat{U}$ to the left, this becomes $\hat{U}\hat{\bar{f}} = \hat{f}\hat{U}$, from which we take just the occupied-occupied and virtual-occupied blocks. Switching to the discrete representation with occupied and virtual blocks denoted by subscripts O and V, this gives:

$$\mathbf{U}_{OO}\bar{\mathbf{f}}_{OO} + \mathbf{U}_{OV}\bar{\mathbf{f}}_{VO} = \mathbf{f}_{OO}\mathbf{U}_{OO} + \mathbf{f}_{OV}\mathbf{U}_{VO} \tag{3.12}$$
$$\mathbf{U}_{VO}\bar{\mathbf{f}}_{OO} + \mathbf{U}_{VV}\bar{\mathbf{f}}_{VO} = \mathbf{f}_{VO}\mathbf{U}_{OO} + \mathbf{f}_{VV}\mathbf{U}_{VO} \tag{3.13}$$

The energy functional for the transformation is minimised by requiring that $\hat{q}_0\hat{\bar{f}}\hat{\rho}_0$ disappears [223], such that after applying $\mathbf{U}_{OO}^{-1}$ from the right, we get

$$\bar{\mathbf{f}}_{OO} = \mathbf{U}_{OO}\bar{\mathbf{f}}_{OO}\mathbf{U}_{OO}^{-1} - \mathbf{f}_{OV}\mathbf{U}_{VO}\mathbf{U}_{OO}^{-1}$$

Inserting this into equation 3.11 then yields

$$\Delta E = \mathrm{Tr}\left\{\bar{\mathbf{f}}_{OO} - \mathbf{U}_{OO}\bar{\mathbf{f}}_{OO}\mathbf{U}_{OO}^{-1} + \mathbf{f}_{OV}\mathbf{U}_{VO}\mathbf{U}_{OO}^{-1}\right\} = \mathrm{Tr}\{\mathbf{f}_{OV}\mathbf{x}_{VO}\} \quad (3.14)$$

where we have used the permutational invariance of the trace and defined $\mathbf{x}_{VO} = \mathbf{U}_{VO}\mathbf{U}_{OO}^{-1}$.

Applying $\mathbf{U}_{OO}^{-1}$ to the right of equations 3.13 and 3.12, and $\mathbf{x}_{VO}$ to the left in the latter, then results in

$$\mathbf{U}_{VO}\bar{\mathbf{f}}_{OO}\mathbf{U}_{OO}^{-1} = \mathbf{f}_{VO} + \mathbf{f}_{VV}\mathbf{x}_{VO}$$
$$\mathbf{U}_{VO}\bar{\mathbf{f}}_{OO}\mathbf{U}_{OO}^{-1} = \mathbf{x}_{VO}\mathbf{f}_{OO} + \mathbf{x}_{VO}\mathbf{f}_{OV}\mathbf{x}_{VO}$$

Subtracting one from the other then leads to the following quadratic equation:

$$\mathbf{f}_{VO} + \mathbf{f}_{VV}\mathbf{x}_{VO} - \mathbf{x}_{VO}\mathbf{f}_{OO} - \mathbf{x}_{VO}\mathbf{f}_{OV}\mathbf{x}_{VO} = 0 \quad (3.15)$$

This can be solved iteratively for $\mathbf{x}_{VO}$, from which the energy can be determined by equation 3.14.

The main advantages of this approach are that the formulation is invariant with respect to change of basis, and that the most expensive step computationally is the matrix-matrix multiplications in the quadratic term above; such multiplications scale as $\mathcal{O}(N^2)$, as opposed to $\mathcal{O}(N^3)$, and in general involve far fewer floating point operations. The disadvantage is, of course, that several iterations may be necessary to converge the coefficients. As a result, if only the second-order correction is needed, it is often cheaper to use the formulation in equation 3.8, up to a certain system size where the asymptotics take effect. As we will see shortly, however, there is a further advantage in the second method, which is that it allows for decomposition of the charge transfer into contributions between fragments. Finally, it should be noted that in the case where the virtual-virtual and occupied-occupied blocks of the Fock matrix are constrained to be diagonal, neglecting the quadratic term in equation 3.15 leads to

$$0 = f_{ai} + f_{ab}x_{bi} - x_{aj}f_{ji} = f_{ai} + (\epsilon_a - \epsilon_i)x_{ai}$$

which rearranges to give $x_{ai} = f_{ai}/(\epsilon_i - \epsilon_a)$. Inserting this in equation 3.14 yields

$$\Delta E = f_{ia}x_{ai} = \frac{|f_{ia}|^2}{\epsilon_i - \epsilon_a} = \Delta E^{(2)}$$

That is, the dual basis correction reduces to the second-order perturbation result. This means that the earlier scheme can be treated computationally as a subset of the latter method, and that inclusion of the quadratic term gives a measure of the difference between the second- and infinite-order corrections.

3.2.2 Directionality

Using equation 3.14, the closed-shell charge transfer contribution is

$$\Delta E_{CT} = 2\text{Tr}\{\mathbf{f}_{OV}\mathbf{x}_{VO}\} \tag{3.16}$$

where the rotation coefficients are either found by iterative solution of equation 3.15, or by following the perturbative procedure outlined in equations 3.8 and 3.9, depending on the truncation level desired.

Charge transfer by definition is a directional quantity, however, and it would be at least qualitatively useful to have a way to analyse in what manner charge is transferred, i.e. between which fragments and to what extent. To do this, we introduce fragment partition operators for each molecular orbital, $\hat{p}_{Xi}$ for occupied orbital ϕ_i on fragment X and $\hat{q}_{Xa}$ for virtual orbital ϕ_a on fragment X. For simplicity, we first project the virtual orbitals into the orthogonal complement of the occupied subspace, i.e. replace ϕ_a with $\overline{\phi}_a$ where

$$|\overline{\phi}_a\rangle = \hat{q}_0 |\phi_a\rangle$$

We then require that

$$\sum_{Xi} \hat{p}_{Xi} = \hat{\rho}_0 \text{ and } \sum_{Xa} \hat{q}_{Xa} = \hat{q}_0$$

with each partition operator itself being a projection; that is, they must be idempotent. Using these, the energy can be partitioned into contributions from fragment Y onto fragment Z as follows:

$$\Delta E_{CT}^{Y \rightarrow Z} = \sum_{i,a} \text{Tr}\left\{\hat{p}_{Yi}\hat{f}\hat{q}_{Za}\hat{X}\right\} \tag{3.17}$$

Note that by summing over all unique Y, Z pairs, including where Y = Z, the total energy from equation 3.16 is recovered.

The choice of partition operators is not unique, in exact analogy with population analyses. This introduces a certain arbitrariness to the procedure, meaning that while it appears to be quantitative, it is only truly useful in a qualitative sense. The simplest choice of partition would be that of Mulliken [216], in which the operators are simply

$$\hat{p}_{Xi} = |\phi_{Xi}\rangle \langle \phi^{Xi}| \text{ and } \hat{q}_{Xa} = |\overline{\phi}_{Xa}\rangle \langle \overline{\phi}^{Xa}| \tag{3.18}$$

which clearly satisfy the required relations; in particular, idempotency follows by the orthonormality of the co- and contra-variant orbitals. The advantage of this choice is that the intrafragment charge transfer terms would be rigorously zero if the virtuals had not been projected, as shown earlier when considering the lack of charge transfer in the ALMO scheme. As in practice the projection only slightly alters the locality, these terms will still be negligibly small. Intuitively, it makes

little sense to talk about transferring electron density to oneself. The resulting energy contributions in the MO basis are then given by

$$\Delta E_{CT,\,M}^{Y\to Z} = \sum_W \langle \phi^{Wk}|\phi_{Yi}\rangle \langle \phi^{Yi}|\hat{f}|\overline{\phi}_{Za}\rangle \langle \overline{\phi}^{Za}|\hat{x}|\phi_{Wk}\rangle$$
$$= \text{Tr}\{(\mathbf{f}_{OV})_{YZ}\,(\mathbf{x}_{VO})_{ZY}\}$$

(3.19)

In Appendix C, it is shown that one can also define partial charges as

$$\Delta Q_{CT,\,M}^{Y\to Z} = \text{Tr}\{(\boldsymbol{\rho}_{OV})_{YZ}\,(\mathbf{x}_{VO})_{ZY}\}$$

(3.20)

This is somewhat more complicated, as the transformed density, $\hat{\rho} = \hat{U}\hat{\rho}_0\hat{U}^\dagger$ needs to be determined. However, the concept of a physical charge moving from one fragment to another in this context - where the molecules are fixed in time and cannot be considered to ionise - is questionable at best. It implies two things: that the electron 'belongs' to some specific subsystem, and as such is labelled; and that it is possible to follow the trajectory of the electron, so as to describe it as moving between subsystems. Clearly, both assertions are physically absurd - electrons are indistinguishable and have no definite position, and the partitioning into subsystems is itself arbitrary. Instead, it is more reasonable to consider charge transfer as the energy lowering caused by the non-local response of the electron density on each fragment when brought into proximity with the other fragments. That is, we have simply split polarisation into local and non-local portions. The first, $\Delta E_{pol.}$, reflects the redistribution of electron density *This volume is effectively* constrained to a finite volume around the fragment, while ΔE_{CT} en-*defined by the extent of* compasses the remaining relaxation outside that volume. *the local atomic orbitals.*

3.3 DISPERSION

The mean-field treatment presented thus far takes no account of dynamical correlation, and therefore is incapable of describing dispersion interactions. In this section, we describe how an exact treatment of dispersion can be formulated in terms of the frequency-dependent electric polarisabilities of the system, and how this leads to the random phase approximation (RPA) approach. As will be demonstrated, this is intimately connected with coupled cluster theory, providing an efficient method of calculating dispersion in such a way that the different interfragment contributions can be isolated. Neither method is new, but the development in this context, and the approach to decomposing the excitations in such a way that they are free of BSSE and able to be computed in a linear-scaling manner, are novel.

3.3.1 Pair correlations

We define one- and two-particle density operators as

$$\hat{\rho}_{pq} = a_p^\dagger a_q \text{ and } \hat{\rho}_{pqrs} = a_p^\dagger a_q^\dagger a_s a_r \qquad (3.21)$$

which represent the number-conserving correlations between different states in the underlying Fock space. The Hamiltonian, from equation 2.10, is thus

$$\hat{H} = h_{pq}\hat{\rho}_{pq} + \frac{1}{2}g_{pqrs}\hat{\rho}_{prqs} \qquad (3.22)$$

highlighting the physical nature of the different couplings. In general the density will fluctuate over time, and the response of the densities to perturbations is by definition the electron correlation.

Consider the wavefunction at time t, $|\Psi(t)\rangle$. In the interaction picture [224], this is propagated from an initial state at time $t = 0$ by the unitary time evolution operator, $\hat{U}(t)$:

$$|\Psi(t)\rangle = \exp(-i\hat{\mathcal{H}}(t))|\Psi(0)\rangle = \hat{U}(t)|\Psi(0)\rangle$$
$$\hat{\mathcal{H}}(t) = \int_{-\infty}^{t} dt'\, \hat{H}(t') \qquad (3.23)$$

Then in general the expectation of any operator, $\hat{A}$, at time t is given by

$$A(t) = \langle\Psi(t)|\hat{A}|\Psi(t)\rangle = \langle\Psi(0)|\hat{U}^\dagger \hat{A}\hat{U}|\Psi(0)\rangle \qquad (3.24)$$

If we consider only the linear response by expanding the exponential to first order, this becomes

$$A(t) = \langle\Psi(0)|[1 + i\hat{\mathcal{H}}(t) + \mathcal{O}(\hat{\mathcal{H}}^2)]\hat{A}[1 - i\hat{\mathcal{H}}(t) + \mathcal{O}(\hat{\mathcal{H}}^2)]|\Psi(0)\rangle$$
$$\approx A(0) + i\langle\Psi(0)|\hat{\mathcal{H}}\hat{A} - \hat{A}\hat{\mathcal{H}}|\Psi(0)\rangle$$

Then the variation in expectation, δA, is simply

$$\delta A = i\int_{-\infty}^{t} dt'\, \langle[\hat{H}(t'), \hat{A}(t-t')]\rangle \qquad (3.25)$$

Next, we write the Hamiltonian as a time-independent, zeroth-order part, and a one-electron perturbation, $\hat{f}(t)$:

$$\hat{H}(t) = \hat{H}_0 - \hat{f}(t) = \hat{H}_0 - f_{pq}(t)\hat{\rho}_{pq}(0)$$

We wish to describe the response of the electron density to this perturbation, such that $\hat{A}(t) = \hat{\rho}_{rs}(t)$ here. As this necessarily commutes with $\hat{H}_0$, equation 3.25 becomes

$$\delta\rho_{rs} = -\int_{-\infty}^{t} dt'\, \mathcal{G}_{pq,rs}^-(t-t')f_{pq}(t') \qquad (3.26)$$

where we have extracted the two-particle retarded Green's function, or time propagator, for the density:

$$G^-_{pq,rs}(\tau) = i \langle [\hat{\rho}_{pq}(0), \hat{\rho}_{rs}(\tau)] \rangle \qquad (3.27)$$

As this describes the linear response in the density, it is also sometimes called the linear response function.

Making the time translation $\tau = t - t'$, equation 3.26 becomes

$$\rho_{rs}(t) - \rho_{rs}(0) = \int_0^\infty d\tau \, G^-_{pq,rs}(\tau) f_{pq}(t - \tau)$$

Fourier-decomposing the perturbation then gives

$$\begin{aligned}
\rho_{rs}(t) - \rho_{rs}(0) &= \int_0^\infty d\tau \, G^-_{pq,rs}(\tau) \int_0^\infty d\omega \, \tilde{f}_{pq}(\omega) e^{i\omega(t-\tau)} \\
&= \int_0^\infty d\omega \, \tilde{f}_{pq}(\omega) e^{i\omega t} \int_0^\infty d\tau \, G^-_{pq,rs}(\tau) e^{-i\omega\tau} \\
&= \int_0^\infty d\omega \, \tilde{f}_{pq}(\omega) G^-_{pq,rs}(\omega) e^{i\omega t}
\end{aligned}$$

where $G^-(\omega)$ is the frequency- (or energy-) dependent Green's function, i.e. the Fourier transformed time-dependent Green's function. The above implies that the Fourier transform of the density is related as

$$\tilde{\rho}_{rs}(\omega) = G^-_{pq,rs}(\omega) \tilde{f}_{pq}(\omega) \qquad (3.28)$$

The Fourier-transformed Green's function, $G^-_{pq,rs}$, is simply the four-point polarisability tensor, or the susceptibility of the electron density to fluctuations in the electric field [225, 226]. Physically, it describes the extent to which the density will polarise, or distort, when exposed to small changes in the interactions, and therefore describes electron correlation. Starting from equation 3.27, integration over a contour an infinitessimal $\eta \to 0_+$ below the real axis yields an explicit form for the polarisability:

Due to its many different names, G^- is often written as α (polarisability) or χ (susceptibility).

$$G^-_{pq,rs}(\omega) = \langle \hat{\rho}_{rs}(0) [\omega + i\eta + \hat{H}]^{-1} \hat{\rho}_{pq}(0) \rangle - \langle \hat{\rho}_{pq}(0) [\omega + i\eta - \hat{H}]^{-1} \hat{\rho}_{rs}(0) \rangle$$

This can be split into real and imaginary parts describing dispersion and absorption, respectively:

$$G^-_{pq,rs}(\omega) = \chi_{pq,rs}(\omega) + i\Pi_{pq,rs}(\omega) \qquad (3.29)$$

Expanding in the complete set of eigenstates of the Hamiltonian, $\{|n\rangle\}$, gives the spectral resolution or Lehmann representation [227] of the polarisability:

$$\chi_{pq,rs}(\omega) = \sum_{n \neq 0} \left[\frac{(\rho_{pq})_{0n}(\rho_{rs})_{n0}}{\omega_n - \omega} + \frac{(\rho_{rs})_{0n}(\rho_{pq})_{n0}}{\omega_n + \omega} \right]$$

$$\Pi_{pq,rs}(\omega) = \pi \sum_{n \neq 0} [(\rho_{pq})_{0n}(\rho_{rs})_{n0}\delta(\omega - \omega_n) - (\rho_{rs})_{0n}(\rho_{pq})_{n0}\delta(\omega + \omega_n)]$$

$$(3.30)$$

where ω_n are the poles of χ, or the resonance frequencies of the density response. In the neighbourhood of these frequencies, the value of χ is taken in the principal value sense.

The delta functions in Π imply that, for any reasonably well-behaved distribution $z(\omega', \omega)$:

$$\frac{1}{\pi} \int_{-\infty}^{\infty} d\omega' \, \Pi_{pq,rs}(\omega') z(\omega', \omega)$$
$$= \sum_{n \neq 0} [(\rho_{pq})_{0n}(\rho_{rs})_{n0} z(\omega_n, \omega) - (\rho_{rs})_{0n}(\rho_{pq})_{n0} z(-\omega_n, \omega)]$$

Choosing $z(\omega', \omega) = (\omega' - \omega)^{-1}$ leads to one half of the Kramers-Kronig (or Hilbert transform) relations [227]:

$$\chi_{pq,rs}(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \, \frac{\Pi_{pq,rs}(\omega')}{\omega' - \omega} \tag{3.31}$$

where P denotes the Cauchy principal value. Thus, knowledge of the imaginary part completely determines the system. In particular, the two-particle transition densities, $d_{pq,rs}$, are given by choosing z to be the Heaviside function [228]:

$$d_{pq,rs} = (\rho_{pq})_{0n}(\rho_{rs})_{n0} = \int_{0}^{\infty} d\omega' \, \frac{1}{\pi} \Pi_{pq,rs}(\omega') \tag{3.32}$$

This implies that $\frac{1}{\pi}\Pi(\omega)$ is the density of states, which describes the availability of states that can absorb an incoming light wave of frequency ω.

3.3.2 Correlation energy and the adiabatic connection

We now demonstrate how the correlation energy can be determined from the polarisability. The correlation energy is defined to be the difference between the true energy of the system, E, and the mean-field approximation E_0 determined from an SCF calculation. Choosing the Fock operator from this calculation as the zeroth-order Hamiltonian, we slowly 'switch on' the perturbation $\hat{V}$, through a parameter λ:

$$\hat{H}(\lambda) = \hat{f} + \lambda \hat{V} \tag{3.33}$$

where the Fock operator is defined by its matrix elements as

$$f_{pq} = h_{pq} + \langle pi \| qi \rangle \tag{3.34}$$

Clearly, this gives $E = E(\lambda = 1)$ and $E_0 = E(\lambda = 0)$, such that the correlation energy is given by

$$\Delta E = \int_{0}^{1} d\lambda \, \frac{\partial E}{\partial \lambda} = \int_{0}^{1} d\lambda \, \langle \Psi(\lambda) | \hat{V} | \Psi(\lambda) \rangle \tag{3.35}$$

This technique is called the 'adiabatic connection' [226, 229, 230], as the interaction is slowly - or adiabatically - tuned from nothing to fully interacting.

Consideration of equation 3.34 gives the interaction operator as

$$\frac{1}{4} \langle pq \| rs \rangle \hat{\rho}_{pq,sr} - \langle pi \| qi \rangle \hat{\rho}_{pq} + \frac{1}{2} \langle ij \| ij \rangle \qquad (3.36)$$

Taking the expectation and inserting in equation 3.35 then yields

$$\Delta E = \int_0^1 d\lambda \left[\frac{1}{4} \langle pq \| rs \rangle \rho_{rs,pq}(\lambda) - \langle pi \| qi \rangle \rho_{pq}(\lambda) \right] \qquad (3.37)$$

Using the anticommutation relations for the creation and annihilation operators, the two-particle density can be written in terms of the one-particle density as

$$\hat{\rho}_{rs,pq} = a_r^\dagger a_s^\dagger a_q a_p = a_r^\dagger \left(a_q a_s^\dagger - \delta_{sq} \right) a_p = \hat{\rho}_{rq} \hat{\rho}_{sp} - \delta_{sq} \hat{\rho}_{rp}$$

This means that the two-particle density matrix is

$$\rho_{rs,pq} = \langle \hat{\rho}_{rq} \hat{\rho}_{sp} \rangle - \delta_{sp} \rho_{rp} = \rho_{rq} \rho_{sp} + d_{pr,qs} - \delta_{ps} \rho_{qr} \qquad (3.38)$$

Next, consider the change in this density as λ slowly increases from zero:

$$\delta \rho_{rs,pq} = (\delta \rho_{rq}) \rho_{sp} + \rho_{rq} (\delta \rho_{sp}) + (\delta \rho_{rq})(\delta \rho_{sp}) + \delta d_{pr,qs} - \delta_{ps} \delta \rho_{qr}$$

The terms involving $\delta \rho$, when averaged over, are presumed to be negligibly small; in the words of Bohm and Pines [231]:

> We distinguish between two kinds of response of the electrons to a wave. One of these is in phase with the wave... which contributes to the organized behaviour of the system. The other response has a phase difference with the wave... this tends to average out to zero when we consider a large number of electrons. This procedure we call the random phase approximation.

Using this in equation 3.37, the correlation energy in the RPA is thus given by

$$\Delta E_{RPA} = \frac{1}{2} \int_0^1 d\lambda \, \mathrm{Tr}\{\mathbf{W}\Delta\mathbf{d}\}$$

$$\Delta d_{pq,rs}(\lambda) = \frac{1}{\pi} \int_0^\infty \left[\Pi^\lambda_{pq,rs}(\omega) - \Pi^0_{pq,rs}(\omega) \right] \qquad (3.39)$$

where we have defined an interaction tensor, $\mathbf{W}$; in the above formalism, this would be $W_{pqrs} = \frac{1}{2} \langle pq \| rs \rangle$. Therefore, the RPA energy is completely determined by knowledge of the polarisabilities. Note that in special cases, such as the uniform gas [232], this energy is exact as the $\delta \rho$ terms are rigorously zero.

3.3.3 Determining the polarisabilities

The zeroth-order polarisability, i.e. that calculated from the SCF wavefunction, is easily determined [229], but it is not possible to write the full polarisability in closed form. Instead, it must be determined in either a self-consistent or perturbative manner. Within the RPA, it is possible to use the former approach, allowing inclusion of terms to infinite order. As a starting point, consider the relation in equation 3.28 as obtained from the Hamiltonian in equation 3.33. For clarity, we write the full and zeroth-order Green's functions as G and G_0, dropping the particle subscripts, so that this is $\tilde{\rho} = G_0 \tilde{f}$. If instead the Hamiltonian was partitioned as

$$\hat{H} = \hat{f}(0) + \lambda W \hat{\rho} - \hat{f}(t) = \hat{f}(0) - \hat{w}(t)$$

where W contains the interaction contribution, then the equivalent to equation 3.28 would be $\tilde{\rho} = G\tilde{w}$. Combined with the earlier relation, this gives $G\tilde{w} = G_0\tilde{f}$. Using the linearity of Fourier transforms,

$$\tilde{w} = \tilde{f} - \lambda W\tilde{\rho} = [1 - \lambda W G_0]\tilde{f}$$

so that eliminating $\tilde{f}$ yields

$$G_0 = G - \lambda G W G_0 \tag{3.40}$$

This is simply the Dyson equation for the propagator [227]. Multiplying on the left by G^{-1} and on the right by G_0^{-1}, then inverting, this becomes

$$G = \left[G_0^{-1} - \lambda W \right]^{-1} = G_0 \left[1 - \lambda W G_0 \right]^{-1} = G_0 + G_0 \sum_{n=1}^{\infty} \lambda^n (W G_0)^n \tag{3.41}$$

Physically, at each order of λ the next order in the interaction is included, screened by the zeroth-order density propagators. It is this inclusion of the screening that makes the RPA such a significant leap forward in solving the many-body problem. Earlier perturbative attempts to determine the correlation energy could not include screening [233, 234], and thus contained spurious divergences; in contrast, the above leads to a geometric series that is absolutely convergent [232]. This can be related to the imaginary part via analytic continuation onto the imaginary frequencies, whereby equation 3.31 results in

$$\int_0^\infty d\omega \, G(i\omega) = \int_0^\infty d\omega \, \Pi(\omega)$$

Using equation 3.39 and integrating over λ then gives the correlation energy in terms of the known zeroth-order polarisability:

$$\Delta E_{RPA} = \frac{1}{2\pi} \sum_{n=2}^{\infty} \frac{1}{n} \int_0^\infty d\omega \, \text{Tr}\left\{ [\mathbf{W}\mathbf{G}_0(i\omega)]^n \right\} \tag{3.42}$$
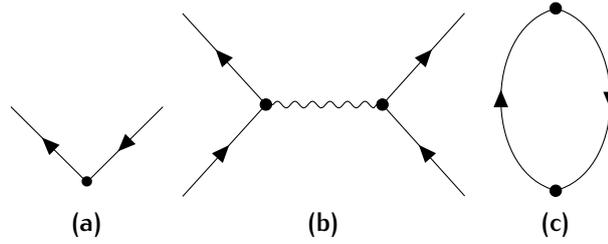
Figure 3.1: Three components of Feynman diagrams. Each vertex (solid dot) represents a term in the interaction Lagrangian, while incoming and outgoing fermions, or more precisely their propagators, are represented by the arrowed lines. The wavy line, or photon line, shows the interaction between processes in the Lagrangian.

We note in passing that this form makes it clear that the RPA energy is equivalent to that obtained in the GW approximation [235] (so-named because of the repeating units of GW in the trace), where the self-energy - the contribution to a particle's energy from its interactions with its environment - is truncated at first order.

It is enlightening in the context of noncovalent interactions to consider the case where the interaction, W, is given by that of two electric dipoles in a Coulombic field, separated by a distance R along the x-axis:

$$W_{ij} = (\nabla_i \times \nabla_j)|\mathbf{r}_i - \mathbf{r}_j|^{-1} = \frac{1}{R^3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

In this instance, the zeroth-order Green's function is simply the diagonal matrix of point dipole polarizabilities, $\alpha_i(\omega)$, so that $\text{Tr}\left\{(WG_0)^2\right\} = \alpha_1\alpha_2\text{Tr}\left\{W^2\right\}$. The remaining trace is simply $(2^2 + (-1)^2 + (-1)^2)R^{-6} = 6R^{-6}$. The second-order interaction energy is therefore

$$\Delta E_{\text{dip.}}^{(2)} = -\frac{1}{6} \cdot 6R^{-6} \cdot \frac{3}{\pi} \int_0^\infty d\omega \; \alpha_1(i\omega)\alpha_2(i\omega) = -\frac{C_6}{R^6} \tag{3.43}$$

where we have extracted the Casimir-Polder integral to get the definition of the C_6 coefficient [10, 236]. This is the familiar leading-order contribution to the dispersion energy. Similarly, by going to third order, one obtains the Axilrod-Teller-Muto term [10]. As such, the RPA approach can be seen to lead directly to the dispersion terms of the interaction energy that we wish to include in the new method.

3.3.4 Connection with coupled–cluster doubles

When the Hartree-Fock solution is taken as the reference for the above, the result is often termed time-dependent Hartree-Fock [229]. This actually goes back as far as Dirac in 1930 [237], but it was the

development of quantum field theory, and in particular the Feynman-Dyson diagrammatic techniques developed for quantum electrodynamics, that led to its application to condensed matter. The approximation was first presented by Bohm and Pines [231, 238–240] in response to attempts by Landsberg [241] and Wohlfarth [242] to describe the high-density electron gas beyond the estimate of Wigner [233, 234]. It was then formulated diagrammatically by Gell-Mann and Brueckner [243] based on Goldstone's Feynman-like diagrams [244]. This allowed for its use in the study of a much wider variety of problems. More recently, it has found considerable renewed interest due to its ability to include otherwise absent nonlocal correlation contributions in density functional theory [226, 230, 236, 245–249], and its intimate connection to coupled-cluster doubles (CCD) theory [250–255]. Not only does the latter imply applicability of the RPA to molecular systems, but it also means that efficient algorithms can be devised utilising the usual CC framework. Additionally in the current context, this will allow us to decompose the correlation energy into physically distinct contributions, as will be described later.

To understand the connection with CCD, it is easiest to use the diagrammatic approach mentioned above. Essentially, these diagrams are graph-theoretical descriptions of the underlying Lagrangian problem. It is easiest to start from the Feynman-type diagrams, the main components of which are shown in Figure 3.1. Each vertex is a term in the Lagrangian, through which particle lines pass - note that we choose time to be the horizontal axis, as is conventional in quantum chemistry. For example, Figure 3.1a shows incoming and outgoing fermions, i.e. one-particle Green's functions. Interactions between terms are then introduced as shown in Figure 3.1b; in the current context, these would represent W, and the subfigure would be read as a fermionic pair 'annihilating' to produce a virtual photon mediated by W, which in turn creates a new fermionic pair. Finally, the non-interacting two-particle Green's function, G_0, is given as the closed loop of two one-particle propagators, as shown in Figure 3.1c; the interacting Green's function G would in general be an infinite sum of similar diagrams with interaction lines included.

Using this, the RPA energy in equation 3.42 can easily be represented as the following closed-loop, or 'ring', Feynman diagrams:

Simple, systematic rules can then be used to translate each diagram into algebraic form, and indeed to perform the summation, making the value of the diagrammatic approach immediately apparent [256].

In electronic structure, where only electrons need be considered, it is simpler to instead use Goldstone diagrams [244], in which case the RPA energy becomes

$$\Delta E_{RPA} = \quad \text{} \quad + \quad \text{} \quad + \quad \text{} \quad + \dots$$

(3.44)

The interpretation of these, and thus also the rules for their evaluation, are different [222]. In the Goldstone diagrams, the lower and upper horizontals are taken to be, colloquially, the ground and excited states, respectively. Upward, 'particle' lines denote an annihilation operator, while downward, 'hole' lines denote a creation operator. The dashed lines correspond to the interaction elements connecting particle-hole pairs. Note that the prefactors from the Feynman diagrams have been absorbed, and resolvent lines (energy denominators) are implied.

Assuming that W is the bare Coulomb interaction - leading to the 'direct' RPA result (dRPA) - and that a standard HF reference is used with orbital energies ϵ_p, the first diagram in equation 3.44 is evaluated as

$$\text{} = \frac{1}{4} \sum_{iajb} \frac{|W_{iajb}|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = \Delta E_{dMP2}$$

That is, the first term in the dRPA expansion corresponds to the direct MP2 energy. Similarly, subsequent terms give higher-order perturbative corrections, such that we have performed a (selective) summation to infinite order, or 'renormalised' the perturbation series. The key to why this is possible is that the topology of all the graphs in the series is the same, i.e. they all have a ring structure [256].

However, evaluating each term in the series in the above way is both impractical and prone to error, and as we have already seen from equation 3.40, it should be possible to determine the energy self-consistently, rather than perturbatively. This is achieved using coupled-cluster theory. The HF SCF calculation gives the best possible single determinant description of the system, $|\Phi\rangle$. In the limit of a complete basis, the true wavefunction could be exactly expanded in a linear combination of all possible excited determinants, $|\Phi_{ij\cdots k}^{ab\cdots c}\rangle$ - that is, determinants where electrons are 'excited' from occupied orbitals $(ij \cdots k)$ into virtual orbitals $(ab \cdots c)$. This is written as follows:

$$|\Psi\rangle = |\Phi\rangle + t_i^a |\Phi_i^a\rangle + t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots = |\Phi\rangle + \sum_{n=1}^{n_{el.}} \hat{T}_n |\Phi\rangle \qquad (3.45)$$

where the t are amplitudes weighting the importance of each determinant, and we have defined an excitation operator, $\hat{T}_n$, that generates the weighted combination of all possible n-fold excited determinants:

$$\hat{T}_n = t_{ij\ldots k}^{ab\cdots c} a_a^\dagger a_b^\dagger \cdots a_c^\dagger a_k \cdots a_j a_i \qquad (3.46)$$

If the amplitudes in equation 3.45 were variationally optimised for all n, the result would be the full configuration interaction solution, which is exact within the space spanned by the basis set used [34]; problematically, doing so is computationally unfeasible for most systems. In the coupled-cluster approach [222], however, the wavefunction is written using an exponential ansätz, $|\Psi\rangle = \exp(\hat{T})|\Phi\rangle$, where $\hat{T}$ is the sum of all excitation operators. To linear order in the exponential, this clearly gives back the CI expansion. The advantage, though, is that when the excitation series is truncated, higher-order excitations are approximately included by virtue of the exponential. In the case of CCD, this results in

$$|\Psi_{CCD}\rangle = \exp(\hat{T}_2)|\Phi\rangle = \left[1 + \hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \ldots\right]|\Phi\rangle \qquad (3.47)$$

Diagrammatically, the partitioned Hamiltonian of equation 3.33 can be written as

$$\hat{H} = f_{pq} a_p^\dagger a_q + W_{pqrs} a_p^\dagger a_q^\dagger a_s a_r = \ \times\text{-----}\bigvee \ + \ \bigvee\text{-----}\bigvee \qquad (3.48)$$

where the cross represents a scalar, in this case the matrix element f_{pq}. Note that we suppress the arrows and vertices for simplicity, but such a diagram implies all symmetrically unique combinations of directions subject to the requirement that each vertex has exactly one incoming and one outgoing edge. Vertices will be explicitly shown when their presence would otherwise not be clear. Requiring that the wavefunction satisfies the Schrödinger equation and projecting on the left by either $|\Phi\rangle$ or all doubly-excited determinants gives the CCD energy and amplitude equations, respectively:

$$E_{CCD} = \langle\Phi|\hat{H}\exp(\hat{T}_2)|\Phi\rangle$$
$$\langle\Phi_{ij}^{ab}|\hat{H}\exp(\hat{T}_2)|\Phi\rangle = E_{CCD}\langle\Phi_{ij}^{ab}|\exp(\hat{T}_2)|\Phi\rangle \qquad (3.49)$$

We use intermediate normalisation, i.e. $\langle\Phi_{ij}^{ab}|\Phi\rangle = 1$.

Assuming strong orthogonality between the occupied and virtual subspaces, the matrix elements above will be nonzero if and only if the operators result in either a net excitation of zero for the energy and two for the amplitudes. Equation 3.48 demonstrates that the Hamiltonian will connect states separated by at most two differences in occupations. Thus, in the energy equation, only the $\hat{W}\hat{T}_2$ contribution survives:

$$E_{CCD} = W_{iajb} t_{ij}^{ab} = \bigcirc\bigcirc \qquad (3.50)$$
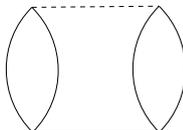
where the amplitudes are represented by particle-hole pairs connected with a solid line, as shown in Figure 3.2a.

For the amplitudes, only the linear term in the exponential is non-zero on the right, while on the left, the Hamiltonian connects the doubly excited determinant to the ground, doubly-excited, and quadruply-excited states. Therefore, these simplify to give

$$\langle \Phi_{ij}^{ab} | \hat{H} \left[1 + \hat{T}_2 + \frac{1}{2} \hat{T}_2^2 \right] | \Phi \rangle = E_{CCD} t_{ij}^{ab} \tag{3.51}$$

The left hand side can thus be split into three distinct contributions, according to their order in $\hat{T}_2$. It is easy to see that the first term is simply W_{iajb}, shown diagrammatically in Figure 3.2b. Similarly, the first-order terms incorporate the whole Hamiltonian, and the second-order terms can clearly only connect through $\hat{W}$, leading to

$$\langle \Phi_{ij}^{ab} | \hat{H} | \Phi_{kl}^{cd} \rangle \, t_{kl}^{cd} \text{ and } \langle \Phi_{ij}^{ab} | \hat{W} | \Phi_{klmn}^{cdef} \rangle \, t_{kl}^{cd} t_{mn}^{ef}$$

respectively. These can be expanded by considering all topologically distinct connections between the amplitude diagram, Figure 3.2a, and the Hamiltonian diagrams in equation 3.48 such that two particle and two hole lines are left disconnected; in the case of the second-order terms, two amplitude diagrams are included.

There is only a single such graph for the one-electron term, shown in Figure 3.2c. As we have only assumed orthogonality between the virtual and occupied subspaces, not within each subspace, this contribution is as follows:

$$\langle \Phi_{ij}^{ab} | \hat{f} | \Phi_{kl}^{cd} \rangle \, t_{kl}^{cd} = \left[f_{ab} \delta_{ij} - f_{ij} \delta_{ab} \right] t_{ij}^{ab} = \epsilon_{ia,jb} t_{ij}^{ab} \tag{3.52}$$

such that the scalar represented in Figure 3.2c is $\epsilon_{ia,jb}$. For the two-electron, first-order term, there are two topologically distinct graphs, given in figures 3.2d and 3.2e, respectively. At second-order, there are four diagrams, shown in figures 3.2f through 3.2i. Note that the disconnected, unlinked diagram in Figure 3.2i is simply $E_{CCD} t_{ij}^{ab}$, and so cancels with the right hand side in equation 3.51. The remaining diagrams can then be classified according to their topology.

Consider, for example, diagram 3.2d. This itself contains the amplitude diagram, highlighting the recursive, self-consistent nature of the CCD amplitude equations. This amplitude could then be expanded using the same diagrams; the first such term, utilising diagram 3.2b, then contracting with the interaction (diagram 3.2b) would lead to the second diagram in the RPA expansion, equation 3.44. Clearly, contracting diagram 3.2b with itself gives the first term in said expansion, while performing the same procedure with diagram 3.2f gives the fourth. In fact, it becomes apparent that by including only diagrams 3.2b, 3.2c, 3.2d, and 3.2f, the entire RPA series is recovered.

Amplitudes and zeroth-order term

(a)　　　　　　　(b)

First-order terms

(c)　　　　　(d)　　　　　(e)
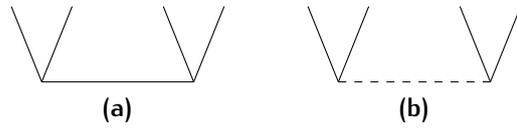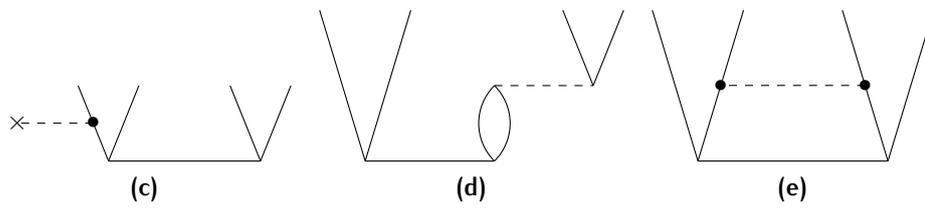
Second-order terms

(f)　　　　　　　(g)

(h)　　　　　　　(i)
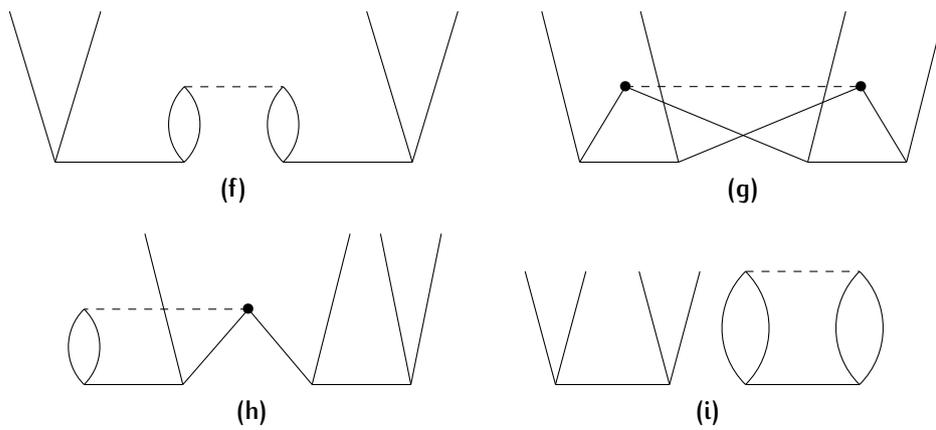
Figure 3.2: Skeleton diagrams for the contributions to the amplitude equations in coupled-cluster doubles theory, separated by order of the excitation operator, $\hat{T}_2$.

These diagrams have 'ring' topologies, as opposed to the remaining ones which have 'ladder' topologies. Only including the ring contributions is called ring coupled-cluster doubles (rCCD), and can be rigorously shown to be equivalent to RPA [252, 253]. This explains why the series is only a *selective* summation to infinite order, as ladder terms are neglected.

The RPA amplitude equations are thus, from equation 3.51:

Taking into account that there are two symmetry-distinct ways to label the particle-hole lines in the second and third diagrams, this translates into algebraic form as

$$\mathbf{W} + [\epsilon\mathbf{t} + \mathbf{t}\epsilon] + [\mathbf{tW} + \mathbf{Wt}] + \mathbf{tWt} = 0 \qquad (3.53)$$

where $\mathbf{t}$ is the amplitude tensor. In the case of dRPA [255] (or drCCD), we will denote the bare Coulomb interaction as $K_{iajb} = (ia|jb)$, and define $\mathbf{A} = \epsilon + \mathbf{K}$. The direct RPA equations are thus

$$\Delta E_{dRPA} = \frac{1}{2}\text{Tr}\{\mathbf{Kt}\}$$
$$\mathbf{R} = \mathbf{K} + \mathbf{At} + \mathbf{tA} + \mathbf{tKt} = 0 \qquad (3.54)$$

where $\mathbf{R}$ is the RPA residual. This is a Riccatti equation, for which many simple and efficient methods exist [257], allowing iterative determination of the RPA energy. In addition, explicitly computing the amplitudes will allow us to decompose this energy into different contributions.

3.3.5 Using ALMOs in RPA

In the above, we have assumed strong orthogonality between the occupied and virtual subspaces, and orthonormality within those subspaces. The bare ALMO creation and annihilation operators, of equation 2.33, do not satisfy this requirement. For the development of the ALMO SCF routine, we half-orthogonalised these in the occupied subspace, and for the charge transfer treatment earlier, we projected the virtual subspace out of the occupied subspace. We do this again here, then orthogonalise the resulting virtual subspace in the same way. That is, we define our creation operators as follows, where the first line is equivalent to equation 2.34:

$$a_{i\sigma}^{\dagger} = \left[\sigma^{-1/2}\right]_{ij} b_{j\sigma}^{\dagger}$$
$$a_{a\sigma}^{\dagger} = \hat{q} b_{b\sigma}^{\dagger} \left[\pi^{-1/2}\right]_{ba} \qquad (3.55)$$

We have defined the virtual-virtual block of the MO overlap matrix as $s_{VV} = \pi$.

Using the operators of equations 3.55 allows us to use all of the results outlined in the previous section, but does not yield any of the advantages that the localised orbitals would - the entire point of using these orbitals is to exploit their locality to achieve a linear-scaling algorithm later. As such, we need to back-transform the equations to the ALMO basis. This is done by careful application of the above relations to equations 3.54 in the case of dRPA. For clarity, we will denote quantities in the transformed basis with a tilde, such that for example the residual is

$$\tilde{\mathbf{R}} = \tilde{\mathbf{K}} + \tilde{\mathbf{A}}\tilde{\mathbf{t}} + \tilde{\mathbf{t}}\tilde{\mathbf{A}} + \tilde{\mathbf{t}}\tilde{\mathbf{K}}\tilde{\mathbf{t}} = 0$$

Therefore, for example, the bare interaction tensor, $\mathbf{K}$, would be related to $\tilde{\mathbf{K}}$ as

$$\tilde{\mathbf{K}} = \sigma^{-1/2}\pi^{-1/2}\mathbf{K}\pi^{-1/2}\sigma^{-1/2}$$

and similarly for the amplitudes and $\mathbf{A}$ tensor. The residual thus becomes

$$\begin{aligned}
\tilde{\mathbf{R}} = \sigma^{-1/2}\pi^{-1/2}\{&\mathbf{K} + \mathbf{A}\pi^{-1/2}\sigma^{-1/2}\sigma^{-1/2}\pi^{-1/2}\mathbf{t} + \mathbf{t}\pi^{-1/2}\sigma^{-1/2}\sigma^{-1/2}\pi^{-1/2}\mathbf{A} \\
&+ \mathbf{t}\pi^{-1/2}\sigma^{-1/2}\sigma^{-1/2}\pi^{-1/2}\mathbf{K}\pi^{-1/2}\sigma^{-1/2}\sigma^{-1/2}\pi^{-1/2}\mathbf{t}\}\pi^{-1/2}\sigma^{-1/2} = 0
\end{aligned}$$

Upon switching to index notation, the residual can be greatly simplified, as the metric terms combine. Also, as the metrics must be positive definite, the term in brackets must equivalently be zero. Therefore the residual equation to be solved in the ALMO basis is

$$\begin{aligned}
R_{iajb} = &K_{iajb} + A_{iakc}\sigma^{kl}\pi^{cd}t^{db}_{lj} + t^{ac}_{ik}\pi^{cd}\sigma^{kl}A_{ldjb} \\
&+ t^{ac}_{ik}\pi^{cd}\sigma^{kl}K_{ldme}\sigma^{mn}\pi^{ef}t^{fb}_{nj} = 0
\end{aligned} \tag{3.56}$$

Similarly, the energy becomes

$$E_{dRPA} = \frac{1}{2}\left[\sigma^{ik}\pi^{ac}K_{kcld}\pi^{db}\sigma^{lj}\right]t^{ab}_{ij} \tag{3.57}$$

The simplicity of the derivation of these equations involving highly nonorthogonal orbitals again demonstrates the utility of the second-quantized approach we have used to develop the ALMO theory. It has also led to the formulation in terms of localised amplitudes, which will form the crux of our linearisation procedure in the next chapter.

3.3.6 Extension to higher orders

The RPA kernel is the simplest form of coupled cluster-type equation; apart from its physical justification for describing dispersion - as was discussed earlier - its main advantage is in the relatively low

Contribute to $\hat{T}_2$ amplitudes

(a) (b)

Contribute to $\hat{T}_3$ amplitudes

(c) (d) (e)

(f) (g) (h)

(i) (j) (k)
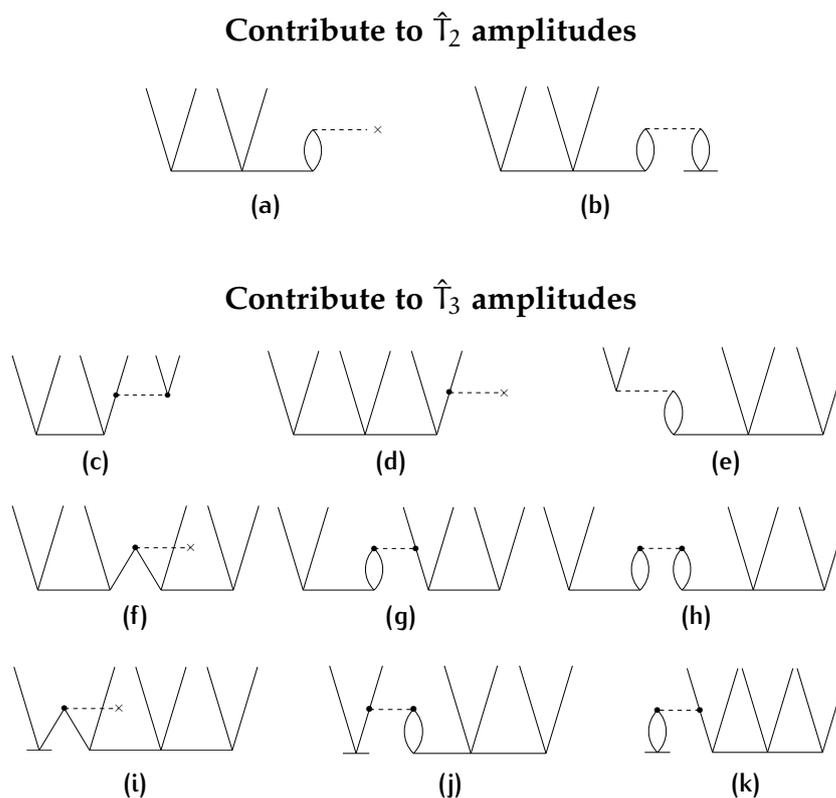
Figure 3.3: All ring-type diagrams arising from the inclusion of the $\hat{T}_3$ excitations in the coupled cluster formalism, categorised by whether they contribute to the $\hat{T}_2$ or $\hat{T}_3$ amplitude equations. Of the latter, only diagrams **(c)** and **(d)** appear up to second order in the wavefunction, so form the basis of an approximate triples correction.

computational cost. Standard CCD methods scale as the sixth-order of the system size, whereas analysis of the dRPA residual equation 3.56, when factored into separate matrix multiplications, merely shows a fourth-order dependence. The full CCD has the advantage of explicitly including exchange, however, and in the next section we will see how inclusion of this in RPA complicates matters slightly, leading to $\mathcal{O}(N^5)$ scaling. If accuracy were more important than efficiency, there is nothing to prevent the extension of the above to include higher orders, i.e. triples and above. The ALMO localisation could be applied in the same manner, but with additional amplitudes being included. These could then be assigned to trios of fragments, similarly to the doubles amplitudes, and decomposed to exploit the locality. This would lead to a factorial increase in the number of interactions, though, and so is not feasible to be done exactly. Investigations into the possible use of stochastic approaches [258] to circumvent this problem are ongoing - in principle, it would require no significant changes to the theory or implementation of the overall method.

A different approach often used in coupled cluster theory is to approximately include the triple excitation terms by perturbation [222,

259–261]. The associated diagrams are not of the ring type, however, so the perturbative triples formalism cannot be directly applied to RPA. Instead, here we will discuss diagrammatically how to systematically include higher-order effects in the RPA method; the present form of correction, to the author's knowledge, has not been attempted before for any form of RPA.

The Goldstone diagrams of the ring type representing the full triples ($\hat{T}_3$) terms in CCSDT are given in Figure 3.3, split by whether they contribute to the $\hat{T}_2$ amplitudes or the $\hat{T}_3$ amplitudes; for the latter, the directionality of the particle-hole lines has been removed to save space, and all valid combinations are implied. These form only a small subset of the total number of diagrams, of which there are six of the first type, and 47 of the second [222]. However, note that the generation of these diagrams can be done systematically, as described earlier, and rapidly truncated by whether they are ring-type diagrams or not. In fact, packages exist that can automatically generate (admittedly somewhat inefficient) implementations of arbitrary-order coupled cluster equations [262–265], which could in principle be modified via this classification. We focus only on the third-order terms here as they are likely to be the most significant [34].

As before, the full triples terms can be read off from the diagrams; this is convoluted for the triples amplitude equations, but reasonably simple for the triples *corrections* to the doubles amplitudes. From Figures 3.3a and 3.3b, we see that when added to the RPA residual of equation 3.53, we get

$$\mathbf{R} = \mathbf{W} + [\epsilon(\mathbf{t}_2 + \mathbf{t}_3) + \mathbf{W}(1 + \mathbf{t}_1)\mathbf{t}_2 + \text{c.c.}] + \mathbf{t}_2\mathbf{W}\mathbf{t}_2 = 0 \qquad (3.58)$$

where $\mathbf{t}_n$ are the n-th order amplitudes, and c.c. stands for complex conjugate. Note the inclusion of the singles and triples amplitudes, which could be found exactly through solution of their respective residual equations. This, however, necessarily couples each residual equation, greatly increasing the complexity of the procedure. The single excitations can be safely ignored in most cases, making negligible contributions - they could even rigorously be avoided by transforming to a Brueckner formulation if desired [222, 266]. The triples could then be perturbatively approximated by careful analysis of the triples residual. The resulting theory, RPA(T), could offer significant improvements in the accuracy of the dispersion energy, similar to how CCSD(T) outperforms CCSD [267–270].

In the standard CCSD perturbative triples correction, we eliminate all diagrams that couple the triples amplitudes with the fluctuation potential [271]. In our case, with only ring diagrams included, this is equivalent to using only the diagrams in Figures 3.3c and 3.3d,

remembering to take all combinations of the particle-hole lines. The resulting equation is

$$\epsilon_{ijk}^{abc} t_{ijk}^{abc} = \hat{P}(k/ij|a/bc) \left[W_{iadb} t_{jk}^{dc} - W_{iajl} t_{lk}^{bc} \right] \tag{3.59}$$

where the operator $\hat{P}(k/ij|a/bc)$ cyclically permutes the indices kij and abc, e.g.

$$\hat{P}(i/jk) = \hat{I} - \hat{\pi}_{ik} - \hat{\pi}_{jk}$$

with $\hat{P}(k/ij|a/bc) = \hat{P}(k/ij)\hat{P}(a/bc)$. The term ϵ_{ijk}^{abc} is the triples equivalent of equation 3.52, which in the canonical scheme would equate to the sum of energy differences between the occupied and virtual orbitals, and in the non-canonical scheme is given by

$$\epsilon_{ijk}^{abc} = \hat{P}(k/ij|a/bc)\epsilon_{ij}^{ab} \tag{3.60}$$

From this, given doubles amplitudes solved in the usual way via equation 3.53, equation 3.59 can easily be solved to give the approximate triples amplitudes. These can then be used in the standard formulation to compute a triples correction to the energy [271]. Alternatively, they could be put back into equation 3.58 to solve for new doubles amplitudes, a process which in principle could iteratively be used to approximate the full triples term.

3.4 EXCHANGE

Direct RPA performs very well in the long range [232, 255, 272], especially as it leads to the correct R^{-6} dependency, as demonstrated in equation 3.43. In the short range, however, it provides a much poorer description; for example, the pair-correlation function in the high-density electron gas becomes negative at small separations [272]. This is due to the missing exchange terms in the Coulomb interaction [273], i.e. the fact that antisymmetrised integrals are not used. At short separations, exchange-repulsion dominates as the distributions of electrons of the same spin attempt to overlap. Several different schemes have been suggested to ameliorate this problem [232, 255, 273], one of the simplest and most successful of which is simply to contract the energy with the antisymmetrised integrals:

$$B_{iajb} = (ia|jb) - (ib|ja)$$

Note that these are in terms of spin-orbitals - in the spin-restricted case, the Coulomb term gains a factor of two. This is called the second-order screened exchange correction (SOSEX) [274], which is identical to equation 3.54, but with the energy evaluated as

$$\Delta E_{RPA+SOSEX} = \frac{1}{2} \text{Tr} \{ \mathbf{Bt} \} \tag{3.61}$$

This includes exchange to the lowest-order in perturbation. If we also replace $\mathbf{K}$ with $\mathbf{B}$ in the amplitude terms, including in $\mathbf{A} = \epsilon + \mathbf{B}$, then we get RPA plus exchange, or RPAx [255]:

$$\Delta E_{\text{RPAx}} = \frac{1}{4} \text{Tr}\{\mathbf{Bt}\}$$

$$\mathbf{R} = \mathbf{B} + \mathbf{At} + \mathbf{tA} + \mathbf{tBt} = 0$$

(3.62)

Note the additional factor of a half, necessary to avoid double counting.

This is equivalent to including all contractions of the amplitudes with the interaction that do not form 'loops', i.e. where one particle and one hole line have been 'exchanged'. That is, in analogy to equation 3.50:
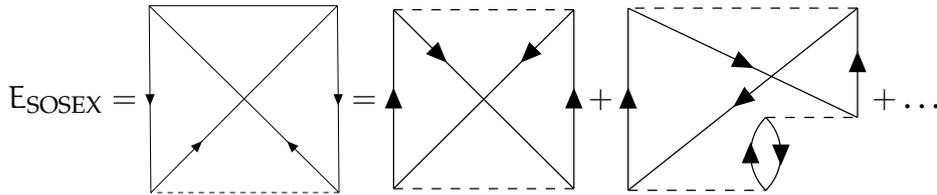
This is true also for RPAx, but with exchange included in the Coulombic screening when determining the amplitudes. The second-order contribution in RPAx is thus

and similarly for higher orders, although ladder terms are of course missing. As such, this approach can be seen to lie somewhere between perturbative methods and CCD.

The main problem with both the SOSEX and RPAx corrections, however, is that they only include higher-order diagrams inexactly, leading to a spurious self-interaction error in the short range [275, 276]. One solution to this would be to generate the relevant diagrams and explicitly add these as a correction [277]; such an approach would greatly increase the complexity of the equations to be solved, and bring it into similar territory as a full CCD calculation. An alternative that could offer lower scaling than either method would be to draw an analogy with the GW approximation truncated at first order, as noted after equation 3.42. An essentially exact second-order screened exchange correction to the GW approximation can be developed by analogy to the above [278].

In the GW approximation [279, 280], we deal with the self-energy, Σ, defined as

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2) = i \int d\mathbf{r}_3 \, d\mathbf{r}_4 \, G(\mathbf{r}_1, \mathbf{r}_3) W(\mathbf{r}_1, \mathbf{r}_4) \Gamma(\mathbf{r}_4, \mathbf{r}_2, \mathbf{r}_3) \qquad (3.63)$$

where for clarity we have switched from a discrete to a continuous representation - i.e. $G_{pq,rs} = G(\mathbf{r}_1, \mathbf{r}_2)$ where pq are orbitals as functions of $\mathbf{r}_1$ and rs as functions of $\mathbf{r}_2$. The function Γ is known as the vertex function, and controls the order to which the GW approximation is taken [281]. At first order it is given by

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \approx \delta(\mathbf{r}_1 - \mathbf{r}_2)\delta(\mathbf{r}_1 - \mathbf{r}_3)$$

which upon insertion into equation 3.63 yields

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2) \approx iG(\mathbf{r}_1, \mathbf{r}_2)W(\mathbf{r}_1, \mathbf{r}_2)$$

From this, the Dyson equation of equation 3.40 can be recast in terms of the self energy by right-multiplying by the fluctuation W, to give

$$\Sigma_0 = \Sigma + i\lambda\Sigma\Sigma_0$$

The derivation of the energy then follows to equation 3.42 as before; however, we now have a procedure for *arbitrary* orders of the self energy, by including higher terms in the vertex function [282–284]. This could lead to an alternative to the higher-order methods described in the previous section, incorporating correlation beyond the random phase approximation [285, 286].

Here, however, we are only concerned with including exchange. To do so, we follow the approach of density functional theory [278], and partition the self energy into Coulomb and exchange, rewriting in terms of the screened fluctuation:

$$\begin{aligned}
\Sigma_c &= \int d\mathbf{r}_3 d\mathbf{r}_4 G(\mathbf{r}_1, \mathbf{r}_2)V(\mathbf{r}_1, \mathbf{r}_3)G(\mathbf{r}_3, \mathbf{r}_4)G(\mathbf{r}_4, \mathbf{r}_3)W(\mathbf{r}_4, \mathbf{r}_2) \\
\Sigma_x &= \int d\mathbf{r}_3 d\mathbf{r}_4 G(\mathbf{r}_1, \mathbf{r}_4)V(\mathbf{r}_1, \mathbf{r}_3)G(\mathbf{r}_4, \mathbf{r}_3)G(\mathbf{r}_3, \mathbf{r}_2)W(\mathbf{r}_4, \mathbf{r}_2)
\end{aligned} \tag{3.64}$$

where V is the bare Coulomb potential, and the screened fluctuation under the full vertex can be written iteratively as

$$W(\mathbf{r}_1, \mathbf{r}_2) = V(\mathbf{r}_1, \mathbf{r}_2) + \int d\mathbf{r}_3 d\mathbf{r}_4 V(\mathbf{r}_1, \mathbf{r}_3)\chi_0(\mathbf{r}_3, \mathbf{r}_4)W(\mathbf{r}_4, \mathbf{r}_2) \tag{3.65}$$

where χ_0 (the zeroth-order polarisability) is defined as a subset of equation 3.31.

In principle it is possible to yield the exact second-order screened exchange by iterative application of equations 3.64 and 3.65. In terms of computational scaling, these are $\mathcal{O}(N^5)$, and with suitable choice of integration grids, this is with a very low prefactor; density fitting could also reduce the cost greatly. The primary problem, however, is relating the quantities calculated in this way to the amplitudes from the rCCD-type methods. As will be seen in the next chapter, these are necessary for the linear scaling with number of fragments that we will achieve. An interesting avenue of further investigation would be to derive, either approximately or analytically, the connection so that we can completely correct for the problems with exchange in the RPA formalism.

3.5 IMPLEMENTATION

3.5.1 ALMO with charge transfer

The final step of the ALMO routine is the perturbative correction and optionally the decomposition into pairwise charge transfer terms. To do this, we need to iteratively solve equation 3.15 for the amplitudes, X_{ai}. This is a particularly simple form of Riccati equation, which has been studied extensively in the field of optimal control [257]. Perhaps the simplest approach utilises the fact that the Fock matrix is diagonally dominant, so that we can guarantee convergence by extracting the diagonal elements as follows:

$$(f_{aa} - f_{ii})x_{ai}^{(n+1)} = -f_{ai} - \sum_{b\neq a} f_{ab}x_{bi}^{(n)} - \sum_{j\neq i} x_{aj}f_{ji} - \sum_{jb} x_{aj}^{(n)} f_{jb}x_{bi}^{(n)}$$

(3.66)

This is repeated until $\left\|x^{(n+1)} - x^{(n)}\right\|$ is less than a given threshold - typically, 10^{-4} in the Frobenius norm is sufficiently tight, usually requiring fewer than ten iterations. Clearly, the most expensive part of this is the quadratic term, which scales as $\mathcal{O}(O^2V^2)$, where O and V are the total number of occupied and virtual orbitals in the system, respectively . While this is formally quartic, in practice it is quadratic as the number of occupied orbitals is only a small fraction of the basis size. As these are simply matrix multiplications the prefactor is anticipated to be very small, such that this will only dominate the SCF calculation in the limit of very large systems. Neglecting the quadratic term, which as noted before leads to the second-order perturbative correction, would reduce the prefactor further, but would not eliminate the $\mathcal{O}(V^2)$ dependency. Regardless, the charge transfer calculation is a much less significant contribution to the cost than the calculation of the integrals.

As mentioned earlier, we enforce strong orthogonality between the occupied and virtual subspaces by projecting the virtual coefficients out of the occupied subspace. This is done in the atomic orbital basis by the mapping $\tilde{V} \leftarrow QV$. In the AO basis, the charge transfer energy and residual then read as

$$E_{CT} = 2\text{Tr}\left\{T^\dagger FQVx\right\}$$ (3.67)

$$R_{CT} = V^\dagger Q^\dagger FT + V^\dagger Q^\dagger FQVx - xT^\dagger FT - xT^\dagger FQVx = 0$$ (3.68)

3.5.2 Dispersion and exchange–dispersion

The dynamical electron correlation in this method is determined using RPA, which encompasses to some extent infinite-order many-body effects on the single particle level. In a naive treatment, the

correlation energy would be determined for each fragment and the complex separately, and the interaction taken as the difference. This would of course reintroduce superposition errors, and would require more expensive monomer calculations.

Here, I discuss two ways the former problem could be alleviated. Firstly, by using range separation, as dispersion is inherently a long-range effect. However, this introduces an arbitrary range-separation parameter, and could possibly lead to discontinuities in the potential. It also requires the evaluation and transformation of a second set of two-electron integrals. A second, in many ways better, approach is to try to classify the different excitation contributions in the RPA framework.

Range separation to eliminate BSSE

The RPAx method, just as for any other correlated method, will likely contain a large basis set superposition error due to the importance of interactions between occupied and virtual orbitals on separate fragments. As these cross-fragment terms contain the very interactions that we are trying to describe, it is not possible to try and restrict contributions in a similar way to the ALMO approach. However, the ALMO reference does provide highly localised orbitals, even after the virtual subspace is projected out of the occupied subspace. This in itself somewhat minimises the error.

In the context of DFT, range-separated RPA has been exploited by several different groups as a way of accounting for the long-range correlation inherently missing from most density functionals [226, 245, 247, 255, 274, 287, 288], while avoiding double-counting of the short-range contributions that are already well described. This is achieved by attenuating the Coulomb interaction by the error function as follows:

$$\hat{V}_{lr} = \frac{1}{2} \sum_{ij} \frac{\mathrm{erf}(\mu r_{ij})}{r_{ij}} \tag{3.69}$$

The parameter μ controls the range-separation, by attenuating how rapidly the error function monotonically increases from zero to unity - it has units of reciprocal Bohr, and roughly speaking $1/\mu$ corresponds to the distance beyond which is considered to be long range. The primary advantage of separating in this manner is that the evaluation of two-electron integrals in this potential requires only minor modifications to existing routines. Range-separation is one way to try and isolate only the interaction contributions, although it is not clear whether this will help in eliminating BSSE. An alternative, perhaps more useful approach will be given in the next section, where we consider how to partition the correlation into physically distinct terms.

Decomposing the excitations

As was demonstrated earlier, RPA entails finding a set of amplitudes for all possible doubly-excited determinants, $|\Phi_{ij}^{ab}\rangle$. As a result, such excitations can involve at most four fragments (one for each index). This implies that, while the method does include many-body dispersion at the particle level, it cannot describe higher than four-body terms at the fragment level. Diagrammatic representations of all possible inter-fragment excitations are shown in Figure 3.4, classified by whether they involve two, three, or four bodies.

We identify five categories of excitation based on what they physically represent. The first, which is not shown in Figure 3.4 is intramolecular, when all orbitals involved are on the same fragment. As these will only be slightly perturbed by the presence of the rest of the complex, it is reasonable to approximate that these do not form part of the interaction energy. Dispersion is taken to be the response of on-fragment excitations to simultaneous excitation within a separate fragment, i.e. the dispersive coupling between the separated fragment electron densities, as related through the adiabatic connection fluctuation-dissipation relations outlined earlier. This corresponds to class a) in the figure, while class b) is exchange-dispersion. This is again a number-conserving coupling of excitations involving exchange between fragments. It should be noted that this is not well-separated from the dispersion term, as the use of a fully anti-symmetrised reference wavefunction means that a certain degree of exchange-repulsion is already included in all of the terms. As such, this is not directly comparable with the equivalent term in SAPT. The dispersive energy contributions are therefore taken from equation 3.62 (in the case of RPAx) as

$$\Delta E_{\text{disp.}} = \sum_X \sum_{Y>X} \sum_{(ia)\in X} \sum_{(jb)\in Y} t_{ij}^{ab} \left[\sigma^{-1}\pi^{-1}\mathbf{B}\pi^{-1}\sigma^{-1}\right]_{ij}^{ab} \qquad (3.70)$$

$$\Delta E_{\text{exch-disp.}} = \sum_X \sum_{Y>X} \sum_{(ib)\in X} \sum_{(ja)\in Y} t_{ij}^{ab} \left[\sigma^{-1}\pi^{-1}\mathbf{B}\pi^{-1}\sigma^{-1}\right]_{ij}^{ab} \qquad (3.71)$$

The remaining classes in Figure 3.4 are labelled as 'ionic' and 'BSSE'. These are both excitations that involve excitation from X to Y, without a corresponding excitation into X. It should be stressed that these excitations categorically do not involve the transfer of electrons, they are rather couplings between different states in Fock space. The term ionic is simply denoting that the excitations are not particle number-conserving within a fragment. These represent a long-range polarisation effect, i.e. a form of 'charge transfer' at the correlated level. In particular, it is supposed that diagrams (e) and (i), which are double excitations from the same fragment into distant virtual orbitals on separate fragments, represent an attempt to improve the description

Two fragments

(a) Dispersion

(b) Exchange-dispersion

(c) Ionic

(d) Ionic

(e) BSSE

Three fragments

(f) Ionic

(g) Ionic

(h) Ionic

(i) BSSE

Four fragments

(j) Ionic
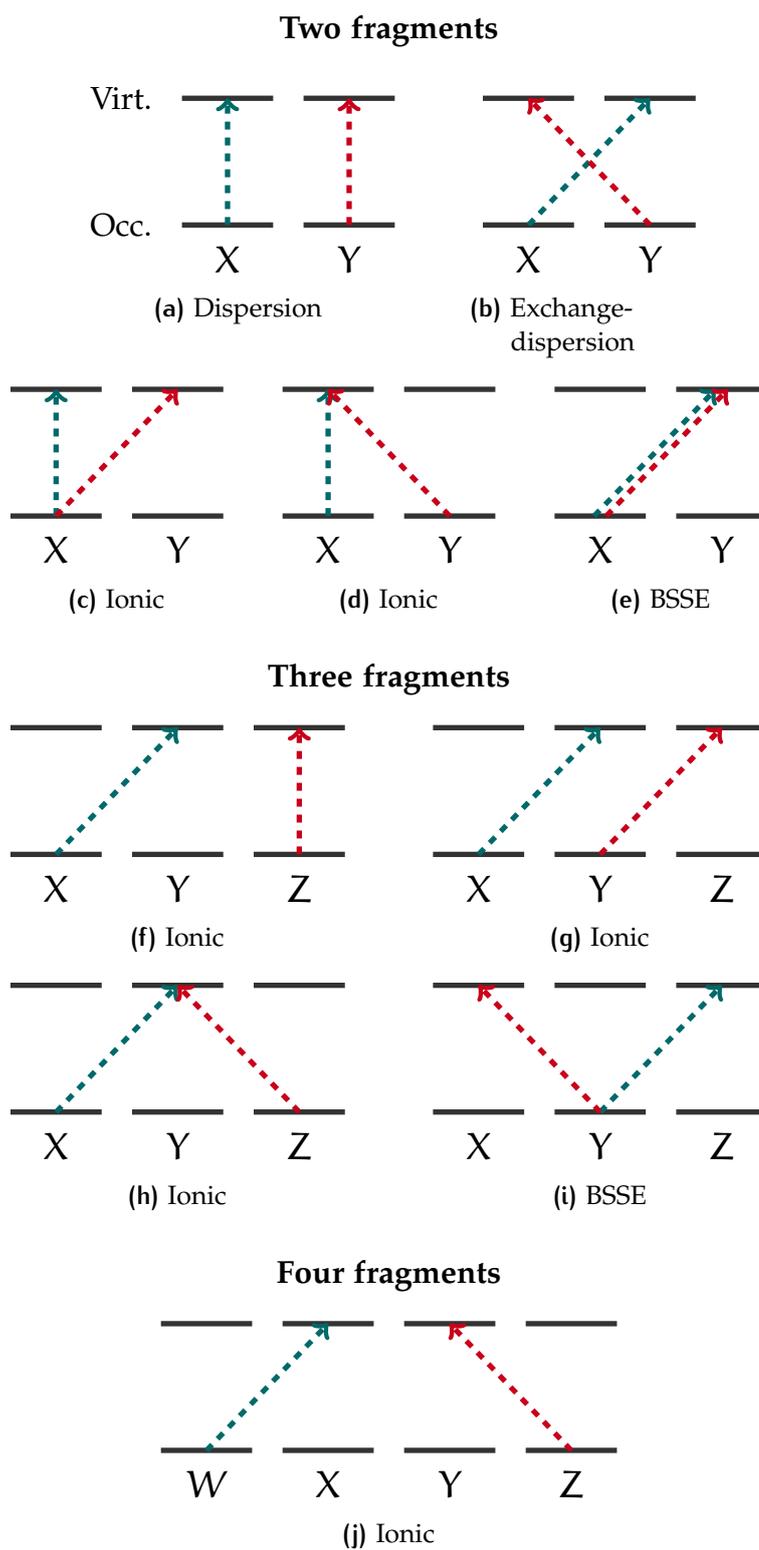
Figure 3.4: Schematics of all unique classes of inter-fragment excitations.

of *intramolecular* correlation using the extended basis, following a similar suggestion by Schütz when developing local MP2 [37]. In reality, while these may be the main contributors to BSSE, any term involving orbitals not on the same fragment will most likely contain some amount of superposition error. This includes the exchange-dispersion term above (but not the pure dispersion), although in practice this turns out not to be the case. As the RPA is done using the ALMO density, albeit with the virtual subspace projected out of the occupied subspace, the ionic terms will not only contain some measure of BSSE, but also a significant amount of the charge transfer effects that would be encompassed at the mean-field level by ΔE_{CT}. This means that only one of the two contributions should be included. As BSSE is known to be substantially larger at the correlated level, we choose to use the latter as defined in equation 3.16, and neglect diagrams (c), (d), (f)—(h), and (j) in Figure 3.4.

4 | A LINEAR SCALING METHOD

The physical justification and theoretical description of the method outlined so far is not the only aspect that needs to be considered. Equally important is the question of how to implement it in such a way that computations are efficient and widely applicable. In this chapter, I outline the practical considerations behind this, paying particular attention to the scaling of the method with system size. As our approach is specialised to fragmented, non-covalent interactions, there are two scaling regimes: scaling with fragment size, and with the number of fragments. The latter magnifies the former considerably, and therefore is in general the largest source of computational cost; effectively, it is a measure of the entire system size. As such, this is the facet that I will focus on improving, culminating in a novel, linear-scaling procedure.

There are essentially five portions of the calculation that incur significant computational expense:

1. evaluation and storage of the two-electron integrals;

2. formation of the Fock matrix;

3. solution of the coupled on-fragment Roothaan-Hall equations;

4. transformation of the two-electron integrals into the molecular orbital basis;

5. solution of the RPA amplitude equations.

The first of these indicates that scaling with processing cost, i.e. the number of floating point operations (FLOPs), is not the only issue, but also the requirements on memory or disk resources. In fact, the latter is often the more pressing concern, as physical resources are inherently finite, whereas time is in somewhat greater supply.

The scaling can be formally measured in terms of the number of atomic basis functions, N, the number of fragments F, and the total number of occupied and virtual molecular orbitals, O and V. In the asymptotic limit, we can consider all fragments to have essentially the same number of orbitals, such that $N = Fn$, $O = Fo$, and $V = Fv$. The per-fragment cost is thus in terms of n, o, and v, whereas the order of

F determines the scaling with overall system size. Description as an $\mathcal{O}(f(N))$ process, where f is some function, implies that asymptotically

$$\text{resource} \rightarrow \alpha f(N) \text{ as } N \rightarrow \infty \qquad (4.1)$$

where α is termed the scaling prefactor. Note that in actual computations, the asymptotic limit is often not reached, such that the prefactor can be as important as the formal scaling.

Now consider each of the above sections. The two-electron or electron repulsion integrals (ERIs) are four-centre quantities over the atomic orbitals, thus formally scaling as $\mathcal{O}(N^4) = \mathcal{O}(F^4 n^4)$. The most expensive part of the Fock build can be seen from equation 2.22 to be the formation of the Coulomb and exchange matrices. Equation 2.30 demonstrates that this involves contraction of the ERIs with the density matrix, involving four distinct indices, making this also an $\mathcal{O}(N^4)$ process. In practice, the prefactor for the exchange contribution is larger than that for the Coulomb matrix, as it involves contraction over 'mixed' indices - that is, one contracted index lies in each of the bra and ket parts of the ERI. The solution of the SCF equations involves matrix diagonalisations, which scale as $\mathcal{O}(n^3)$ per fragment, giving overall $\mathcal{O}(Fn^3)$, making this linear in total system size. This is an improvement already implicit in the model, as in a full calculation this would be expected to cost $\mathcal{O}(N^3) = \mathcal{O}(F^3 n^3)$. As has been discussed earlier, the integral transformation is $\mathcal{O}(N^5)$. However, as only the $(ia|jb)$ integrals are needed, the most expensive quarter transformation is $\mathcal{O}(ON^4) = \mathcal{O}(F^5 o n^4)$, such that the on-fragment scaling is less severe than the cost with increasing number of fragments. Finally, equation 3.62 shows that the most expensive step in the RPA amplitudes is the contraction of three four-index tensors:

$$[\mathbf{tBt}]_{ij}^{ab} = t_{ik}^{ac} B_{kcld} t_{lj}^{db} = t_{ik}^{ac} [\mathbf{Bt}]_{kj}^{cb}$$

Naively, this appears to be $\mathcal{O}(O^4 V^4)$, but factorising as shown results in two $\mathcal{O}(O^3 V^3)$ contractions. This is therefore $\mathcal{O}(F^6)$ overall compared to the less expensive $\mathcal{O}(o^3 v^3)$ per fragment. Moreover, the residual amplitude equations need to be performed over multiple iterations, making this easily the most expensive part of the calculation.

4.1 DENSITY FITTING

As mentioned earlier, the first and fourth points above can be ameliorated by using density fitting (DF). This requires the evaluation of two- and three-centre Coulomb integrals, $(P|Q)$ and $(\mu\nu|P)$, the latter being given in equation 1.11. We use capital letters to denote functions in an auxiliary basis set (ABS), $\{\chi_P\}$, while lowercase refers to either

the atomic or molecular orbital bases (OBS) as per usual. A method for the evaluation and subsequent transformation to the MO basis of the DF ERIs, which makes full use of their inherent symmetries, is given in Algorithm 4.1 below. In this way, we have that the necessary

---

Algorithm 4.1 Conventional density-fitting of two-electron integrals

1: **for all** μ, $\nu \leqslant \mu$ in OBS, P, Q $\leqslant$ P in ABS **do**
2:     $B_{\mu\nu}^{P} \leftarrow (\mu\nu|P)$
3:     $G_{PQ} \leftarrow (P|Q)$
4: **end for**
5: Cholesky decompose $\mathbf{G} = \mathbf{LL}^{\mathsf{T}}$
6: Solve $\mathbf{LL}^{-1} = \mathbf{I}$
7: $\mathbf{b} \leftarrow \mathbf{BL}^{-1}$
8: **for all** i in occ., a in virt. **do**
9:     $\tilde{k}_{i\nu}^{P} \leftarrow \sum_{\mu} b_{\mu\nu}^{P} C_{\mu}^{i}$
10:     $\tilde{b}_{ia}^{P} \leftarrow \sum_{\nu} \tilde{k}_{i\nu}^{P} C_{\nu}^{a}$
11: **end for**

ERIs in the AO and MO bases can be formed via equation 1.12 as

$$(\mu\nu|\lambda\tau) = \left[\mathbf{bb}^{\mathsf{T}}\right]_{\mu\nu}^{\lambda\tau} \text{ and } (ia|jb) = \left[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{\mathsf{T}}\right]_{ia}^{jb} \tag{4.2}$$

Inspection of the loop structure in Algorithm 4.1 shows that the second line requires $\mathcal{O}(N^2M)$ computations, while the third and seventh are nominally $\mathcal{O}(M^2)$, with the matrix-matrix multiplication being the cheaper of the two. The Cholesky decomposition and subsequent solution of the lower triangular matrix equation is $\mathcal{O}(M^3)$ but with a very small prefactor, while lines nine and ten are $\mathcal{O}(ON^2)$ and $\mathcal{O}(OVN)$, respectively. We therefore conclude that the overall scaling is cubic, and for most systems will be dominated by the evaluation of the three-centre integrals; note, however, that for very large systems, the Cholesky decomposition will become the bottleneck. Regardless, this is a sizeable improvement on the quartic and quintic evaluation and transformation steps that would otherwise be required. This procedure can be used in both the ALMO and RPA portions of the calculation, but in general different auxiliary bases will be needed for each. The cost of using density fitting is therefore a minor increase in programmatic complexity and the need to evaluate two such sets of integrals, although only one set needs to be transformed. An additional advantage in the ALMO case, however, is that the decomposition $\mathbf{b}$ can be used directly removing the need to ever form and store the four-index integrals.

Another major advantage to using DF is the reduction in memory cost, from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2M)$. For example, in the case of water dimer in the aug-cc-pVDZ basis [128] with the matched JK-fitting set

of Weigend [151, 162, 289] (comprising 205 and 750 functions in the orbital and auxiliary sets, respectively), approximately 8 MB of storage are needed for the density-fitted integrals compared to 88 MB for the full set of ERIs. Moving to the water pentamer, this becomes 0.12 GB versus 3.3 GB. However, the cubic scaling in both FLOPs and memory still becomes problematic when larger complexes are considered - a cluster of one hundred water molecules would require almost a terabyte of memory, which is completely unfeasible. The usual solution would then be to avoid storage in memory either by reevaluating the integrals in each iteration and using them as needed (called the 'direct approach' [290]), or to store the integrals on disk in batches [291], reading them in each iteration (the 'out-of-core' approach). While efficient screening can be achieved by neglecting shell-blocks that would not make a significant contribution to the Fock matrix [34, 292], formation of the Coulomb contribution to the Fock matrix would require two integral evaluations per iteration, as will be described later. Coupled with the formally cubic processing cost, the direct approach quickly becomes impractical. The out-of-core procedure on the other hand removes the need for reevaluation of the integrals, but at the cost of substantial communication between the processor and physical storage. The associated overhead per iteration can often be several orders of magnitude larger than the cost of the rest of the calculation.

4.1.1 Sparsity

It would seem that the severe cost with increasing system size is unphysical, as intuitively one expects the overlap between the charge distributions on each fragment to vanish rapidly with distance. The electron repulsion integral between two one-particle densities $\rho_p(\mathbf{r}_1)$ and $\rho_q(\mathbf{r}_2)$ is given by

$$V_{pq} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, \frac{\rho_p(\mathbf{r}_1)\rho_q(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{4.3}$$

Assuming these densities to be spherical Gaussians, with centres $\mathbf{P}$, $\mathbf{Q}$ and exponents p, q, it can be shown that (see Appendix D)

$$V_{pq}^0 = \sqrt{\frac{4\alpha}{\pi}} F_0(\alpha R_{PQ}^2)$$

where $\alpha = pq/(p+q)$ and $R_{PQ} = |\mathbf{P} - \mathbf{Q}|$. This is written in terms of a Boys function, which are more generally given as [34]:

$$F_n(x) = \int_0^1 dt \, t^{2n} \exp(-xt^2) \tag{4.4}$$

and V_{pq}^n for arbitrary angular momenta Gaussians can be written in terms of this.

Clearly, the integrand in equation 4.4 vanishes as $\exp(-\alpha R_{PQ}^2)$, such that we expect a fairly rapid decay of the integral itself. For large R_{PQ}, we can use this to extend the region of integration to the whole real line, allowing for analytical evaluation:

$$V_{pq}^n \approx \sqrt{\frac{4\alpha}{\pi}} \int_0^\infty dt\, t^{2n} \exp(-\alpha R_{PQ}^2 t^2) \sim \frac{2}{\alpha^n R_{PQ}^{2n+1}} \qquad (4.5)$$

Thus, the worst case is for spherical Gaussians, where the integral decays as R_{PQ}^{-1}, i.e. in line with the Coulombic potential itself. This implies that ERIs will be nonvanishing over long distances, especially for very diffuse (small) exponents. This is why the direct approach remains costly for large systems despite prescreening the integrals, and also explains why it is harder to reduce the scaling of the Coulomb part of the Fock matrix compared to the exchange part, as will be described later. Nonetheless, for reasonably large systems the integral tensors should become highly sparse, with each fragment having an effectively fixed size domain over which integrals do not fall below a given threshold. In this way, it should be possible to make the evaluation and storage of the ERIs asymptotically linear scaling. One possibility would be to directly use compressed sparse matrix formats [124], such as the popular compressed row storage data structure, which can be very successfully used if only matrix multiplications and certain decompositions are needed. However, this is not compatible with Fock-building implementations that require access to individual integrals, as expensive binary searches need to be performed to locate the integral.

4.1.2 Fragment–blocked storage and evaluation

To solve this problem, I have devised a custom data structure and accompanying algorithm that makes full use of the fragmented structure of the system. Consider a slice of the three-index DF tensor **B** along the ABS axis, P. A typical sparsity pattern for this matrix is shown below:

where the blocks are fragment by fragment, i.e. $\mathbf{1} \equiv \mathbf{B}_{XX}^{K}$, $\mathbf{2} \equiv \mathbf{B}_{XY}^{K}$, etc. Note that the block indexing reflects the symmetry between the two OBS indices, such that at most $N_{Blocks} = N(N+1)/2$ blocks need be considered. The gradation represents the maximum integral value in that block: the diagonal, darker blocks contain mostly non-negligible integrals, while the lighter, off-diagonal blocks contain mostly near-zero integrals.

The new data structure comprises two arrays of size N_{Blocks}: one is an array of zero-sized matrices, the other a list of boolean values reflecting whether the corresponding block is zero. For the first row of blocks in the structure above, this would be as follows:

$$\text{Blocks} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & 3 & 4 & 5 & \ldots \end{bmatrix}$$
$$\text{Zeros} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & \ldots \end{bmatrix}$$

where one is true and zero is false. This approach can trivially be expanded to include the ABS blocks by 'z-marching' along the index P - this involves indexing as shown above within each P-slice, then incrementing the ABS index and repeating. In this way, only the non-vanishing blocks need be stored, but these can still be rapidly accessed individually due to the redundant, zero-sized matrices. This structure therefore does not reflect the true sparsity of the integrals as zero elements within each nonzero block are explicitly stored, unlike in sparse matrix structures where these would be discarded. However, as the number of significant blocks per fragment will eventually reach a fixed size, the storage and evaluation will still reach linear scaling. The data redundancy is then a small price to pay for the greatly increased speed and compatibility with later routines.

All that remains is to determine a suitable screening criterion for each block, $\mathbf{B}_{XY}^{Z}$. To do this, we use the fact that the electron repulsion integral in equation 4.3 is simply the inner-product induced on the Hilbert space by the Coulomb metric r_{12}^{-1}, allowing us to apply the Cauchy-Schwarz inequality [34]:

$$|V_{pq}| \leqslant |V_{pp}|^{1/2}|V_{qq}|^{1/2} \tag{4.6}$$

Specialising to the building of the Fock matrix, the initial density, $\mathbf{P}$, can also be used coupled with the expected distance scaling from equation 4.5. Taking the maximum values over each block then leads to

$$\left\| \mathbf{B}_{XY}^{Z} \mathbf{P}^{XY} \right\|_{\infty} \lesssim 2 \left\| \mathbf{g}_{XYXY} \right\|_{\infty}^{1/2} \left\| \mathbf{G}_{ZZ} \right\|_{\infty}^{1/2} \left\| \mathbf{P}^{XY} \right\|_{\infty} R_{(XY)Z}^{-1} \tag{4.7}$$

The initial density must be of reasonable quality, or screening may be inaccurate.

where $R_{(XY)Z}$ is the distance from fragment Z to the midpoint of fragments X and Y. The infinity norms for the 'diagonal' ERIs can very efficiently be computed once at the beginning of the calculation, as is done for conventional integral evaluation routines. The full $\mathbf{G}$ matrix needs to be computed regardless, as shown in Algorithm 4.1, such

that this screening has very little overhead. The fragment-blocked DF integral algorithm is given in Algorithm 4.2. Comparing with Algo-

---

Algorithm 4.2 Fragment-blocked density-fitting of ERIs

1: $\epsilon \leftarrow$ integral threshold
2: Blocks $\leftarrow$ array of integral blocks, $\mathbf{B}^Z_{XY}$
3: Zeros $\leftarrow$ array of booleans, all set to true
4: **for** each unique XY pair **do**
5:   Compute $\mathbf{g}_{XYXY}$
6:   $g_{XY} \leftarrow \|\mathbf{g}_{XYXY}\|^{1/2}_{\infty}$
7: **end for**
8: **for all** P, Q $\leqslant$ P in ABS **do**
9:   $G_{PQ} \leftarrow (P|Q)$
10: **end for**
11: Cholesky decompose $\mathbf{G} = \mathbf{LL}^\mathsf{T}$
12: Solve $\mathbf{LL}^{-1} = \mathbf{I}$ and store $\mathbf{L}^{-1}$
13: **for** each unique XYZ triple **do**
14:   $R_{(XY)Z} \leftarrow |(\mathbf{X} - \mathbf{Y})/2 - \mathbf{Z}|$
15:   **if** $2g_{XY}\|\mathbf{G}_{ZZ}\|^{1/2}_{\infty}\|\mathbf{P}^{XY}\|_{\infty} R^{-1}_{(XY)Z} > \epsilon$ **then**
16:     Zeros[INDEX(X, Y, Z)] $\leftarrow$ false
17:     Compute $\mathbf{B}^Z_{XY}$
18:     Blocks[INDEX(X, Y, Z)] $\leftarrow \mathbf{B}^Z_{XY}$
19:   **end if**
20: **end for**

---

rithm 4.1, there are a number of additional steps, but far fewer integral evaluations. Note that the Cholesky decomposition is still necessary, and eventually the cubic scaling of this will dominate the otherwise linear scaling algorithm. One slight difference is the need to store $\mathbf{L}^{-1}$ in line 12, instead of the symmetrising step in line 7 of the original algorithm. The reason for this will become clear later.

4.2 THE FOCK MATRIX

The ERIs and the density matrix at each iteration of the SCF procedure are used to form the Fock matrix, as given in equation 2.30. The core Hamiltonian involves only two indices and does not change throughout the procedure, so requires negligible computational effort. The expensive portions are the Coulomb and exchange matrices. From equation 2.30 it can be seen that in the conventional approach these would both entail quartic dependence on the number of basis functions. We therefore describe methods to reduce the cost for each of these in turn, starting with the more difficult exchange contribution.

4.2.1 The exchange matrix

Within the density-fitting framework described in algorithm 4.1, the exchange matrix can be rewritten as

$$K_{X\mu,Y\nu} = \sum_{W,Z} \langle \mu^X \tau^W | r_{12}^{-1} | \lambda^Z \nu^Y \rangle P^{W\tau,Z\lambda} \approx \sum_{W,Z} b^Q_{X\mu,W\tau} b^Q_{Z\lambda,Y\nu} P^{W\tau,Z\lambda}$$

From this, it immediately becomes apparent why this term is more expensive than the equivalent Coulomb one: the contraction with the density occurs across the two three-index tensors, seemingly resulting in an $\mathcal{O}(N^4 M)$ cost, which is a factor of M worse than the conventional counterpart. This can be improved by using the MO coefficients directly instead of the density, although this is complicated in the present instance by the nonorthogonality of the orbitals. We can get around this by decomposing the metric as $\sigma^{-1} = \mathbf{s}\mathbf{s}^\mathsf{T}$, and then half-transforming the coefficients so that equation 2.45 becomes

$$\mathbf{P} = \mathbf{T}\sigma^{-1}\mathbf{T}^\mathsf{T} = \mathbf{Ts}\left(\mathbf{Ts}\right)^\mathsf{T} = \tilde{\mathbf{T}}\tilde{\mathbf{T}}^\mathsf{T}$$

Note that the resulting orbitals are no longer strictly localised, but in practice are still well localised. Then, the density-fitted exchange can be written as

$$K_{X\mu,Y\nu} \approx \left(\sum_W b^Q_{X\mu,W\tau} \tilde{T}^{W\tau\cdot}_{\cdot i} \right) \left(\sum_Z b^Q_{Y\nu,Z\lambda} \tilde{T}^{Z\lambda}_{\cdot i} \right)^\mathsf{T} = \tilde{k}^{i,Q}_{X\mu\cdot} \tilde{k}^{Q,i}_{\cdot Y\nu} \qquad (4.8)$$

This factorises the construction into an AO to MO half-transformation step, followed by a contraction over the MO and auxiliary basis indices, both of which are $\mathcal{O}(ON^2 M)$. Thus the cost is slightly worse than cubic, but still substantially better than either quartic or quintic.

However, cubic scaling will still result in the exchange matrix becoming the bottleneck in the calculation, especially as it has to be performed once per iteration. Exchange is an inherently short-range effect [10], such that it seems sensible to try and use the inherent localisation of the ALMO orbitals to reduce the cost. A 'local exchange' approach was first proposed in the context of density-fitted calculations by Polly *et al.* [122] and has more recently been improved by Köppl and Werner [123]; we use their method as a starting point. The idea is that each localised orbital only has significant overlap with other orbitals within a fixed extent. Therefore, for each occupied molecular orbital, i, an orbital domain can be assigned, [i], comprising the atomic orbitals that will give non-vanishing contributions. As the system size increases, the domain sizes eventually become fixed (due to the localisation), such that the calculation eventually becomes linear scaling.

The DF framework requires the definition of three such domains per occupied MO, adding a substantial layer of complexity to the procedure. To see this, we split equation 4.8 into its two steps, and partition in terms of each i. The transformation step is

$$\tilde{k}_{X\mu\cdot}^{i,Q} = \sum_{W\tau \in [i]_{MO}} b_{X\mu,W\tau}^{Q} \tilde{T}_{\cdot i}^{W\tau\cdot} \, , \, X\mu \in [i]_{AO}, \, Q \in [i]_{ABS} \qquad (4.9)$$

From this, it can be seen that to achieve linear scaling, it is necessary to restrict the indices not only in the summation, but in the definition of the tensor itself. This results in separate domains for: the atomic orbitals associated with i in the tensor $\tilde{k}$, $[i]_{AO}$; the auxiliary functions in the same tensor, $[i]_{ABS}$; and the atomic orbitals with which the MO has significant overlap, implying non-negligible coefficients, $[i]_{MO}$. The definition of each of these will in general need to be different, as will be discussed shortly. The exchange matrix is subsequently constructed as

$$K_{X\mu,Y\nu} \approx \sum_{i} \sum_{Q \in [i]_{ABS}} \tilde{k}_{X\mu\cdot}^{i,Q} \tilde{k}_{\cdot Y\nu}^{Q,i} \, , \, X\mu, Y\nu \in [i]_{AO} \qquad (4.10)$$

In this way only a fixed-sized sum is performed per orbital, resulting in linear scaling.

Apart from the additional complexity noted above, a clear disadvantage of this approach is that the results so obtained are likely to be highly sensitive to how the domains are defined. This is where the fragment-localisation of the current method becomes particularly useful. Instead of considering the significance of each individual atomic orbital and auxiliary basis function with each MO, the contribution of a fragment as a whole to each MO can be used as the criterion. This will in general result in the inclusion of orbitals that are not necessary, and thus a somewhat slower approach to the linear regime, but the well-separated nature of the fragments precludes the exclusion of important contributions, resolving the strong dependence on whatever threshold is used. In the original local exchange methods, it has been found that while the errors in the density from the local approximation are small, it is necessary after self-consistency has been achieved to recalculate the Fock matrix once in full in order to determine an accurate energy [122, 123]. As will be shown later, the current approach removes this necessity, thus alleviating the need for an expensive quartic step in each calculation.

The atomic orbitals of a given fragment, X, are included in $[i]_{MO}$ if they make a significant contribution to the density on i. That is, defining

$$\rho_{X \to i} = \sum_{\mu \in X} \tilde{T}_{\cdot \mu}^{i\cdot} \tilde{T}_{\mu\cdot}^{\cdot i} \qquad (4.11)$$

we include $\{X\mu\} \subset [i]_{MO}$ if $\rho_{X \to i} > \epsilon_{MO}$. A reasonable threshold seems to be $\epsilon_{MO} \approx 10^{-6}$, determined by tests on small clusters, as described

in the next chapter. It is then presumed that, almost by definition, if the density of i has a significant contribution from $\{X\mu\}$, then this will also be true in the $\tilde{k}$-vector. Therefore $[i]_{MO} \subset [i]_{AO}$. However, as was noted previously, the Coulomb force decays slowly with separation, such that $[i]_{AO}$ will in general need to be considerably larger than $[i]_{MO}$, even for the short-range exchange term. To do this, it is necessary to define the extent of the orbital. This can be done by assigning spheres to each fragment, such that any orbital on that fragment will have negligible overlap with any function outside the sphere. This will clearly be controlled by the most diffuse basis function on each fragment, ϕ_μ. The maximum overlap with any given function is by definition that with itself, i.e. its norm. For a spherical Gaussian (which we take to be a reasonable approximation in all cases), this can be written as

$$\|\phi_\mu\| = 4\pi \left(\int_0^{r_\mu} + \int_{r_\mu}^\infty \right) dr \ r^2 \exp(-2\mu r^2) = N(r_\mu) + e(r_\mu)$$

The extent, and thus the sphere, is defined to be the radius r_X such that $e(r_\mu \geqslant r_X) < \epsilon_{AO}$. The overlap with any function located outside that extent will thus necessarily be less - and in general substantially less - than this threshold. Then, for i located on fragment Y, we include all the atomic orbitals from fragment X in $[i]_{AO}$ if $r_X < |Y - X|$. The above integration can easily be performed, yielding

$$e(r_\mu) = \left(\frac{\pi}{2\mu} \right)^{3/2} \left\{ \operatorname{erfc}\left(\sqrt{2\mu} r_\mu \right) + \frac{4\mu}{\sqrt{\pi}} \exp\left(-2\mu r_\mu^2 \right) \right\}$$

Assuming that $r_\mu \gg 1$ Bohr, which will in general be true for any basis with diffuse functions, the asymptotic expansion of the complementary error function, $\operatorname{erfc}(x) \approx \exp(-x^2)/(x\sqrt{\pi})$, leads to the above condition becoming

$$r_X \approx \left[\frac{\ln(\sqrt{2}\pi) - \ln(\mu\epsilon_{AO})}{2\mu} \right]^{1/2} \tag{4.12}$$

Again, a reasonable threshold is $\epsilon_{AO} \approx 10^{-6}$.

Finally, the auxiliary set domains are found by a population analysis, as suggested originally by Polly and coworkers [122]. For i on fragment Y, the auxiliary basis of fragment X is included in $[i]_{ABS}$ if the partial charge associated with charge transfer from i to X is greater than a threshold, ϵ_{ABS}. However, it was shown earlier that a Mulliken analysis, equation 2.52, would lead to no charge transfer and thus the auxiliary domain would consist of only the functions on that fragment. This would clearly lead to large errors in any system

Algorithm 4.3 The selection of domains for each occupied molecular orbital in the local exchange procedure.

1: Determine extent, r_Y, of each fragment Y (equation 4.12)
2: Compute $\mathbf{S}^{1/2}\tilde{\mathbf{T}}$
3: **for** each occupied molecular orbital Yi **do**
4:     Add all AOs on fragment X to $[i]_{MO}$ and $[i]_{ABS}$
5:     **for** each fragment $X \neq Y$ **do**
6:         **if** $\rho_{X \to i} > \epsilon_{MO}$ (equation 4.11) **then**
7:             Add all AOs on fragment X to $[i]_{MO}$
8:         **end if**
9:         **if** $q_{i \to X} > \epsilon_{ABS}$ (equation 4.13) **then**
10:             Add all auxiliary functions on fragment Y to $[i]_{ABS}$
11:         **end if**
12:     **end for**
13:     $[i]_{AO} = [i]_{MO}$
14:     **for** each fragment X not included in $[i]_{MO}$ **do**
15:         **if** $|\mathbf{Y} - \mathbf{X}| < r_Y$ **then**
16:             Add all AOs on fragment X to $[i]_{AO}$
17:         **end if**
18:     **end for**
19:     Form the inverse fitting metric for $[i]_{ABS}$
20: **end for**

---

containing significant delocalisation. Therefore, it is necessary to use a Löwdin analysis instead. Defining

$$q_{i \to X} = \sum_{\mu \in X} \left[\mathbf{S}^{1/2}\tilde{\mathbf{T}} \right]_{\mu i}^2 \tag{4.13}$$

the condition becomes $q_{i \to X} > \epsilon_{ABS}$. In the original local exchange method [122, 123], this domain would be further extended by including either all functions within a distance cutoff, or from all atoms connected to the atom of the current orbital. This is unnecessary here, however, as the localisation is on a fragment-by-fragment rather than atom-by-atom basis, and the fragments are by assumption not covalently bound.

The domain fitting process is summarised in algorithm 4.3. In general, the domains should need to change at each iteration as the quantities in equations 4.11 and 4.13 will change, further complicating the procedure. However, given a high quality initial guess to the density, as is obtained by using the converged fragment coefficients as outlined in algorithm 2.1, it is sufficient to select the domains in the first iteration and keep them fixed throughout. One additional complication, however, is that the restriction of the fitting basis to $[i]_{ABS}$ for each occupied orbital means that the fitting metric, $\mathbf{G}$, from algorithm 4.1 is no longer a valid metric in the restricted space. As

such, the relevant blocks from $[i]_{ABS}$ need to be formed into a smaller matrix, $\mathbf{G}^i$, which is then Cholesky decomposed in the usual way as $\mathbf{G}^i = \mathbf{L}^i \left(\mathbf{L}^i\right)^T$. The integrals used in the construction of the local exchange matrix are thus not $b_{\mu\nu}^Q$, but rather equation 4.9 is rewritten as

$$k_{X\mu\cdot}^{i,Q} = \sum_{W\tau\in[i]_{MO}} (\mu^X\tau^W|Q)\tilde{T}_{\cdot i}^{W\tau\cdot} \, , \, X\mu \in [i]_{AO}, \, Q \in [i]_{ABS} \qquad (4.14)$$

Then, an intermediate step between the transformation and the contraction is added, where the metric is included:

$$\tilde{\mathbf{k}}^i = \mathbf{k}^i \left[\mathbf{L}^i\right]^{-1} \qquad (4.15)$$

4.2.2 The Coulomb matrix and the blocked ERIs

The density-fitted Coulomb matrix is in principal much simpler than the exchange matrix. It is given by

$$J_{\mu\nu} = b_{\mu\nu}^Q \left[\mathbf{b}^T\mathbf{P}\right]^Q \qquad (4.16)$$

That is, it can be formed by two $\mathcal{O}(N^2M)$ matrix-matrix products, if the $\mu\nu$ indices are compressed into a single matrix index. This is clearly far less expensive than the conventional exchange contribution, although it is still cubic. It is thus tempting to try and apply the local approximation here also, but as has been mentioned earlier, the Coulomb term is not a short-range effect. This means that the orbital domains that would be necessary to avoid substantial errors would be prohibitively large, such that no savings in cost would be seen even for large systems. The second issue is that, if the local exchange method is being used, then it is necessary to store the raw three-centre ERIs rather than those symmetrised with the metric; certainly, it is not desirable to store both sets, given the cubic memory requirements. This itself raises the problem of making efficient use of the fragment-blocked ERIs in the Fock building procedure, as this precludes the simple matrix multiplications in equation 4.16. Algorithm 4.4 outlines how the majority of these problems can be alleviated; it assumes that the ERIs have been computed via algorithm 4.2, and that the inverse metric $\mathbf{G}^{-1}$ is available.

The integral blocking therefore not only reduces the memory imprint, but also the computational cost of the Fock build, as large sections of the inner loops can be neglected in both the exchange and Coulomb parts of the algorithm. In particular, the first and third steps for the Coulomb matrix asymptotically scale linearly with system size as the number of nonzero ERI blocks per fragment becomes constant.

Algorithm 4.4 Fock build using blocked DF-ERIs and the local exchange approximation

---

1: $\mathbf{F} \leftarrow \mathbf{h}$
2: **for** each occupied MO, Yi **do**
3:      $\mathbf{k}^i \leftarrow \dim([i]_{AO}) \times \dim([i]_{ABS})$
4:      **for all** $ZQ \in [i]_{ABS}$, $X\mu \in [i]_{AO}$ **do**
5:          **if** ERI block XYZ is not zero **then**
6:              Add contribution from $X\mu, ZQ$ to $\mathbf{k}$ (equation 4.14)
7:          **end if**
8:      **end for**
9:      Compute $\tilde{\mathbf{k}}^i$ by equation 4.15
10:      Add $\tilde{\mathbf{k}}^i \left[\tilde{\mathbf{k}}^i \right]^T$ to $\mathbf{K}$
11: **end for**
12: **for all** nonzero ERI blocks, XYZ **do**
13:      $\mathbf{B}_Z \leftarrow \mathbf{B}_Z + \mathbf{P}_{XY}(XY|Z)$
14: **end for**
15: $\tilde{\mathbf{B}} \leftarrow \mathbf{G}^{-1}\mathbf{B}$
16: **for all** nonzero ERI blocks, XYZ **do**
17:      Add $(XY|Z)\tilde{\mathbf{B}}_Z$ to $\mathbf{F}$
18: **end for**

This implies that the dominating step becomes the intermediate multiplication by $\mathbf{G}^{-1}$, which is formally $\mathcal{O}(M^2)$, albeit with a very small prefactor. This step is likely to be very rapid despite the quadratic dependence. While $\mathbf{G}$ will become sparse in the macroscopic limit, this in general does not imply that its inverse will be sparse - in fact, a sparse matrix very often has a dense inverse [124]. Therefore it is not clear how the expense of this step could be reduced if it were seen to eventually become the bottleneck.

4.3 PAIRWISE RPAXD

The calculation of the RPA dispersion energy requires integrals transformed to the molecular orbital basis, which can be achieved using the density-fitting described in algorithm 4.1. The ERIs are then given by equation 4.2. Specifically, for dRPA as described in equation 3.54, we have that $\mathbf{K} = \tilde{\mathbf{b}}\tilde{\mathbf{b}}^T$. This means that, as with the Fock build above, we can avoid ever needing to form the full four-index integral tensor. Instead, we expand the amplitude equations as follows:

$$0 = \mathbf{K} + \mathbf{A}\mathbf{t} + \mathbf{t}\mathbf{A} + \mathbf{t}\mathbf{K}\mathbf{t} = \epsilon\mathbf{t} + \mathbf{t}\epsilon + (1+\mathbf{t})\mathbf{K}(1+\mathbf{t})$$

Therefore, writing $\mathbf{u} = (1+\mathbf{t})\tilde{\mathbf{b}}$ results in the much simpler

$$\epsilon\mathbf{t} + \mathbf{t}\epsilon = -\mathbf{u}\mathbf{u}^T$$

However, in the context of RPAx (equation 3.62), this is not possible due to the exchange terms. Instead, we must directly form the tensor **B** as

$$B_{iajb} = \tilde{b}_{ia}^P \tilde{b}_{jb}^P - \tilde{b}_{ja}^P \tilde{b}_{ib}^P \tag{4.17}$$

Note that in the spin-restricted case, the first term (the direct integral) gains a factor of two.

Having to store the full tensor introduces a quartic memory requirement, although in practice it is essentially quadratic as it is $\mathcal{O}(O^2V^2)$. It is not possible to use the blocking technique described in algorithm 4.2, as the virtuals must be projected out of the occupied subspace before the integral transformation step, which will alter the localisation. In addition, the requirements on the auxiliary basis to minimise the density-fitting errors are completely different for the integral transformation as compared to the Fock build, meaning that a separate ABS must be used. This both increases the program complexity and means that a second set of integrals must be computed. The advantage of this, however, is that it means there is no change in cost if range separation is desired - we simply follow the same algorithm, but with the two- and three-center integrals over the full Coulomb operator replaced by integrals over the operator in equation 3.69. Thus incorporating range separation is as simple as having a flag to switch the potential used, noting that the algorithms to compute the different integrals are essentially identical.

As noted earlier, the amplitude equations are Riccati equations, which can be solved analogously to the charge transfer equations. That is, we use the fact that ϵ is almost diagonal, such that the diagonal terms are dominant, guaranteeing convergence. For the direct RPA equations, this becomes

In the canonical orbitals, ϵ would be diagonal, making the result simpler to find.

$$\left(\epsilon_{ii}^{aa} + \epsilon_{jj}^{bb}\right)[t^{(n+1)}]_{ij}^{ab} = -\sum_{k\neq i, c\neq a} \epsilon_{ik}^{ac}[t^{(n)}]_{kj}^{cb} - \sum_{k\neq j, c\neq b}[t^{(n)}]_{ik}^{ac}\epsilon_{kj}^{cb}$$
$$- \left[\mathbf{u}^{(n)}(\mathbf{u}^T)^{(n)}\right]_{ij}^{ab} \tag{4.18}$$

Similarly, in RPAx the tensor **A** is diagonally dominant, such that if we define **A**$'$ to be **A** with all diagonal elements set to zero, and let $\mathbf{W}^{(n)} = \mathbf{B}t^{(n)}$, the amplitude equation becomes

$$\left(A_{ii}^{aa} + A_{jj}^{bb}\right)[t^{(n+1)}]_{ij}^{ab} = -\left[\mathbf{B} + \mathbf{A}'\mathbf{t}^{(n)} + \mathbf{t}^{(n)}\mathbf{A}' - \mathbf{t}^{(n)}\mathbf{W}^{(n)}\right]_{ij}^{ab} \tag{4.19}$$

The most expensive computation in each is the final, quadratic term. For dRPA, this is $\mathcal{O}(O^2V^2M)$, whereas for RPAx it is $\mathcal{O}(O^3V^3)$; the reduction to quintic scaling in the former is a result of never having to explicitly form the four-index integral tensor.

As several iterations will be necessary to achieve reasonable levels of convergence, this sixth-order dependence is prohibitive for anything except reasonably small systems, and will rapidly dominate

---

Algorithm 4.5 Pairwise RPAxd procedure

1: $\rho \leftarrow$ distance cutoff
2: $\epsilon \leftarrow$ amplitude convergence threshold
3: **for** each unique fragment pair XY **do**
4:     **if** $|X - Y| \leqslant \rho$ **then**
5:         Form σ^{XY}, $\mathbf{P}_{XY}$ from $\mathbf{T}_{XY}$ using equations 2.42, 2.45
6:         Project virtuals, $\tilde{\mathbf{V}} \leftarrow (\mathbf{I} - \mathbf{S}_{XY}\mathbf{P}_{XY})\mathbf{V}_{XY}$
7:         Compute and transform integrals, using algorithm 4.1
8:         **if** dRPA or SOSEX **then**
9:             $(t^{(0)})_{ij}^{ab} \leftarrow (\epsilon_{ii}^{aa} + \epsilon_{jj}^{bb})^{-1} K_{iajb}$
10:         **else**RPAx
11:             $(t^{(0)})_{ij}^{ab} \leftarrow (A_{ii}^{aa} + A_{jj}^{bb})^{-1} B_{iajb}$
12:         **end if**
13:         **while** $\left\| t^{(n+1)} - t^{(n)} \right\| > \epsilon$ **do**
14:             Update amplitudes using equation 4.18 or 4.19
15:         **end while**
16:         Calculate energies using equations 3.70 and 3.71
17:     **end if**
18: **end for**

---

any calculation. The cost can be reduced to linear scaling, however, by making use of the decomposition described in equations 3.70 and 3.71, and figure 3.4. From these it can be seen that within the framework of double excitations, to which rCCD (and therefore RPA) is by definition restricted, only dispersive terms between pairs of fragments can be computed. This does not mean that only two-body contributions are included, as RPA implicitly includes contributions to infinite order at the orbital level. At the fragment level, however, the energies are essentially pairwise. When coupled with the fact that the orbitals are absolutely localised, a local correlation approach is natural, where the orbital domains are restricted. This is logically achieved by performing separate RPA calculations for each pair of fragments. If the ALMO orbitals were used directly, this would be exact, as the integral transformation could essentially be done fragment by fragment. However, the requirement for strong orthogonality between the occupied and virtual subspaces - ensured by projecting the virtual orbitals out of the occupied subspace - means that the pairwise approach is an approximation. On the other hand, by projecting out of a smaller occupied space (restricted to the pair of fragments under consideration), the localisation of the virtual orbitals is better maintained. This means that the decomposition in figure 3.4 is more accurate, as there is less overlap between terms.

Moreover, as dispersion goes asymptotically as R^{-6} where R is the separation between fragments, the procedure can be made to be linear scaling by having a distance cutoff beyond which pairs are not consid-

ered. The appropriate value for such a cutoff will in general depend on the system, and in particular, on the degree of delocalisation in the system. In addition, there is no fixed definition of distance between two molecules. A simple choice would be centre-of-mass separations, but these do not take into account the extent of a molecule. E.g. two water molecules whose centres are 5 Å apart will only be very weakly interacting if at all, but a water and an adenine residue at 5 Å may have points of contact much closer together. Therefore, we approximate the molecules as spheres with a radius defined by half the maximum distance between atoms in the molecule. The separation is then measured as the centre-of-mass distance, minus the radii of the two fragments. Of course, this does not account for any anisotropy in the molecules, but gives a good coarse indicator of when an interaction will be negligible; tests in the next chapter suggest that a cutoff of 15 Bohr leads to only excluding negligible dispersion terms. The resulting 'pairwise RPAxd' procedure, where the 'd' denotes restriction to the dispersion and exchange-dispersion excitations, is given in algorithm 4.5

4.4 PARALLELISATION

While the consideration thus far has been generating an algorithm that scales linearly with the number of fragments, this has occasionally been at the expense of the scaling prefactor; that is, while twice as many fragments takes roughly twice as much resource, this is not much use if a pair of fragments is prohibitively expensive. This is particularly problematic when large fragments are used, especially when using RPAx, which could scale as poorly as $\mathcal{O}(n^5)$ within the fragments. In this section, I detail how extensive use of parallel computing can alleviate these problems in all but memory cost.

At all stages of the calculation, the per-fragment or fragment-pair steps are independent except through the total electronic density. As such, the method is particularly suited to large-scale distributed memory parallelism, with fragments farmed to separate processes. The individual calculations, requiring pooled memory, can then be treated with shared memory - multithreaded - parallelism. Further to this, the extensive linear algebra components can be accelerated using graphical processing units (GPUs). Each form of parallelisation requires separate technical considerations, which will be discussed in detail here. The combined scheme, shown in Figure 4.1, allows for much larger systems to be considered on a reasonable timescale, involving many thousands of molecules, a feat which has not been achieved before for a fully quantum, *ab initio* method.
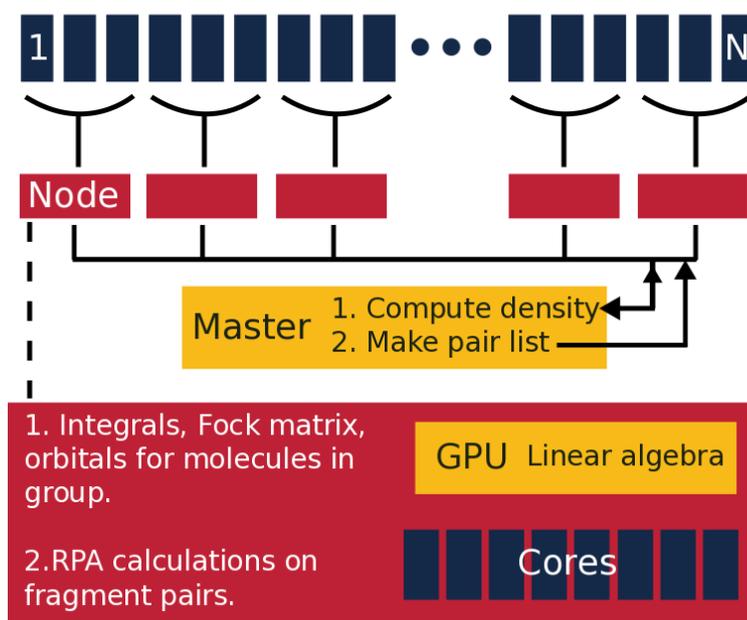
Figure 4.1: A combined scheme for the parallelisation of the ALMO+RPA method, using shared- and distributed-memory parallelisation, in tandem with GPU accelerators.

4.4.1 Multithreading

As shown in Figure 4.1, the primary idea is to perform the collection step (the construction of the total density) on the master node, then farm out the individual fragment calculations to separate nodes. These separate calculations can then be multithreaded, minimising the need for data replication: only the total density and the data for the fragments on any given node need to be stored on that node. The fine-level multithreading could then be treated in two ways. The simplest approach is to make use of existing, highly-optimised linear algebra packages to thread the matrix multiplications and diagonalisations [293]. This does not, however, make any use of the underlying structure of the algorithm, and would compete with the use of GPUs, which are themselves attuned to such tasks [294–296]. Instead, we adapt the algorithms presented in previous sections to maximise the parallelism of the overall routine.

Two-electron integrals

For the two-electron repulsion integrals (ERIs), we have the added complexity of there being three distinct kinds: two-, three-, and four-index integrals. The threading for each follows a similar pattern, however, in that we balance over the blocks of the integral matrices. This is the division of labour with the lowest overhead; that is, the additional cost associated with threading is minimised. It is not the best way

The one-electron integrals can be threaded in exactly the same manner, although their cost is negligible in comparison.

to balance the parallelisation load, though, as integrals over functions with high angular momenta can take considerably longer than those with low angular momentum. In general, there are considerably more of the latter, which is why we have chosen to balance in this way.

The general scheme is as follows: the basis sets are divided into contracted angular momentum shells, and in the inner loop the shell doublet, triplet, or quartet is computed depending on the thread ID. This acts as a heuristic way to balance the load, as it avoids all the shells of a particular angular momentum being placed on one thread. The scheme is summarised in Algorithm 4.6.

Algorithm 4.6 The thread load balancing algorithm for the two-electron integrals involving two, three, or four indices.

1: For each basis set, divide into angular momentum shells, s^m
2: Create an integral engine and initialise counter for each thread
3: **for** each shell, $s^m_{i_m}$, in the basis on index $m = 1, \ldots, n$ **do**
4:     **if** counter % $N_{threads} =$ thread ID **then**
5:         Compute $\left(s^1_{i_1} \cdots s^n_{i_n} \right)$
6:     **end if**
7:     Increment the counter
8: **end for**

An alternative specific to the fragment-blocked integral method of Algorithm 4.2, would be to parallelise directly by the blocking, chunking the integral domains up in to threads. The shell data for *every* fragment is needed on every node in either scheme, so this requires minimal extra data transfer. However, there are two problems: it ignores the un-blocked algorithms, such as the two-centre integrals and if density fitting is not used; and it requires communication between nodes to ensure that there is no duplicated effort in calculating integrals between pairs of domains. The above approach instead parallelises at the lowest level, and is agnostic to the type of two-electron integral required.

Fock build

The Fock build is the most resource-intensive part of the entire calculation, and presents unique challenges to parallelise. This is further complicated by savings made possible in the local Fock build described earlier. As such, we present two separate multithreaded routines, depending on whether the full or local exchange is required.

The full exchange algorithm is given in Algorithm 4.7, where equation 4.8 is split into two half transformations. The first transformation must be completely collected before the second, which forms the main bottleneck in the parallelisation; thus we have opted for the simplest and most efficient scheme, whereby the first atomic orbital index is

---

Algorithm 4.7 Multithreaded, density-fitted full exchange build.

1: Divide number of orbital basis functions on the first index of ERIs into n_{threads} roughly equal chunks, $\{A_i\}$
2: **for** each thread i **do**
3:     **for** first ERI index, μ, in $\{A_i\}$ **do**
4:         Half-transform 3-index ERIs as $(\mu j|K) = (\mu \nu|K)T_j^\nu$.
5:     **end for**
6: **end for**
7: Collect all half-transformed integrals, $(\mu j|K)$.
8: **for** each thread i **do**
9:     **for** first ERI index, μ, in $\{A_i\}$ **do**
10:         Form full exchange contributions as $k_{\mu\nu} = (\mu j|K)(\nu j|K)$
11:     **end for**
12: **end for**

---

chunked into roughly equal sizes. This strategy can then be used in both halves. There is no particular problems with load balancing, as we are essentially using a natural division of two matrix-matrix multiplications. However, by doing it simply, rather than relying on existing, complex threading algorithms for such tensor contractions, we are allowing the use of further acceleration in the raw multiplications using GPUs later. The Coulomb portion of the Fock build is not shown above as it is a simple, single step at the end, where the same threading scheme can be used.

In the local Fock build algorithm, given in Algorithm 4.8 as a direct adaptation of Algorithm 4.4, the complexity switches from the exchange portion to the Coulomb portion. The former can be directly balanced over the occupied molecular orbitals, as each of these dictates the orbital domains involved in the computation. This requires shared memory access to all of the three-index integrals on every node, which could in principle be a problematic communication issue; however, in reality, the fragment blocking minimises this problem. The Coulomb matrix, on the other hand, requires splitting into two distinct steps where the half-transformed **B** matrix needs to be collected before the second transformation can take place. This is only a one-index quantity, however, of dimension N_{ABS}, which again minimises storage and communication requirements. An alternative would be to thread at the level of the tensor contractions, but this would require the storage and communication of multiple three-index quantities, so is not desirable when used in tandem with distributed parallelism.

Algorithm 4.8 Multithreaded, density-fitted Fock build with local exchange.

1: $\mathbf{F} \leftarrow \mathbf{h}$
2: **for** each occupied MO, Yi **do**
3:     **if** $i \% n_{\text{threads}} = $ thread ID **then**
4:         $\mathbf{k}^i \leftarrow \dim([i]_{\text{AO}}) \times \dim([i]_{\text{ABS}})$
5:         **for all** $ZQ \in [i]_{\text{ABS}}, X\mu \in [i]_{\text{AO}}$ **do**
6:             **if** ERI block XYZ is not zero **then**
7:                 Add contribution from $X\mu, ZQ$ to $\mathbf{k}$ (equation 4.14)
8:             **end if**
9:         **end for**
10:         Compute $\tilde{\mathbf{k}}^i$ by equation 4.15
11:         Add $\tilde{\mathbf{k}}^i \left[\tilde{\mathbf{k}}^i \right]^{\mathsf{T}}$ to $\mathbf{K}$
12:     **end if**
13: **end for**
14: Chunk fragments on node into n_{threads} blocks
15: **for** each fragment, f_1, in current thread's chunk **do**
16:     **for** each fragment $f_2 \leqslant f_1$ **do**
17:         **for all** fragments f_k **do**
18:             **if** xyK block $(f_1 f_2 | f_k)$ is non-zero **then**
19:                 $\mathbf{B}_{f_k} \leftarrow \mathbf{B}_{f_k} + \mathbf{P}_{f_1 f_2}(f_1 f_2 | f_k)$
20:             **end if**
21:         **end for**
22:     **end for**
23: **end for**
24: $\tilde{\mathbf{B}} \leftarrow \mathbf{G}^{-1}\mathbf{B}$
25: **for** each fragment, f_1, in current thread's chunk **do**
26:     **for** each fragment $f_2 \leqslant f_1$ **do**
27:         **for all** fragments f_k **do**
28:             **if** xyK block $(f_1 f_2 | f_k)$ is non-zero **then**
29:                 Add $(f_1 f_2 | f_k)\tilde{\mathbf{B}}_{f_k}$ to $\mathbf{F}$
30:             **end if**
31:         **end for**
32:     **end for**
33: **end for**

RPA

The pairwise RPA can primarily be parallelised by farming pairs of fragments to separate nodes. Within each separate calculation, however, there are two resource-intensive portions: the transformation of the AO integrals to the MO basis, and the the iterative calculation of the amplitudes. The latter involves multiple tensor contractions that are best handled by GPUs.

Algorithm 4.9 Multithreaded AO to MO density-fitted integral transformation

1: **for** each ABS index P % n_{threads} = thread ID **do**
2:     **for** each occupied orbital, i **do**
3:         **for** each OBS function pair μ, ν **do**
4:             $(i\nu|P) \leftarrow (i\nu|P) + T_i^\mu(\mu\nu|P)$.
5:         **end for**
6:         **for** each virtual orbital, a, and OBS function ν **do**
7:             $(ia|P) \leftarrow (ia|P) + V_a^\nu(i\nu|P)$
8:         **end for**
9:     **end for**
10: **end for**
11: Form $b_{iaQ} = (ia|P) \left[\mathbf{G}^{-1/2} \right]_{PQ}$

The density-fitted integral transformation involves two half transformations, as described in Algorithm 4.1. Consideration of the equations shows that these two transformations can be treated together by threading over the ABS axis of the tensors, which also happens to be the largest axis. In this way we arrive at Algorithm 4.9. After this is completed, we need to form the excitation matrices, **A**, **B**, or **K**, depending on which flavour of RPA is required. This can easily be coarsely multithreaded by dividing the index quartets $(iajb)$ into n_{threads} chunks, as there is no codependency between terms.

It is in principle possible to divide the computation of the amplitudes into multiple tensor contractions, which could then be handled by separate threads. However, there are serious memory considerations in this approach, as each four-index tensor would require $\mathcal{O}(o^2v^2)$ storage. For a large fragment this could become a serious issue. The alternative is to have each thread accessing the same tensor storage, but this introduces a race condition. Adding mutex blocks to avoid any races slows down the parallelisation inexorably unless the individual contractions are themselves large. Thus, the best way to multithread the amplitude iterations would be to have a switch where, for small molecules, separate storage is used for each step and then brought together at the end, while for large systems, a single storage source is accessed with mutex conditions to avoid a race. If there is a GPU available, however, each separate thread would then be

Lack of co-dependency implies embarrassingly parallel, as each computation requires no information about other computations.

competing for that resource to compute the contraction, such that the multithreading is useless. Overall, this section is the most complex situation to multithread, and requires considerably more consideration in future.

4.4.2 Distributed processing

The distributed parallelisation, using the message passing interface (MPI), is by comparison relatively simple, due to the fragmented nature of the method. The ALMO+RPA method as a whole follows these steps:

1. Perform Hartree-Fock calculation on each fragment;

2. Loop until convergence achieved:

    2.1. Form the total density, $\mathbf{P}$,

    2.2. Build Fock matrix,

    2.3. Form the fragment-localised Fock matrices,

    2.4. Diagonalise the fragment-localised Fock matrices;

3. Compute perturbative correction;

4. Compute RPA dispersion for each pair of fragments.

The first step is distributed to separate processes by dividing the total list of fragments into roughly equal-sized chunks and sending each chunk to one node. The HF calculations are then themselves multithreaded, in the manner described above for the Fock build and integral routines. Typically, for a node with c cores, we attribute two threads per core, such that if n_{threads} are desired per process, we can run

$$n_{\text{processes}} = \text{floor} \left(2c/n_{\text{threads}} \right)$$

fragment calculations at once. Ideally, each node would have sufficient memory to deal with such a load, but in reality estimates would need to be made as to the memory requirements so as to avoid competition between processes. Another concern is that some fragments will be larger and thus take proportionately longer than others; a perfect load-balance would take this into account when chunking up the fragments, so that one process is not left hanging after all the others are complete. Unfortunately, the predicted cost of a calculation depends heavily on the architecture of the system and the exact method (e.g. density-fitted or not) being used. One simple solution is to order the list of fragments by number of electrons, then devise the chunks to have similar total numbers of electrons. This is essentially a bin packing problem [297], for which no closed solution exists (it

Number of electrons is a much better measure of system size than number of atoms, as it dictates the number of occupied orbitals.

is NP-complete), but good heuristic algorithms exist, such as the LPT algorithm [298, 299].

The only step in the above that cannot reliably be distributed is the inherently shared-memory one: the formation of the density. However, this comprises three matrix multiplications and a Cholesky decomposition (to form the inverse metric), which will only become the bottleneck in the asymptotic limit; moreover, these linear algebra routines are readily parallelised using either multithreading or GPUs. The more important point is that this shared memory needs to be communicated to all processes in the subsequent calculation. An order N^2 communication could potentially be time consuming, but only needs to happen once in each iteration.

For the Fock build and fragment diagonalisations, the same fragment-based distribution of processes can be used. The latter are completely independent of one another, but the Fock build needs to be collected at the end from all processes. This is a further order N^2 communication. At this point, the use of a global framework, such as GlobalArrays Toolkit [300], would be desirable. However, this is not yet compatible with our choice of multithreading (C++11 threads) and GPU distribution (CUDA [301]), the advantage from which is much greater than the added communication cost, at least on any of the systems we have thus far considered. In future, this should be addressed, as on truly huge systems (of the order of tens of thousands of atoms), the interconnect used, and thus the communication rate, will have an extreme effect on the efficiency of the parallelism.

The perturbative correction again comprises only a few order $OV \sim N$ matrix multiplications, and so the overhead involved with distributing it would be greater than the speedups from doing so. Instead, we compute this on the master process, relying on the efficiency of shared-memory parallelisation of the linear algebra. Finally, for the pairwise RPA, each pair calculation is completely independent of each other pair. As such, on the master process, we determine a list of all pairs, then prune this list based on the distance cutoff. Of the remaining pairs, we chunk the list into roughly equal sizes - again, this can be done either coarsely by number of pairs, or with more precision by total number of electrons - and distribute each chunk to a single node. On that node, we can run $n_{processes}$ pairs simultaneously. The energies from each calculation can then be sent back to the master process as each completes, at essentially zero cost.

4.4.3 GPU acceleration

The explosion in popularity of computer gaming over the last few decades has resulted in the rapid advancement of technologies capable of rapidly rendering high-quality graphics [302]. At their heart,

$$R_{iajb} = K_{iajb} + A_{iakc}\sigma^{kl}\pi^{cd}t_{lj}^{db} + t_{ik}^{ac}\pi^{cd}\sigma^{kl}A_{ldjb}$$
$$+ t_{ik}^{ac}\pi^{cd}\sigma^{kl}K_{ldme}\sigma^{mn}\pi^{ef}t_{nj}^{fb} = 0$$

Multiplicity

$$\left[Z^1\right]_{il}^{ad} = \left(\epsilon_{ik}^{ac} + K_{iakc}\right)\sigma^{kl}\pi^{cd} \qquad\qquad 2$$

$$\left[Z^2\right]_{kj}^{cb} = \pi^{cd}\sigma^{kl}\left(\epsilon_{lj}^{db} + K_{ldjb}\right) \qquad\qquad 2$$

$$Y_{iajb} = Z_{il}^{ad}t_{lj}^{db} \qquad\qquad 3$$

$$\left[R^4\right]_{iajb} = Y_{iald}\sigma^{mn}\pi^{ef}t_{nj}^{fb} \qquad\qquad 1$$

Figure 4.2: A decomposition of equation 3.56 into the different tensor contractions, ranked by the multiplicity (the number of appearances) of that contraction in the equation.

such renderers rely on ultrafast matrix-matrix multiplications and other common linear algebra operations [303]. The scientific community quickly realised that these new technologies were therefore ideally suited for accelerating many of the most computationally intensive parts of scientific codes [304, 305]. Nowadays, graphical processing units are routinely used in applications ranging from large-scale molecular dynamics simulations [306–309], to the numerical solution of partial differential equations [310–312].

A GPU generally comprises several thousand small processing cores highly optimised for specific floating point operations [303]. There is limited on-board random access memory directly connected to these densely packed cores; this memory and processing is entirely separate from that of the motherboard, or central processing units. At a granular level, a matrix multiplication can be divided into a chain of several floating point multiplications and additions, which can then be collected into the local memory to form the result. Further, all tensor contractions can be partitioned in the same manner, greatly increasing the applicability of GPUs [313, 314]. This does however introduce many unique complications in comparison to traditional programming: the processor instruction set and architecture is entirely different to that of the central processing; the memory is limited and thus needs rapid and repeated communication with the central memory; and the efficiency of the GPU cores is due to fine-tuning towards specific data types, particularly single-precision floats.

The first of these problems is addressed by the high-level libraries provided by the GPU developers, in particular CUDA for NVIDIA

units [301], and OpenCL more generally [315]. These can then be wrapped in an architecture-agnostic manner into linear algebra packages. In this work, we focus on the Eigen library [316], which forms a core part of Google's TensorFlow environment for the parallelisation of tensor operations [317]. This only has limited support for OpenCL, which unfortunately limits the usefulness to mainly NVIDIA GPUs. The advantage is that the individual operations we require - primarily tensor contractions - have already been optimised, and we do not need to consider the finer details of instruction sets and assigning tasks to individual cores on the GPU itself.

Our attentions instead lie on minimising the need for communication between the GPU and central memory, and on maintaining precision. The latter is difficult, in the sense that the accuracy of quantum-chemical methods demands double precision, which is generally only available on expensive, server-class GPUs. However, recent research has demonstrated that there is no loss of accuracy in using single-precision for the majority of the iterations in a coupled cluster routine, switching to double precision only at the very end [318–320]. This suggests that we can still envisage considerable speed-ups from the use of general purpose (GP) GPUs so long as the final steps are performed by conventional means when the double precision units are not available.

Any order n^2 matrix multiplication (n is the number of basis functions on a single fragment) can easily be transferred fully onto the GPU, so long as the individual fragments are not overly large. The exact definition of 'large' of course depends on the basis set used, and the GPU itself. On the assumption that most GP-GPUs have at least 8 Gb of memory, however, and assuming single precision floats (32 bits), we can estimate that the limiting size is roughly 93,000 basis functions. Even for a large basis, e.g. aug-cc-pVQZ, this equates to 1,700 carbon atoms, which lies well beyond the type of fragments this method is designed to consider. Therefore, we need only focus on consideration of the order n^4 quantities, specifically the tensors associated with the pairwise RPA amplitude equations.

An analysis of equation 3.56 is shown in Figure 4.2. From this, it becomes clear that the main speedup is to be gained from using the GPU for the contraction of the effective integrals, Z_{il}^{ad}, with the amplitudes, t_{lj}^{db}, as this is used three times throughout the calculation. It might appear that the fourth term, in purple, would benefit more as it is ostensibly order o^4v^4, but in reality it is two separate order o^3v^3 contractions that only appear once each.

There is an added complexity, however, in that whichever of the multiplicity two contractions goes on to be used in the multiplicity three step could itself benefit, as it must be done first. Certainly in a situation where there is one GPU per fragment pair in the RPA calcu-

lation, it would be sensible to compute this on the GPU, retaining the memory for the subsequent contraction with the amplitudes. In this case, the multiplicity one step could then also be performed on the GPU, as no further memory communication would be required. In the case where multiple processes are sharing a GPU, there is greater benefit in allowing each process to utilise the GPU for the single multiplicity three step. The first process to reach that point copies data to the GPU and performs the contraction; meanwhile, if a second process reaches the same point, it can divert to the next contraction while waiting for the GPU, then a memory swap can occur. In an ideal system where all fragments are of the same size, this will result in perfect resource sharing. Of course, this will not be the case in most instances, and will introduce a bottleneck. There is no perfect solution to this problem, and the approach here is to ensure that the GPU is being used maximally, i.e. to prefer that some process waiting occurs so long as the GPU is always in use. The reason for this choice is that the many thousands of cores on the GPU naturally offer a greater speedup than any multithreading could provide, such that the waiting time is deemed negligible.

The handling of the data transfer and contraction is performed by the Eigen library internally. Unfortunately, this library does not currently provide robust non-GPU tensor support, which creates a programmatic problem. Purpose-built tensor contraction libraries, such as the Cyclops Tensor Framework [321], significantly outperform Eigen on traditional CPU-based systems. Therefore, the presence of a GPU must be determined at *compile-time*, rather than runtime, so that the appropriate library can be chosen; this means that certain portions of the calculation not specialised to the GPU can actually suffer performance issues when a GPU is present. This is a serious issue that needs to be addressed in future, for the scheme in Figure 4.1 to be fully realised.

4.5 SUMMARY

The scalings of the different components of the calculation, with respect to both fragment and overall system size, are given in Table 4.1. It demonstrates both the advantages and limitations of the new method - all of the major steps have been reduced to essentially linear scaling, with the exception of charge transfer and the Coulomb part of the Fock matrix. The former could be made linear in the same way as the RPA portion, by applying the procedure for relevant pairs of fragments, but the associated prefactor is too small to warrant doing so. The charge transfer contribution is only formally quadratic due to a single matrix multiplication, but in reality is

Table 4.1: Summary of the formal scaling of the different parts of the ALMO+RPAxd calculation, along with their conventional counterparts.

Procedure	Overall	Fragment	Conventional
Full ERIs	—	n^4	N^4
DF ERIs	Fn^2m	n^2m	N^2M
Cholesky	—	—	M^3
Diagonalisation	Fn^3	n^3	N^3
Coulomb	FMn^2	—	N^4
Exchange	Fn^4	n^4	N^4
Charge transfer	V^2	v^2	—
RPA	Fo^3v^3	o^3v^3	O^3V^3

asymptotically linear due to the sparsity of the matrices and the blocking of the ERIs. The disadvantages, on the other hand, are twofold: the density-fitting procedure requires the $\mathcal{O}(M^3)$ Cholesky decomposition of the Coulomb metric in the auxiliary basis; and the scaling within a fragment is still dominated by the o^3v^3 dependence of the RPA calculation. Despite the efficiency of standard algorithms for the former, it will dominate for very large systems, although this may be alleviated by the efficient use of parallelisation. The second problem implies that the method is well-specialised for extended systems of small to medium sized molecules - this is a design choice, as it is intended to be applied to the rapid, BSSE-free calculation of noncovalent interactions in many-body systems. In principle, it would be possible to apply existing local treatments to reduce the per-fragment scaling as the methods used are adaptations of standard procedures, but this would lose the unique advantage of starting from the ALMO approximation. As such, this possibility is not investigated here.

5 | TESTS AND BENCHMARKS

In this chapter, the methods and implementations of the previous three chapters are robustly tested and benchmarked. I demonstrate that the described asymptotic scaling is in practice achieved rapidly, and that the accuracy of the method over standard sets of supermolecular complexes is similar to that of coupled cluster, at a fraction of the cost. I also highlight and discuss some of the shortcomings, with possible explanations and strategies for iteratively improving the ALMO+RPA procedure in future.

The methods outlined have been implemented in GAMMA, a suite of quantum chemistry programs developed solely by the author. The density-fitted integrals were evaluated using a combination of custom routines and the LIBINT package of Valeev and coworkers [322]. The Cyclops Tensor Framework [321] was used for the tensor contractions in the RPA calculations, while the EIGEN library [316] was used for all other linear algebra operations. Additional reference calculations in the following were performed in GAMMA, and verified using the MOLPRO 2015.1 package [323]. Unless noted otherwise, all calculations were done using the aug-cc-pVDZ (aVDZ) orbital basis [128, 129, 324], with the associated JKFit [151, 162, 289, 325, 326] and MP2Fit [325, 327] auxiliary bases for the Fock- and RPA-integrals, respectively. Unless stated otherwise (c.f. parallelisation), all timing and memory benchmarks were determined on a single processor on the same desktop machine, with reference coupled-cluster timings taken from the MOLPRO calculations. In the following, we apply the ALMO+RPAxd method to a variety of chemical systems, demonstrating and assessing the accuracy of the method, the errors associated with the various approximations, and the scaling with fragment and system size, including the use of parallelisation. Finally, the energy decomposition analysis is used to probe the nature of the interactions, and compared with existing results. The latter were determined using density-fitted SAPT2 [86] calculations performed in Psi4 version 1.1 [328].
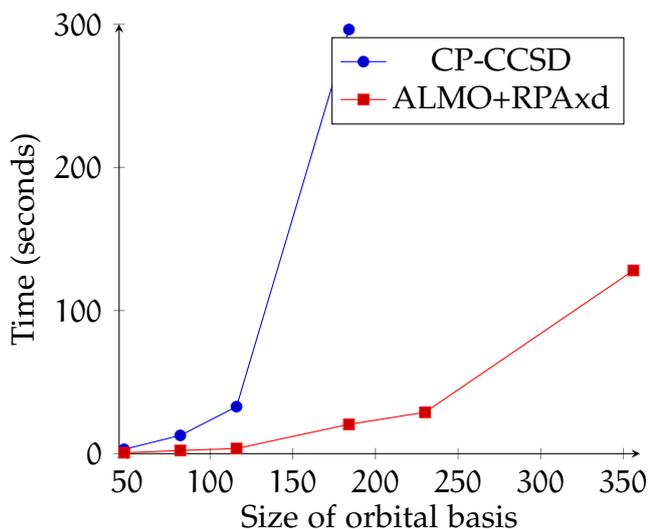
Figure 5.1: The total time taken to calculate the interaction energy for the water dimer in basis sets of increasing size. The overall scaling for ALMO+RPAxd is cubic, compared to the sixth-order scaling of CCSD, leading to substantial speedups.

5.1 SCALING

The scaling with number of basis functions per fragment is demonstrated by the water dimer with the progression of orbital bases cc-pVnZ and aug-cc-pVnZ, n =D, T, Q [128, 324]. This is shown in Figure 5.1. The scaling with fragment size for the ALMO+RPAxd method is approximately cubic, in agreement with Table 4.1; for fragments with larger occupied spaces than water, the prefactor might be expected to be more severe. Most importantly, the time required per fragment quickly becomes multiple orders of magnitude smaller than for CP-corrected CCSD, while achieving essentially the same accuracy, as will be demonstrated later.

To study the scaling with overall system size, we consider two extreme cases: a linear chain of hydrogen fluorides, and a series of clusters of water molecules. The former is the best possible scenario, with a single additional hydrogen bond being added with each new fragment, and with interactions happening along a single axis. The geometries for these chains were made by optimising the dimer at the CCSD/aVDZ level, then replicating the intra- and inter-molecular separations for each new monomer added to the chain. The water clusters, on the other hand, are known to be particularly difficult [329], as they are tightly bound, with each new fragment being close to and interacting with multiple other fragments. The geometries for these were optimised, for up to ten water molecules, using M06-2X [310] with an aVDZ basis. The larger clusters were sampled from molecular dynamics simulations of a 2.3 nm cubic box with a density of
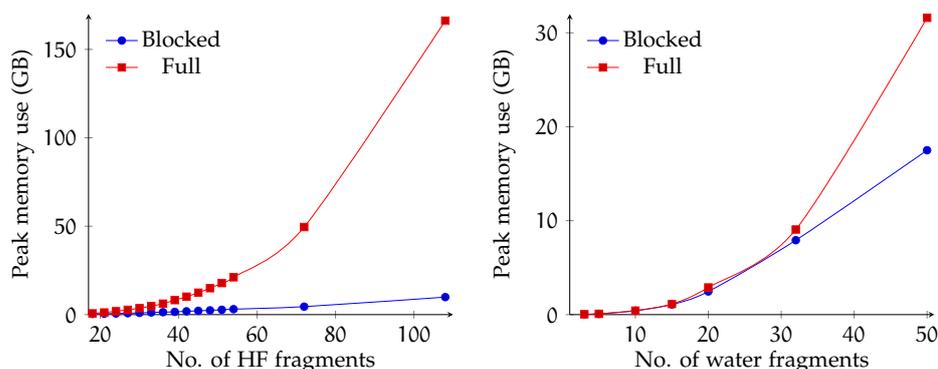
Figure 5.2: The peak amount of memory required during the ALMO+RPAxd calculation on either the linear hydrogen fluoride chains (left) or water clusters (right), showing the improvements resulting from using the blocked ERIs (Algorithm 4.2) as opposed to the full ERIs (Algorithm 4.1).

roughly 997 kg m^{-3}, performed in GROMACS 2016.4 [330] using the CHARMM36 [331] version of the TIP4P water model [332].

Firstly, the efficacy of the blocked DF ERI scheme outlined in Algorithm 4.2 is shown in Figure 5.2. In the case of the HF chain, the memory dependence can be seen to become linear very rapidly. This allowed for the largest system considered here - 108 HF molecules, corresponding to approximately 3,500 orbital basis functions and 12,000 auxiliary functions - to be computed on a desktop computer, using just under 10 gigabytes, compared to the over 150 GB that would be required for all the DF ERIs to be stored. The water clusters show a much slower approach to the linear regime, only becoming sub-quadratic for the largest system of 50 monomers. This is a reflection of the denseness of the integral tensor for these systems, as all the water molecules are very close together and the Coulomb interaction decays slowly. Despite this, substantial memory savings become apparent for the larger clusters.

In contrast, the integrals for the HF chain reach a sparsity of 95% for 108 fragments; in reality, the sparsity is higher than this due to the coarseness of the blocking. This is reflected in the time taken to build the Coulomb matrix, shown in Figure 5.3, which seems to be increasing linearly with system size. In fact, it is roughly quadratic, but with a very small prefactor - the construction is so rapid that the $\mathcal{O}(M^2)$ matrix multiplication is dominating. The one-dimensional, linear system is ideally suited for the local exchange approach, with the domains reaching a roughly fixed size for as few as 9 monomers, as can be seen in Figure 5.4. As would be expected intuitively, the core orbitals and those closer to the end of the chain have smaller domains, while the more diffuse, higher-energy orbitals in the centre extend further out. Notably, the MO and ABS domains stay localised
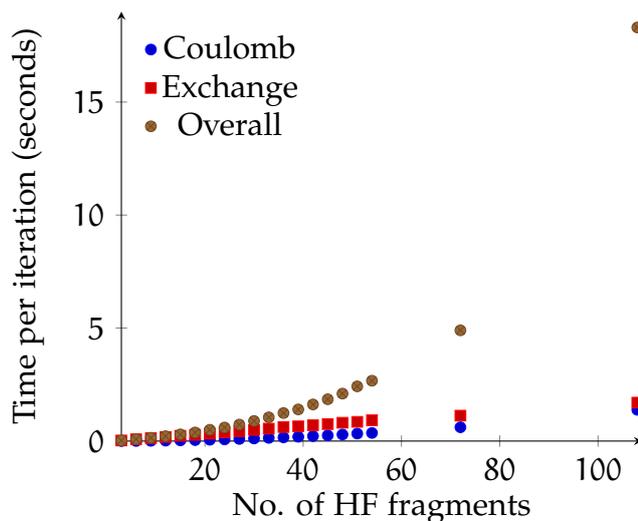
Figure 5.3: The time taken to build the Coulomb and exchange portions of the Fock matrix in each ALMO iteration for linear chains of hydrogen fluoride molecules. Both seem to have rapidly reached the linear scaling regime, to the point that the matrix multiplications (essentially the remainder) take up the majority of the computation time per iteration.
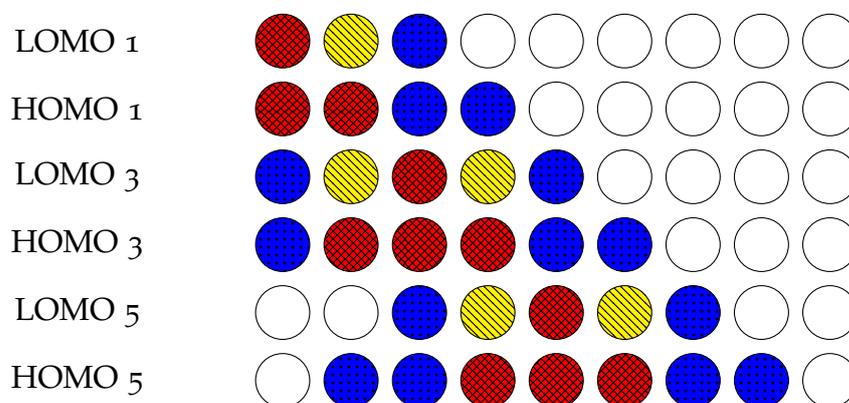
Figure 5.4: A visualisation of the orbital domains for the lowest (LOMO) and highest (HOMO) occupied molecular orbitals on fragments 1, 3, and 5 (from left to right) in a linear chain of nine hydrogen fluorides. The fragments are represented by the circles, where red with a crosshatch pattern denotes that a fragment's atomic orbitals are in all three domains for that molecular orbital, yellow with diagonal lines that they are in the AO and ABS domains, and blue with dots implies they are only in the AO domain.
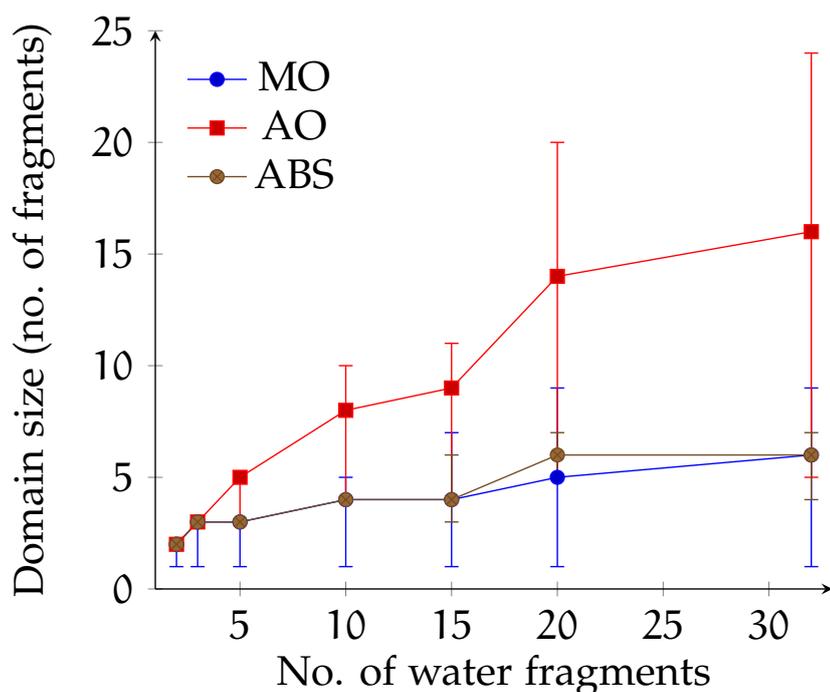
Figure 5.5: The average domain sizes, in terms of number of fragments, for clusters of water molecules, along with bars representing their range in size. While the MO and ABS domains increase slowly, the AO domains cover a large portion of the entire system, even for larger clusters.

to only one to three fragments, while the AO domains extend much further out. The result is true linear scaling in the construction of the exchange matrix, also shown in Figure 5.3. This means that, even for approximately 3500 orbital basis functions, the combined Fock build takes less than five seconds per iteration. Again, the matrix multiplications become the dominant contribution, in this case the formation of the fragment Fock matrices (equation 2.54). At that point, there is very little use in trying to find further savings.

The domains for the water clusters, however, do not reach a fixed size so quickly. Figure 5.5 shows that in this case, while the average size of the MO and ABS increases slowly, becoming fixed at around six fragments, the AO domains are essentially extended over the whole system even for large cluster sizes. The variation in all domain sizes is also quite sizeable, with some MO domains covering a third of the system. This suggests considerable density is being transferred between water molecules, as measured by equations 4.11 and 4.13. Nonetheless, the power scaling of the Fock build does steadily decrease from cubic towards linear. This can be seen in Figure 5.6, where the percentage of time taken on each part of the calculation is shown for both the chain and the clusters. Whereas for the former, the Fock build percentage steadily decreases reflecting its reduced scaling
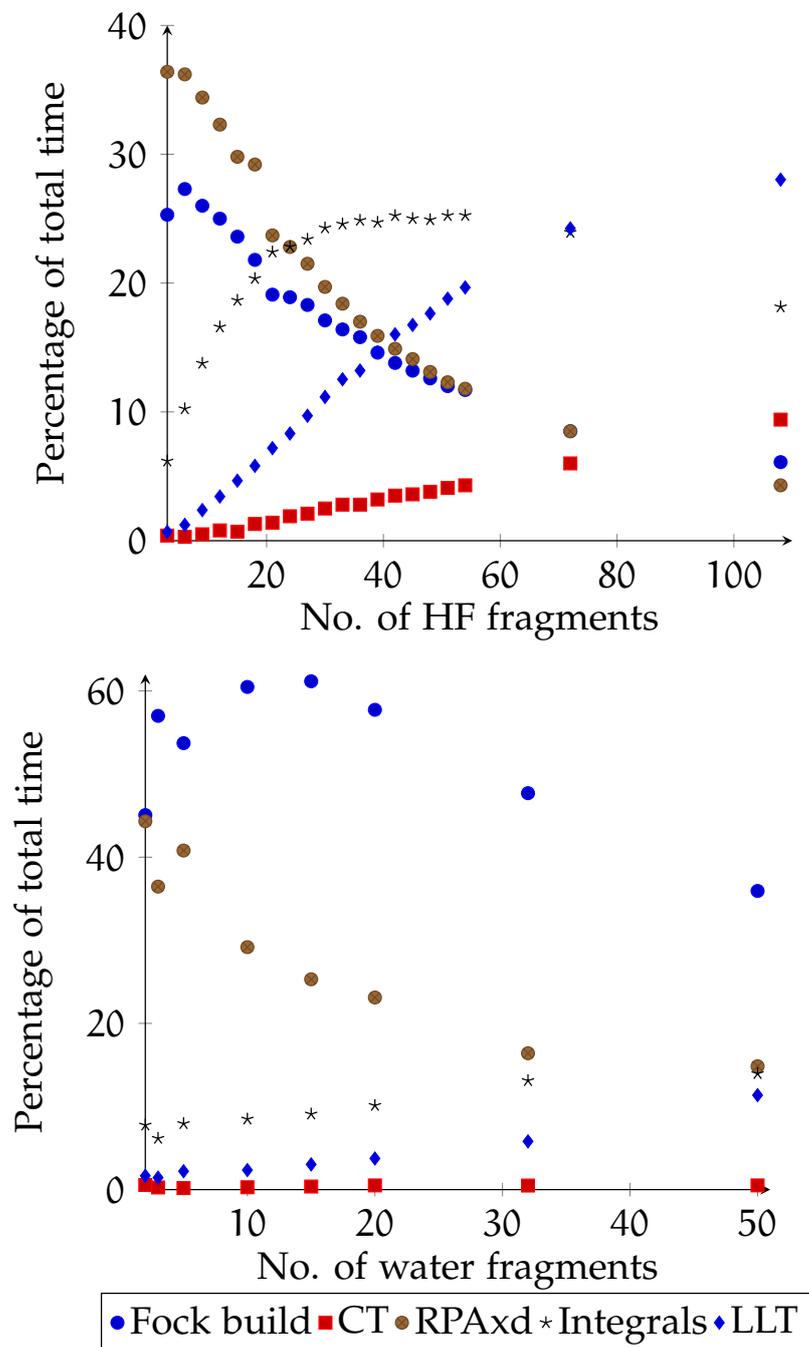
Figure 5.6: The percentage of total computation time spent in each part of the ALMO+RPAxd calculation, for varying sizes of HF chains (top) and water clusters (bottom). LLT is shorthand for Cholesky decomposition.

Table 5.1: The errors per fragment (in kcal/mol) associated with the density fitting and local approximations for various lengths of linear hydrogen fluoride chains, along with the times (in seconds) for each ALMO iteration using the full versus the local Fock build.

No. fragments	DF error	Local error	t_{Full}	t_{Local}
3	0.0021	0.0069	0.08	0.03
6	0.0017	0.0065	2.8	0.11
9	0.0015	0.0064	18.9	0.18
12	0.0016	0.0064	68.8	0.26

in comparison to the dominating step, for the water clusters the Fock build is consistently the most expensive part of the calculation, only beginning to decrease at around twenty monomers. In both cases the cost for the integrals can be seen to plateau, and then begin to decrease for the linear chain, demonstrating that this is becoming linear, while the Cholesky decomposition and charge transfer - cubic and quadratic, respectively - are the only contributions increasing in significance. In particular, the pairwise RPA percentage is monotonically decreasing, going from being one of the most expensive steps to one of the least for both systems. Collectively, consideration of the two limiting cases has shown that linear or near-linear scaling has been achieved for all segments except the charge transfer and Cholesky decomposition, as predicted by Table 4.1. In particular, it appears that the latter will quickly become the bottleneck in very large systems. Irrespective of this, however, the 108 HF and 32 water calculations both took under ten minutes on a single processor.

5.2 ERRORS

While the density fitting, local Fock build, and pairwise RPA procedures have clearly been successful in reducing the cost of the calculation, they are approximations and thus have errors associated with them. It is therefore important to calibrate the various thresholds used so as to obtain the optimum balance of accuracy and cost. Table 5.1 summarises the average error in the interaction energy per fragment due to the density fitting and local approximations, and the savings in time taken per iteration by using the latter, for the first few HF chains. This shows that the error due to both, when reasonable thresholds are used, is minimal, totalling less than 0.1 kcal/mol in the final interaction energy, which is well within the expected limits of accuracy of the method. More importantly, this comes with savings of over two orders of magnitude in the computation time, even for as few as nine fragments. It has been observed in previous versions of the local
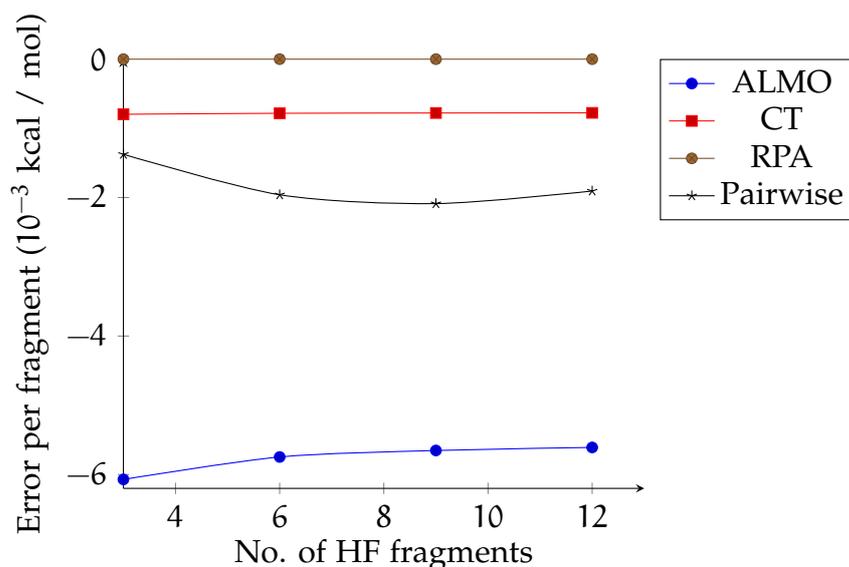
Figure 5.7: The error per fragment in the different contributions to the interaction energy caused by using the local approximation in the Fock build and the pairwise approximation in the RPAxd calculation, for different sizes of hydrogen fluoride chains.

exchange approximation that when diffuse functions are included in the basis set, as has been done here, relatively large errors are seen in the correlation energy [123]. Additionally, as noted earlier, a full Fock build was necessary to accurately reproduce the energies. Table 5.1 and Figure 5.7 demonstrate that for the more coarsely partitioned, fragment-based approach used here, this is categorically not the case. Essentially all of the error due to the local approximation is found in the ALMO result, with a small amount in the perturbative correction, but this error is less than 0.01 kcal/mol per fragment. A full Fock build is therefore entirely unnecessary. We find this to hold generally true, for all but purely dispersively-bound systems; these will be discussed in more detail in the benchmarking section of this chapter. Also shown in Figure 5.7 is the error due to using the pairwise RPA method as opposed to using the full orbital space; it is even smaller than the local error, most likely due to the decomposition of the excitations into essentially pairwise terms, and the inherent localisation of the orbitals.

Finally, Figure 5.8 shows how the error depends on the ϵ_{MO} and blocked DF thresholds. We note separately that the error due to the ϵ_{ABS} threshold was found to be negligibly small in all cases for values of 0.05 or below. Both figures show that the largest error is always in the ALMO energy, suggesting that the energy is more sensitive than the density, in line with previous studies [122, 123]. The RPA calculation, for which new integrals are computed but the ALMO density used, shows the least dependence and sub 0.001 kcal/mol
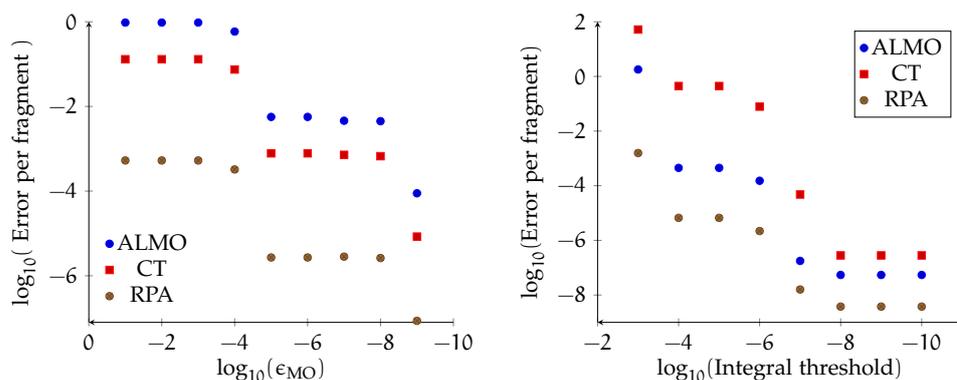
Figure 5.8: The error in the interaction as a function of the MO domain threshold (left) and integral screening threshold (right), taken as the average from calculations on a chain of 24 hydrogen fluorides and a cluster of ten water molecules. The error per fragment is in kcal/mol, while both thresholds have units of Hartree (as they are energy thresholds).

errors. For the other contributions, the figures show a series of steps, where decreasing the threshold allows for new fragments or integral blocks to be included. From these, it appears that a choice of ϵ_{MO} around 10^{-6} and an integral threshold of 10^{-9} essentially eliminates the errors.

5.3 PARALLELISATION

The parallelisation of the ALMO+RPA algorithm, as described in section 4.4, is multi-faceted. It uses three distinct forms of parallelism - multithreading, distributed processes, and GPU acceleration - ultimately in tandem. In this section, I show the performance of these approaches separately, and in tandem. We are somewhat limited by hardware constraints, however, and so have not yet been able to realise the combination of all three at once. In figure 5.9, the multithreading is demonstrated on the same systems that were used earlier to show scaling with number of fragments: a chain of 54 hydrogen fluorides, and a cluster of 32 water molecules. These are at the two extremes of when the fragmented approximation works, and while both comprise many *small* molecules, they clearly show the effects of the threading.

Both systems show almost ideal speedups at two threads, with each tailing off at different rates as the number of threads increase. In particular, the HF chain, which is dominated by the RPAx portion of the calculation, shows only about a five-times speedup at 8 threads, compared to slightly over six-times for the water cluster. As the latter is dominated instead by the ALMO iterations, this suggests that the
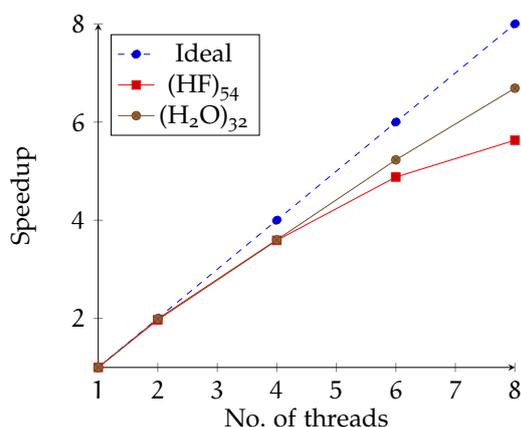
Figure 5.9: The combined speedup in computation time for the integral evaluation, Fock build, and RPA portions of the calculation using multithreading. Results are given for the $(HF)_{54}$ chain and $(H_2O)_{32}$ cluster. The calculations were performed on the same compute node, separately, with no other calculations running. The timings for the serial calculations were 3.3 and 18.1 minutes for the fluoride chain and water cluster, respectively.

threading is most effective in the integral and Fock-build portions of the calculation. This is to be expected, as only the integral transformation of the RPA calculations has been multithreaded, and we would expect to see greater speedups with the distributed parallelism.

To really test the MPI code across multiple nodes, however, we need a much larger system than either of the above. To this end, we have chosen the cluster of 202 water molecules, the geometry again taken from the molecular dynamics simulations described earlier, shown in figure 5.10. This has a high density of hydrogen bonds in a relatively small volume, providing a complex Fock build and a large number of non-zero pairwise RPA calculations.

We chose 202 molecules as this was the maximum size that would fit in each node's memory, at roughly 62 GB.

Figure 5.11 shows how the timings for the ALMO iterations, density-fitted integrals, and pairwise RPA calculations are affected by using multiple processes across two nodes, and with variable numbers of threads per process. The greatest total speedup is seen with 16 processes at 4 threads each, reducing the total time for the calculation from 2 hours 20 minutes to 10 minutes 11 seconds (approximately 14 times speedup). This is marginally better than 8 processes with 8 threads - the same overall number of threads - by about 61 seconds. This can be explained by a more detailed study of the individual portions of the calculation.

In contrast to figure 5.9, we see that the distributed parallelism has less effect on the ALMO iterations, but a much greater effect on the integrals, and particularly the RPA calculations. This is due to the almost embarrassingly parallel nature of the latter. Almost ideal speedups are maintained up to 8 processes (7.3 times). When com-
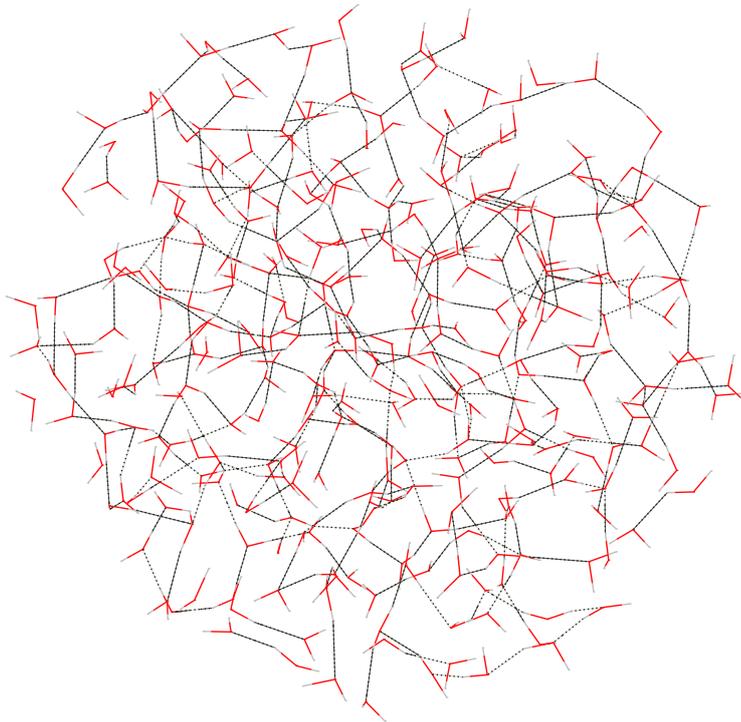
Figure 5.10: A densely packed cluster of 202 water molecules as used in the MPI benchmarking calculations. Hydrogen bonds between water molecules are shown as dotted lines. Due to the density, this is expected to be a worst-case scaling scenario.
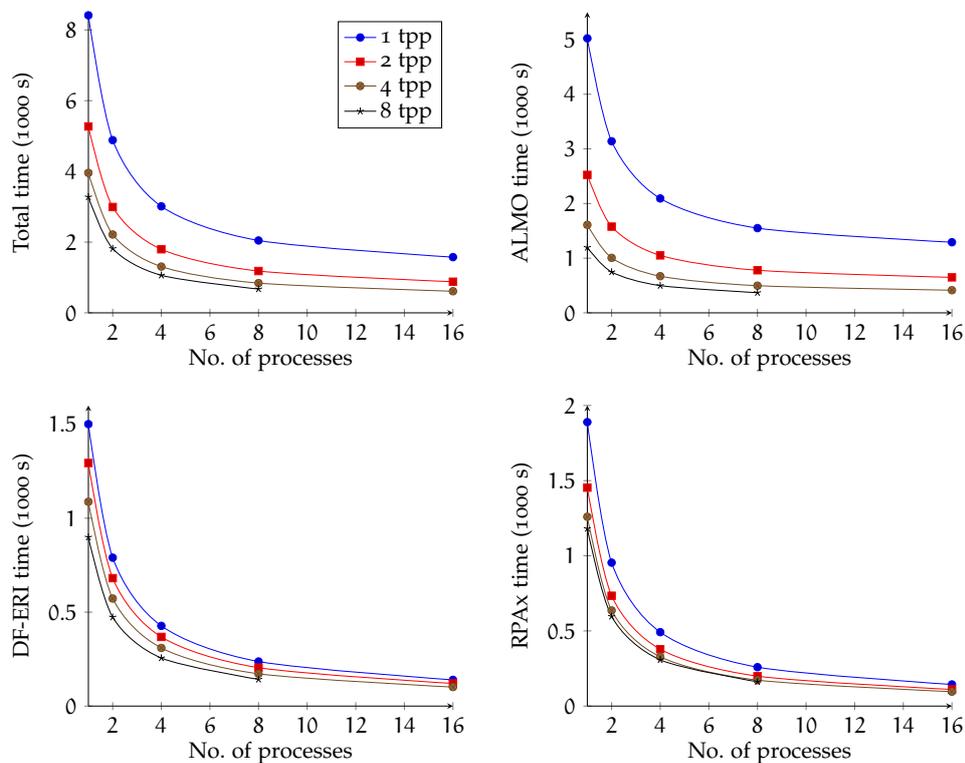
Figure 5.11: Time taken in different parts of the ALMO+RPAx calculation on a cluster of 202 water molecules, using combined distributed- and shared-memory parallelism. The x-axis shows the number of separate processes (distributed), with each line showing the number of threads-per-process (tpp). These calculations were run on up to two compute nodes, with 16 cores and 64 GB of memory each; at two threads per core, the 16 processes at 8 tpp was the only combination not possible. The total time unparallelised is 8413 seconds (2 hours 20 minutes), reducing to 611 seconds (10 minutes) with 16 processes at 4 tpp.

bined with four threads per process this improves to a roughly 11 times speedup.

The balance of times however is roughly 4 : 1 : 1 ALMO : integrals : RPA. While there is nowhere near the same level of improvement in the ALMO iterations - largely due to the unavoidable collection points where the density and total Fock matrices are constructed - we still see a 3.2 times speedup going from 1 to 8 processes, improving to 10.1 times when using 4 threads per process. Overall, this balance of shared- and distributed-memory parallelism seems to work best, allowing for calculations on very large systems - in this case with over 8,000 orbital basis functions - in negligible time.

The final facet is the use of GPUs to accelerate the linear algebra operations, in particular the tensor contractions in the integrals and RPA calculations. In an ideal setup, we would have several nodes with ultrafast interconnects, each with their own GPU accelerators. We only had access to a desktop computer with a single GP-GPU, however, so the validation done here is expected to be a lower bound on the possible improvements. It should be noted that, despite the NVIDIATitan Xp only allowing for single-precision floats, we saw no reduction in the numerical stability of the algorithm, most likely as only a subsection of the operations were actually performed on the GPU.

Figure 5.12 shows the combined effects of multithreading and GPU acceleration on two model systems: the benzene dimer, and a cluster of 101 water molecules. The latter was chosen for similar reasons to the larger water cluster used previously in the MPI study. The benzene dimer, on the other hand, represents a small system of two fairly large monomers (42 electrons, 192 basis functions), which would not benefit at all from the distributed parallelisation outlined earlier. We therefore expect this system to give the best indication of the improvements possible in using GPUs in the RPA tensor contractions.

As can be seen in the figure, the benzene dimer benefits much more from the GPU than the water cluster - a total speed up of slightly over two times on one thread. The water cluster, on the other hand, has much more pronounced improvements from the multithreading, with the GPU only offering a modest speedup of about 1.3 times. Inspecting further, we see this is due to the GPU having only a small effect on the ALMO iterations while having a much greater impact on the RPA calculations. As the latter dominate the benzene dimer computation time, this sees greater improvement.

In all cases, there is little improvement in going from 4 to 8 threads. This is in contrast to figure 5.9, where the change is below ideal, but still significant. There are two possible explanations for this: firstly, the computer on which the current calculations were performed only possesses four cores, such that 8 threads is the maximum with hyper-
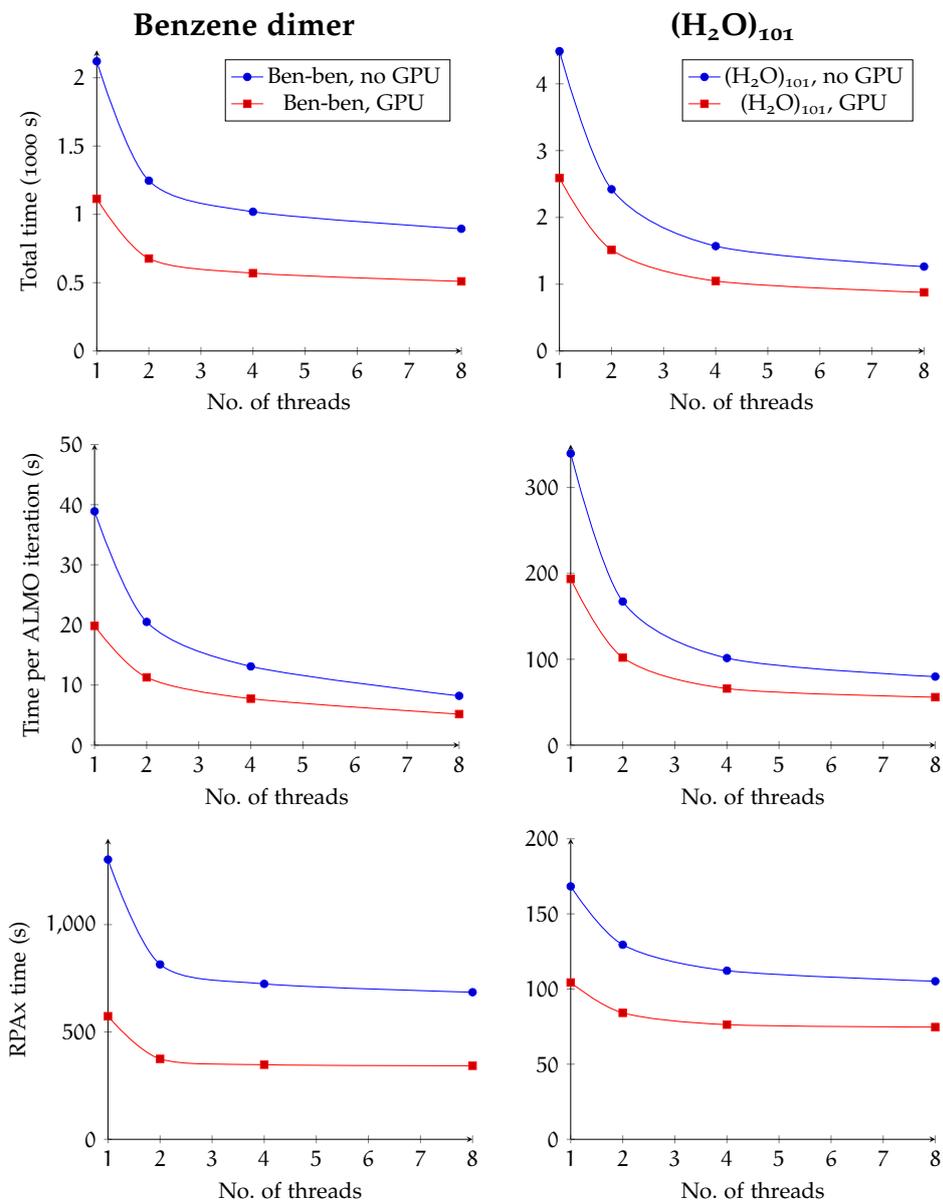
Figure 5.12: Timings for parts of the ALMO+RPAx calculation on two model systems: the benzene dimer (large monomer, small number of fragments) and a cluster of 101 water molecules (small monomer, large number of fragments). Comparisons are given between multi-threading with and without the addition of a GPU accelerator. All calculations were performed on the same computer with 4 cores, 24 GB of memory, and an NVIDIA Titan Xp.

threading; the calculations in figure 5.9 were performed on a node with 16 cores, so that hyperthreading was not required at 8 threads. The second explanation is that the contribution from the ALMO iterations is simply less significant, and as these benefit most from the multithreading, a less pronounced effect is seen. This explanation makes sense for the benzene dimer, but less so for the water cluster, which one would expect to behave in a similar manner to the 32 molecule water cluster earlier.

In summary, the three-fold parallelisation scheme seems to be effective, with each part working cohesively together. There is still considerable work that needs to be done, however, in refining the individual parts to work on a wide range of systems. Of particular difficulty is the messaging between nodes in the distributed parallelisation; there was a noted decrease in efficiency when switching from only 8 processes (requiring only a single node) to 16 processes. On top of this, we are yet to test the interaction of the distributed processing with the use of a GPU. As these both contribute significantly to the increased efficiency of the RPA calculations, it is possible they will compete, rather than work together. Moreover, we are limited by the use of a third party library with limited architecture support for the GPU acceleration. Future work would need to focus on fine-tuning the distribution to work more effectively on any given hardware.

5.4 POTENTIAL ENERGY CURVES

'Noncovalent interaction' is a broad title covering many chemically and physically distinct classes of system. To assess the new method, it is therefore important to select a range of typical and difficult examples. In this regard, we consider: the water dimer, a widely studied example of hydrogen bonding; $FCl\cdots OH_2$, an unusually strong halogen bond; the very weakly bound dispersive helium-helium and helium-neon dimers. The natural points of comparison for the ALMO and RPA approaches are counterpoise-corrected HF, MP2 and CCSD calculations; the latter are equivalent in the sense of primarily treating double excitations.

The potential energy curves for the first of these, the water dimer, are shown in Figure 5.13. The upper figure demonstrates how the charge transfer correction to the ALMO energy is necessary to redress the large errors compared to CP-corrected HF. The ALMO+CT result lies slightly below the HF curve, implying that the CT term contains a small amount of BSSE. This is in line with expectations, and while unfortunate, the error is on the order of 0.1 kcal/mol, which is only a small percentage of the total interaction energy. Importantly, the overall shape of the curve is reproduced correctly, whereas the raw ALMO
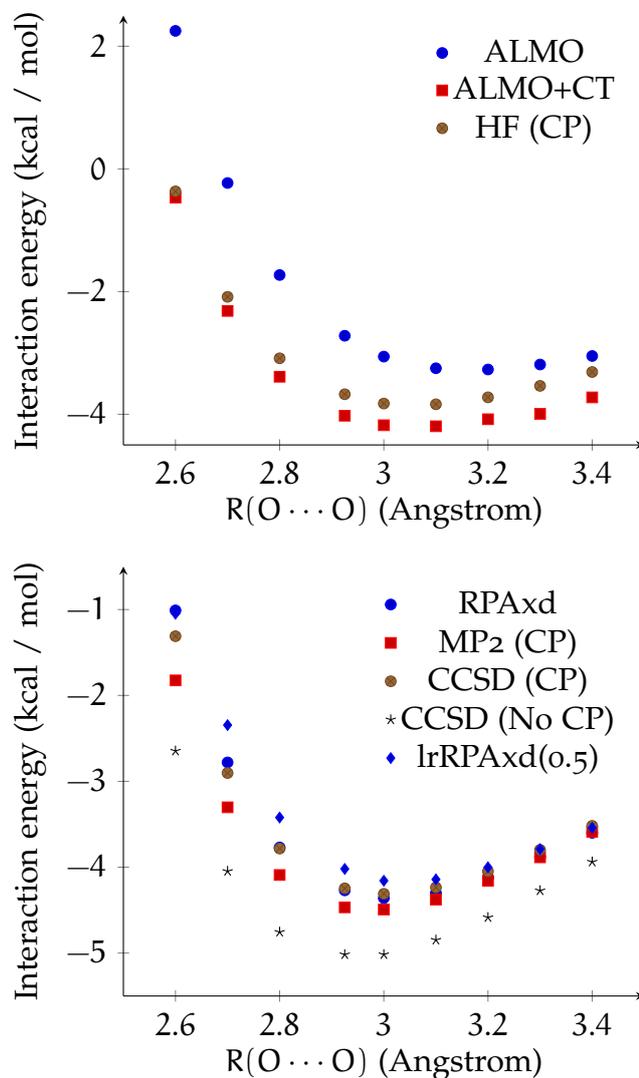
Figure 5.13: Potential energy curves for the water dimer in the aVDZ basis, as calculated with different methods. The upper figure shows the mean-field results, and the importance of including the charge transfer correction to the ALMO energy. The lower figure shows the result of including correlation, and demonstrates the large deviations caused by not removing the basis set superposition error.
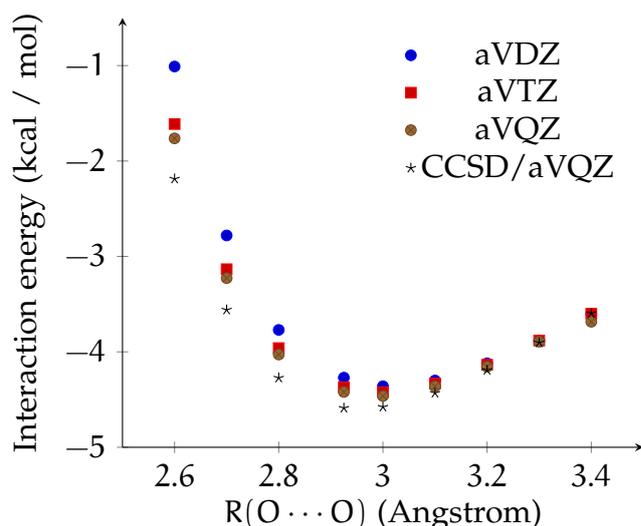
Figure 5.14: Potential energy curves for the water dimer calculated using the ALMO+RPAxd method with the series of orbital bases aVnZ, n =D, T, Q. For comparison, the CCSD/aVQZ result is also shown, to which the former appear to be converging.

curve leads to a significant overestimation of the equilibrium bond distance. Upon including correlation, it can be seen that the RPAxd approach leads to excellent agreement with the CP-corrected CCSD result. In particular, it consistently performs better than MP2, giving a root-mean-square error of 0.12 kcal/mol compared to CCSD as opposed to 0.27 kcal/mol for MP2. Finally, it appears that the BSSE has been successfully eliminated, with the not CP-corrected CCSD curve lying on average more than 0.8 kcal/mol below its corrected counterpart. This demonstrates again that any BSSE present in the CT term is minimal. The other possible approach to removing BSSE - namely the range-separated integrals, lrRPAxd(μ), where μ is the range separation parameter of equation 3.69 - is also shown, but performs worse than the RPAxd variant, overcorrecting for the error and yielding an incorrect shape. When combined with the arbitrariness in the choice of range-separation parameter, it seems that this approach is not to be preferred.

From Figure 5.13, it appears that the method performs worst in the short-range region. It has been noted previously that RPAx overestimates the exchange-repulsion [232, 255], and it is possible that this effect is also being seen here. Figure 5.14 shows that the description in this region is highly basis-set dependent, whereas the rest of the curve is not, suggesting that a larger correlation space is necessary to get a good description of the short-range exchange-dispersion terms. Notably, however, the position of the minimum is fairly insensitive to the quality of basis set, and the error compared to CCSD appears to stay roughly constant at each level. This is promising with regards
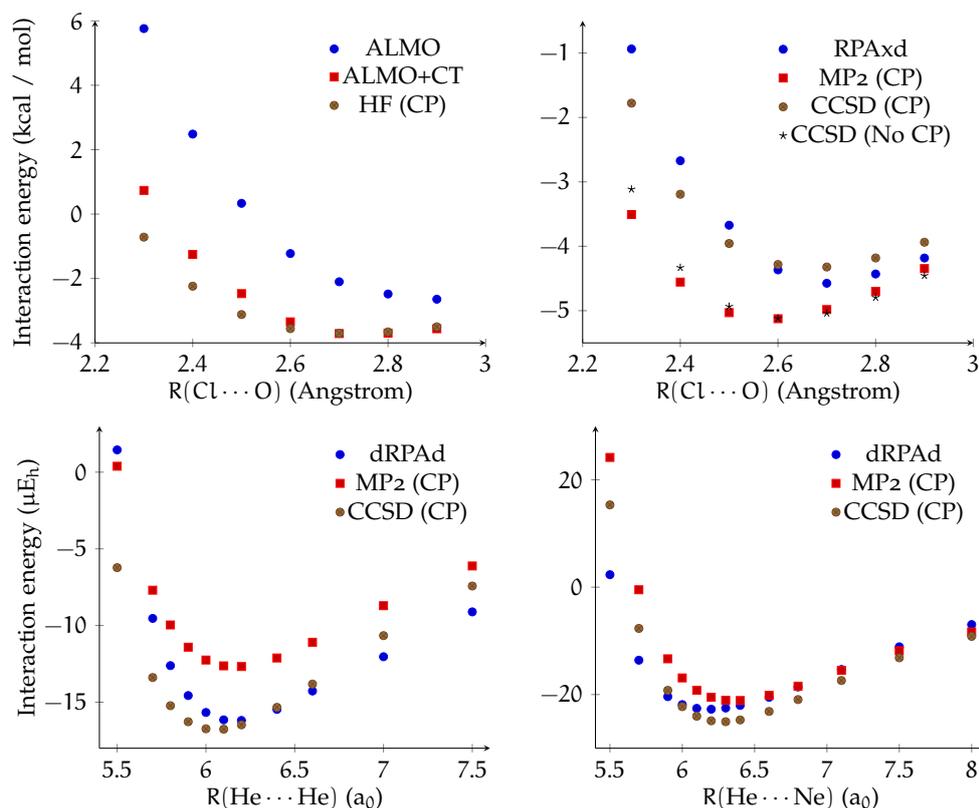
Figure 5.15: Potential energy curves for $FCl\cdots OH_2$ using mean-field (top left) and correlated (top right) methods, and for the helium-helium (bottom left) and helium-neon (bottom right) complexes. These again demonstrate the importance of including charge transfer, and the excellent agreement of ALMO+RPAxd with CCSD results.

to eventually using the method for rapid structure prediction of large systems.

The curves for the halogen-bonded system and the noble gas dimers are shown in Figure 5.15. The former corroborates the observations for the water dimer, with the inclusion of charge transfer reproducing almost exactly the CP-corrected HF result, and RPAxd achieving roughly 0.1 kcal/mol accuracy compared to CCSD. Interestingly, the CP-corrected MP2 result overbinds this complex so severely that it gives broadly similar results to the uncorrected CCSD curve; the latter again demonstrates the presence of substantial BSSE at this level, and the successful avoidance of this in the new method. The dispersively bound noble gases present perhaps the toughest challenge. Despite the interactions being on the micro-Hartree scale, however, excellent agreement is still seen with the CCSD results, and in the case of the helium dimer, this represents a significant improvement on the MP2 curve. It should be noted that the dRPAd variant is used in these calculations, as this eliminates the overestimation of exchange

repulsion. The error is small enough to be well within chemical accuracy, as shown in the previous two systems, but on the minute energy scales here, such errors can become significant. An additional reason for using the direct RPA dispersion energy is that it will allow us to compute directly dispersion coefficients later, as described in equation 3.43.

5.5 BENCHMARKS

Intermolecular interactions are many, varied, and ubiquitous. It is therefore vital to benchmark thoroughly on as broad a range of complexes as possible. This is not possible, in general, for complexes involving more than two fragments, as accurate calculations on these are too difficult; that is the primary motivation behind the new method. High-accuracy benchmark sets do exist for two-fragment systems, though, and we will focus on two such sets: the S66 database of biologically-relevant interactions [267], and the X40 set of halogen-bonded systems [333]. The former are particularly interesting, as a potential application of the ALMO+RPA method is in investigating solvent effects in biological systems at an unprecedented level of accuracy. The halogen bonds are equally very relevant in the fields of crystal and material design [334, 335], primarily in the solid state - with the connection of the RPA method to the GW approach as noted in Chapter 3, it is easy to conceive of adaptations to the method that would allow it to be used in the solid state.

I have calculated interaction energies for every system in both sets using examples of all the standard quantum chemical methods, in the first instance using aVDZ basis sets. Methods in which BSSE would be present (HF, DFT, MP2) have been counterpoise corrected. The geometries were optimised at the CCSD/aVDZ level, and errors are calculated relative to this level, as ALMO+RPA is inherently a double-excitation method. Neither it nor MP2 contain the physics of triple excitations in the diagrammatic sense, and so comparisons to CCSD(T) would make it impossible to distinguish errors due to the triples correction, as opposed to errors inherent to the method. For reference, the mean-absolute errors of CCSD relative to CCSD(T) on the S66 set are 0.41 kcal/mol at the aVDZ level.

RPA does contain three-body contributions, as will be shown later, but these are distinct from the excitations, which refer to the number of particle-hole lines in the diagrams.

Figure 5.16 shows the error distributions for various methods over the S66 set. From this, it is clear that of the ALMO methods, the SO-SEX dispersion term gives the best overall results. In fact, it appears to give the best in comparison to any of the other methods: its error distribution is narrow, and roughly normally distributed around zero. This suggests that it is not biased towards over- or under-estimation, and gives consistently good results over the whole set. The mean

Figure 5.16: Error distributions, relative to CCSD/aVDZ interaction energies, for the S66 database, as calculated using ALMO with three variants of RPA. These are shown with errors for DF-MP2 and DF-SAPT2, which both have second-order dispersion, and M06-2X, a popular density functional for intermolecular interactions.
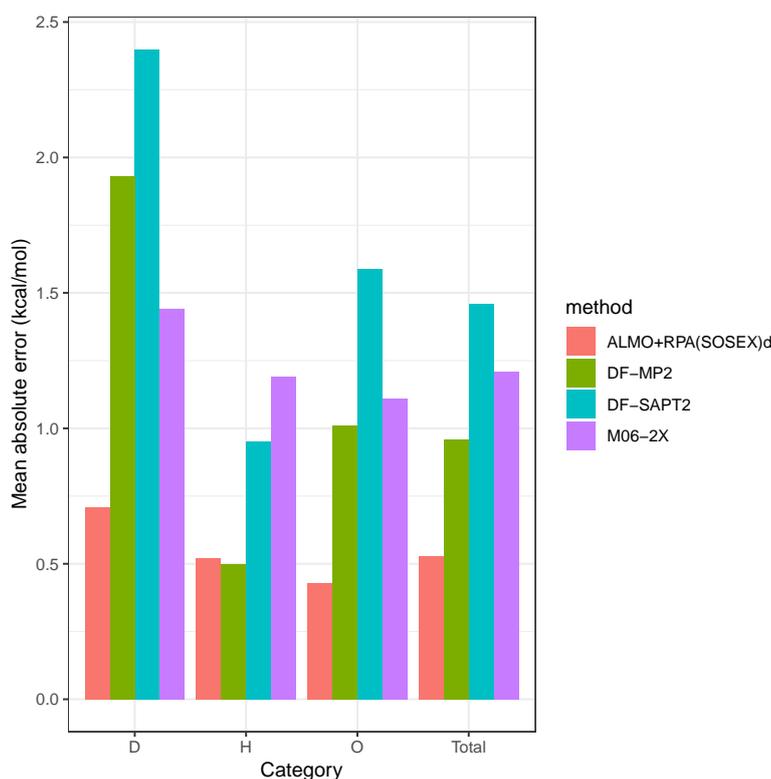
Figure 5.17: The mean absolute errors for different methods with an aVDZ basis, across the S66 database, relative to the CCSD/aVDZ results. The errors are further split by interaction type, as in the original S66 paper [267]: hydrogen-bonded (H); dispersion dominated (D); and other (O).

absolute error (MAE) is 0.51 kcal/mol. The dRPA variant does similarly well, but the distribution is slightly broader, presumably due to the lack of accurate exchange. The RPAx dispersion is significantly shifted towards underestimation (positive errors) for a certain subset of systems, reducing its success. In comparison, both DFT and MP2 significantly overestimate the interaction energies, and both have tails extending to an absolute error of over 4 kcal/mol. Finally, SAPT2 has a very broad distribution, again tending towards overestimation of the interaction energies.

To investigate the performance of each method further, we note that the S66 dataset was originally split into three categories [267], based on the dominant physical contribution to the interaction. Alternatively, they were split into three (related) 'types' of complex: hydrogen bonded, dispersively bound, and other (somewhere in between). The errors for each method on each subcategory are shown in Figure 5.17. This more clearly shows that, overall, ALMO+RPA(SOSEX)d significantly outperforms the other methods, halving the MAE of the nearest competitor (MP2).

Even more reassuringly, it performs consistently well on all three types of system. In contrast, MP2 does as well for the hydrogen-bonded complexes, but considerably worse for the dispersively-bound ones; in fact, for the latter, it is outperformed by DFT. SAPT also does poorly for these complexes, and looking at the decompositions (given in Appendix E, it becomes clear this is due to the perturbative estimation of the dispersion in both MP2 and SAPT2 resulting in a significant overestimation of the interaction. RPA, particularly the SOSEX variant, in contrast contains higher-order terms in the coupled cluster expansion, and is determined self consistently.

Surprisingly, however, the RPAx variant does not perform as well, despite this supposedly giving the best description of the exchange. It does, in fact, give marginally better errors for the hydrogen-bonded and mixed types, but in the dispersive systems, which generally have larger separations between the fragments, it overcorrects the exchange. This is a known problem with the RPAx variant: while it fixes the short-range exchange problems of dRPA, it does not reach the same long-distance limit as the latter, which we have already noted is exact. This is due to missing terms in the diagrammatic expansion [275], and as noted in Chapter 3, could be corrected for.

A further consideration is how much of the error is due to the ALMO approximation, as opposed to that due to only including ring diagrams in the dispersion energy. Comparison of the ALMO energies to counterpoise-corrected Hartree-Fock results shows an MAE of 0.95 kcal/mol, which reduces to 0.33 kcal/mol when the charge transfer correction is included. Determination of the BSSE from the counterpoise correction shows that the mean BSSE in the total energy at the HF and CCSD levels is -0.69 and -1.39 kcal/mol, respectively. This suggests we have successfully eliminated BSSE by using the ALMO approximation.

The errors across the X40 database, shown in Figure 5.18, are much the same again. The SOSEX variant once more outperforms all other methods, albeit with a slightly more skewed distribution of errors towards underestimation. SAPT2 performs much better on these systems than on the S66 set, perhaps due to there being far fewer dispersive complexes. There are still large, negative outliers for both it and MP2, however. The overall MAEs are 0.45, 0.79, 0.86, and 0.97 kcal/mol for ALMO+SOSEX, MP2, SAPT2, and M06-2X, respectively. In combination with the S66 data, this suggests that the ALMO+RPA approach is robustly accurate, and within chemical accuracy of the benchmark CCSD energies.

Of course, the argument for MP2 and DFT are that they are fairly cheap methods to use, and they seem to perform well enough for most systems. As these are all two fragments, we are not in the linear scaling regime of the ALMO+RPA method, so it is not clear whether
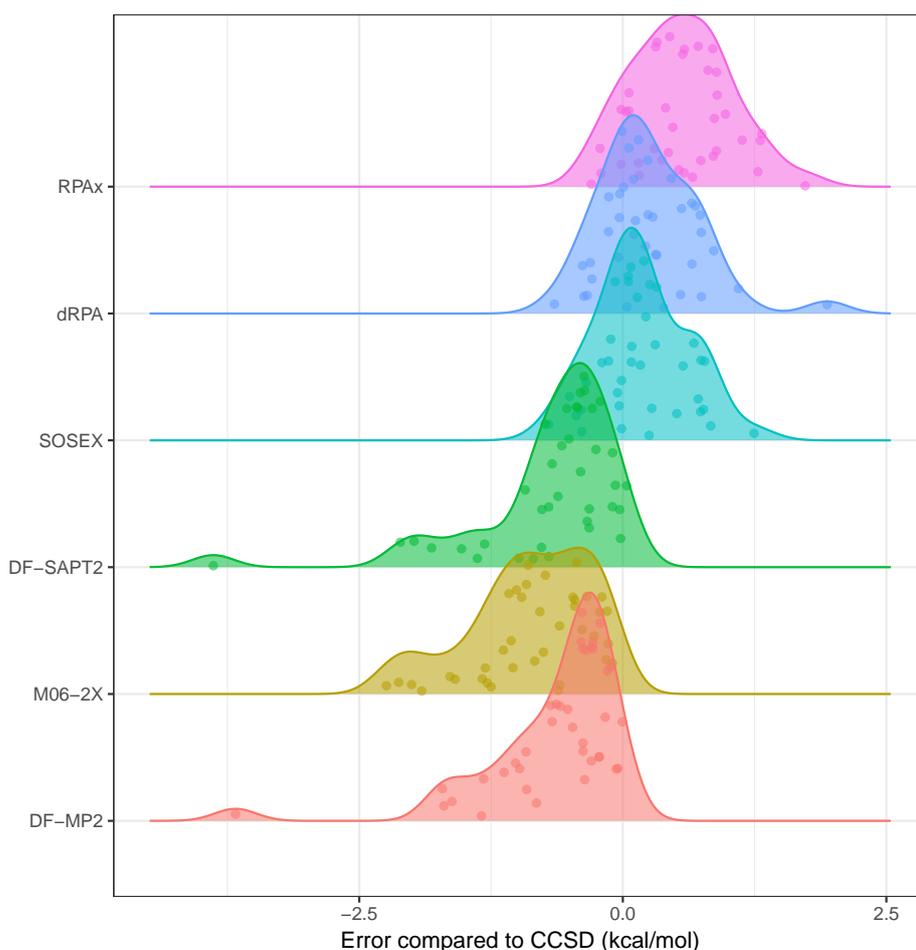
Figure 5.18: Error distributions, relative to CCSD/aVDZ interaction energies, for the X40 database of halogen-bonded interactions. As in Figure 5.16, a selection of methods are shown.

Table 5.2: Ratios of total wall time for various methods to that for ALMO+RPAxd, with and without the local exchange correction. Ratios are averaged across the S66 test set, with calculations carried out on a single core (32 GB of memory). The average ALMO+RPAxd computation time across the set was 447 seconds with the exchange correction, and 354 seconds without.

	DF-MP2	CCSD	CCSD(T)	M06-2X	DF-SAPT2
With X-correct	2.2	21.7	76.5	2.0	4.9
Without X-correct	2.9	27.9	99.6	2.5	6.4

there will be any significant savings, cost-wise, in using the current method. However, timings over the S66 set, all performed in isolation on the same compute node, with the same number of threads (one), show that even ALMO+RPAx - the most expensive variant - is on average 2 to 100 times faster than these other methods. The full set of ratios are shown in Table 5.2.

In the table, timings are given as ratios to ALMO timings with and without the X-correction. This is a correction to the exchange matrix that can be made at the end of the ALMO iterations, using the converged density, that does a full Fock build without the local approximation. The density is insensitive to the local correction, but this is not always true for the energy. Our particular flavour of local exchange is fairly robust, however, as it is coarsely divided over fragments rather than individual orbitals. For all but the dispersion-type systems, the X-correction makes up less than 3% of the total interaction energy. However, as already noted, the dispersive systems have larger separations, so either the distance threshold in the domain selection needs to be increased (which will increase the compute time), or an X-correction needs to be performed. In practice, the increase in cost is similar between the two approaches. Even with the correction, however, Table 5.2 shows that it is still better than twice as fast as MP2.

It should also be noted that these speedups will only increase as the number of fragments increases, and the linear scaling truly takes effect. Similarly, increased basis set size will not greatly affect the Fock build, and will affect RPA calculations on pairs of molecules far less than it would affect a correlated calculation on the full system. We have also performed calculations on the S66 set, using the benchmark geometries from the original paper, at the aVTZ level. These yield an MAE of 0.60 kcal/mol compared to complete basis set limit CCSD results, showing that accuracy is systematically maintained.

The remaining question that we have not so far answered is how well it performs when there are more than two bodies, and many-body effects become important. These are particularly difficult, as they are generally much smaller than two-body effects, and we are taking energy differences. The advantage of RPA over MP2, though, is that it does include three-body effects in the dispersion energy, whereas MP2 cannot. However, the pairwise approximation may detract from this slightly.

To investigate this, I calculated the three-body interaction energies of a set of 9 conformers of the benzene trimer, taken from Reference [336]. These are shown in Figure 5.19, along with the corresponding energies. Overall, we see that it does as well or better than MP2 in all cases except conformer 7 (bottom left), and is usually closer to the CCSD result. In fact, for two of the conformers (5 and 9), it is

	1	2	3
CCSD(T)	-0.034	0.057	0.134
ALMO+RPAx	-0.063	-0.039	0.088
MP2	-0.063	-0.042	0.065
CCSD	-0.043	0.033	0.123

	4	5	6
CCSD(T)	0.059	0.049	-0.046
ALMO+RPAx	0.030	0.047	-0.055
MP2	0.003	-0.0077	-0.064
CCSD	0.043	0.036	-0.051

	7	8	9
CCSD(T)	0.060	0.017	-0.009
ALMO+RPAx	0.081	0.036	-0.009
MP2	0.067	0.040	-0.004
CCSD	0.068	0.020	0.008
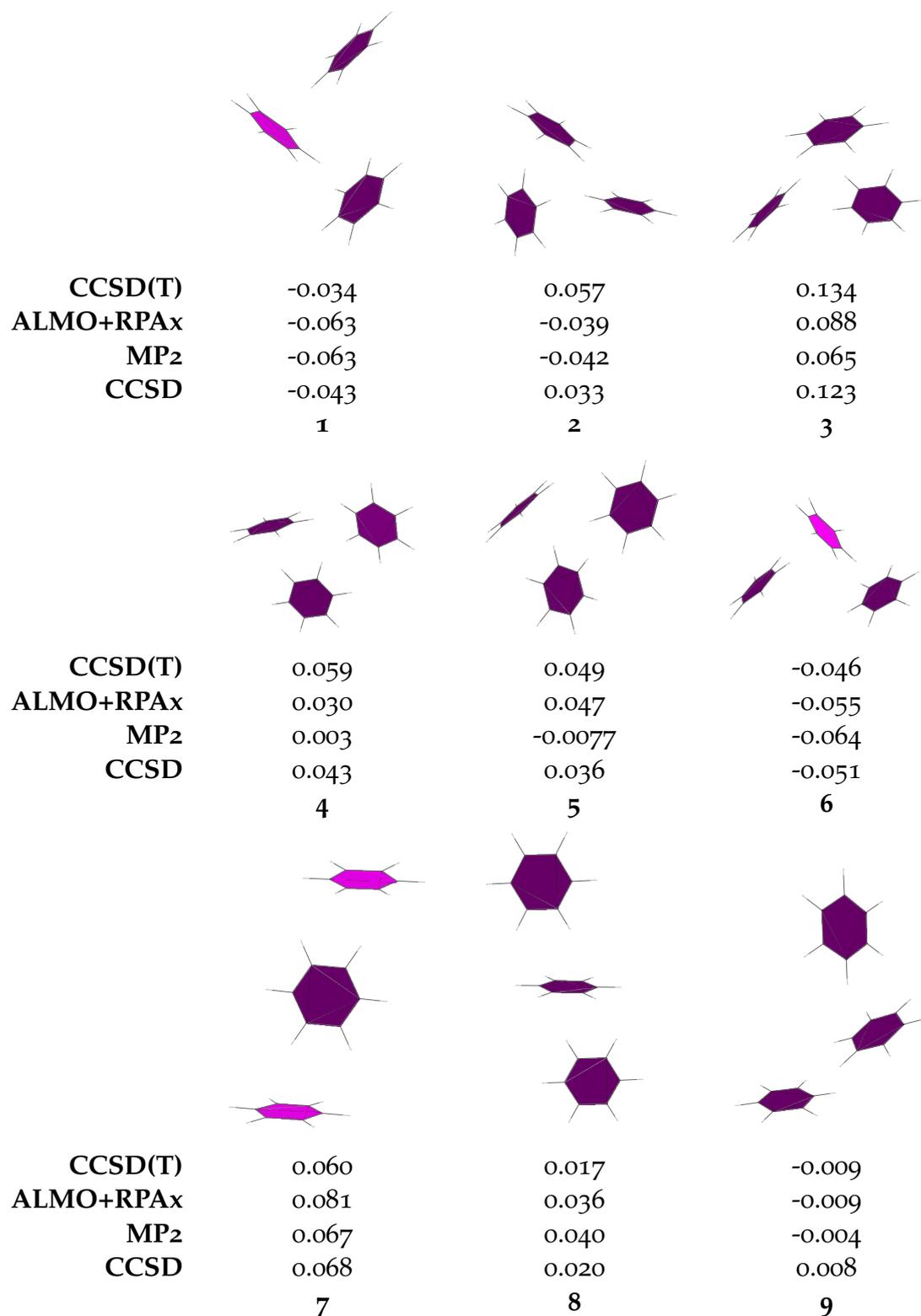
Figure 5.19: The three-body contributions to the interaction energy, in kcal/-mol, for 9 conformers of the benzene trimer. These were all calculated with the aug-cc-pVDZ basis set.
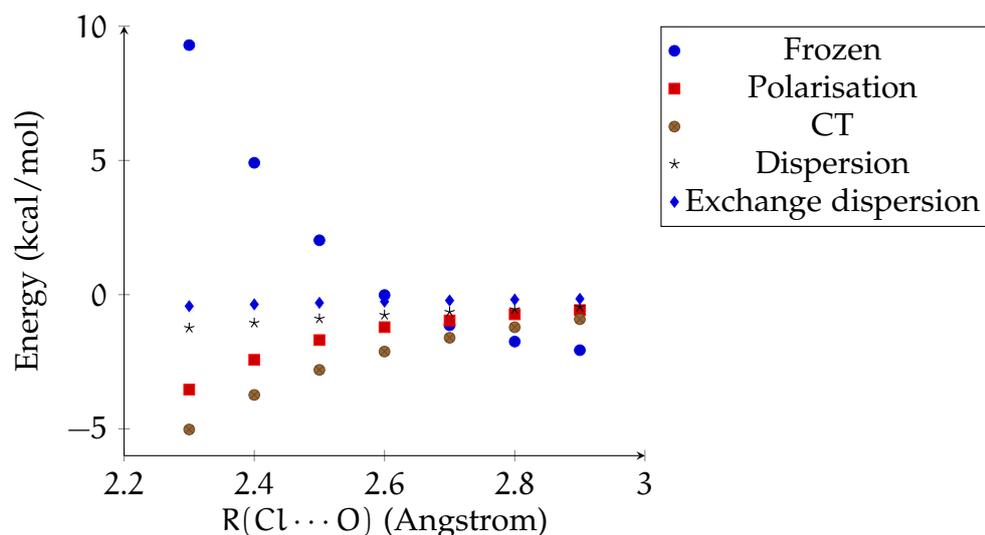
Figure 5.20: The ALMO+RPAxd energy decomposition analysis for the FCl···OH$_2$ complex. While electrostatics dominate at short separations, the interaction around equilibrium (approximately 2.65 Å) is dominated by the charge transfer term.

closer to the CCSD(T) result than CCSD, although this is likely due to fortunate error cancellation. The MAEs compared to CCSD(T) are 0.011, 0.025, and 0.036 kcal/mol, for CCSD, ALMO+RPAx, and MP2, respectively. This demonstrates that the new method does seem to accurately describe three-body interactions, with accuracy somewhere between CCSD and MP2, as is to be expected.

5.6 ENERGY DECOMPOSITION

One of the reasons for choosing the FCl···OH$_2$ complex for testing is that it has been demonstrated that MP2 and DFT perform poorly for these types of halogen-bonded system [12]. It has been suggested that this is due to the significant degree of charge transfer going from the water to the chlorine [12, 14, 15], in contrast to the assertion made by many that halogen bonds are predominantly electrostatic in nature [202, 337]. The energy decomposition for this system is shown in Figure 5.20. From this, it is clear that, while in the very short range the interaction is dominated by electrostatic repulsion, it is the charge transfer term that accounts for the majority of the interaction in the vicinity of the minimum. This is reflected in the correct prediction of the equilibrium bond distance, which from a sixth-order polynomial fit to the curve in Figure 5.15 is found at 2.65 Å. It should be noted that the CT and polarisation terms are not completely well-separated due to the need to project the virtuals out of the occupied space when
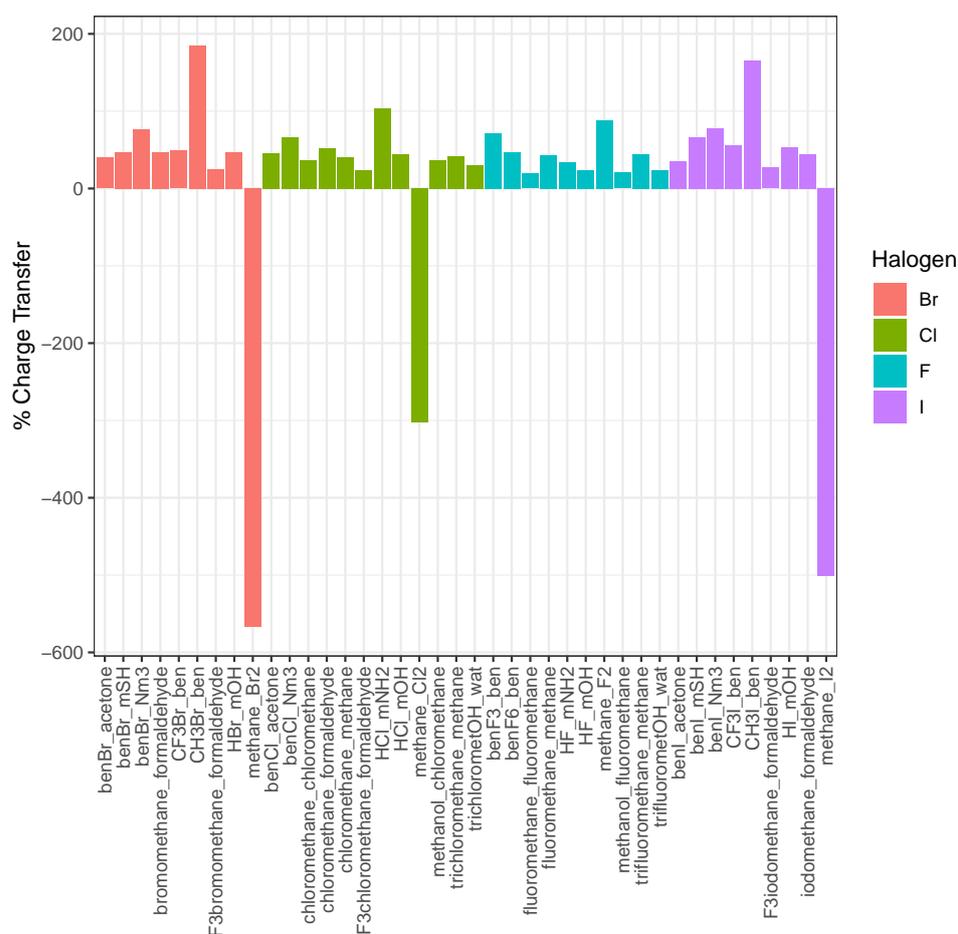
Figure 5.21: The ALMO charge transfer contribution, as a percentage of the total interaction energy, for all systems in the X40 set of halogen bonds [333], grouped by halogen bond donor.

calculating the former, and the fact that the two effects are physically concomitant.

Halogen-bonded systems in general form a good test for the charge transfer component. As well as the above system, in Figure 5.21, the charge transfer terms for the X40 dataset of halogen bonds are shown as a percentage of the total interaction energy. This shows a distinct trend towards larger components in the heavier atoms (bromine and iodine), with far smaller components in particular for fluorine. Of additional note is the interactions of methane with each dihalogen show a significant negative percentage. As the charge transfer energy is always negative, this is telling us that the total interaction is positive. It therefore seems unlikely that significant charge transfer is occurring. Inspection of the interaction energies, however, indicates that the charge transfer terms are small in absolute energy terms, but that the interaction is close to zero, hence the large percentage; these can be seen in Table 5.3. Indeed, one expects the interaction between a

Table 5.3: The charge transfer and dispersion contributions to the total interaction energy (all in kcal/mol) of three dihalogens with methane, from the X40 database.

	Cl_2	Br_2	I_2
CT	-0.266	-0.278	-0.291
Disp.	-0.089	-0.261	-0.331
Total	0.088	0.049	0.058
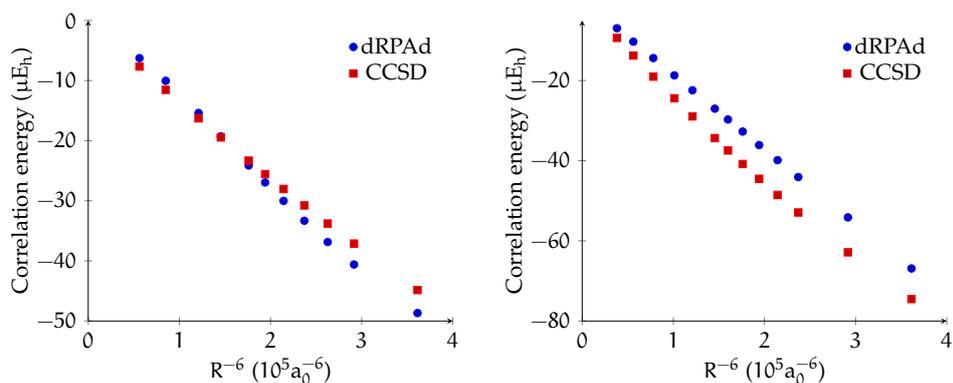
Figure 5.22: The correlation (dispersion) energy as calculated using direct RPAd and CCSD, for the helium-helium (left) and helium-neon (right) complexes, as a function of R^{-6}, where R is the separation. The correct linear behaviour is seen.

neutral dihalogen and methane to be largely dispersive, as there are no significant multipoles present. Again, there is some overlap between the dispersion and charge transfer terms, due to the projection of the virtuals. While this is fairly small, the charge transfer energies in these cases are small, leading to what seem like anomalies.

The dispersive contribution to the above, while vital to achieving agreement with the CCSD result, is not dominant. In the noble gas dimers, however, dispersion accounts for all of the attractive interaction, with the electrostatic contributions giving an entirely repulsive curve. The separations are large enough that the dominating factor is the R^{-6} dependence arising from the Casimir-Polder integral [10]. Whether the correct behaviour is seen for these systems therefore presents an excellent test of the validity of both the choice of RPA as the correlated method, and the decomposition into dispersive excitations described in Figure 3.4. Plots of the correlation energy versus R^{-6} are shown in Figure 5.22. For both dimers, the trend is clearly linear, demonstrating the correct physical behaviour. The C_6 coefficients can be crudely estimated from the gradient of these lines, yielding 1.37 and 2.91 a.u. for He_2 and He$\cdots$Ne, respectively. This is in comparison to the CCSD results of 1.29 and 2.72 a.u. The accepted literature values [203] are 1.46 and 3.08 a.u. However, we have only used
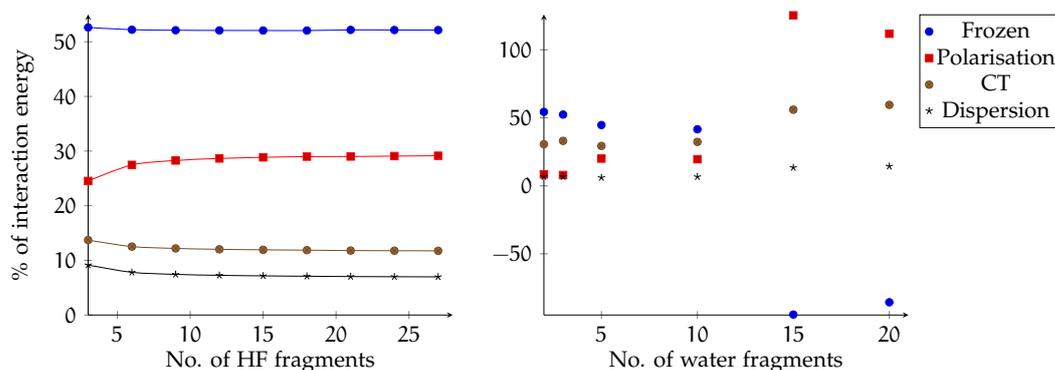
Figure 5.23: The ALMO+RPAxd energy decomposition analyses for chains of hydrogen fluoride molecules (left) and water clusters (right). The former demonstrates cooperativity between the hydrogen bonds, due to increased polarisation. Note that there is a drastic change in geometries going between water clusters of more or less than ten monomers, with a corresponding change in the energy composition.

very small basis sets here, so do not expect quantitative agreement. The key point is that the qualitative behaviour is correct and there is good agreement with the equivalent aVDZ coupled-cluster result; in the limit of much larger basis sets, it is likely that better results would be obtained, as per Figure 5.14.

Perhaps the most useful and novel facet of the new method, however, is the ability to perform an energy decomposition on large, many-body systems. In particular, it is possible to see how the different contributions vary with increasing system size. This is shown for the HF chains and water clusters in Figure 5.23. For the former, the interaction energy per hydrogen bond steadily decreases to a limiting value, in agreement with previous studies [338, 339]. The decomposition shows that this is primarily due to an increase in polarisation; as a new fragment is added, this polarises the fragment next to it, which in turn further polarises the next one along, and so on. This is a well-known example of cooperative hydrogen bonding, and essentially results in a molecular wire with an anode at the fluorine end and a cathode at the terminating hydrogen end.

The shuttling of electron density along the chain results in an increased cohesion. While for these calculations the separation between fragments and geometry within each molecule has been fixed, this suggests that if the system were allowed to relax, the monomers would get slightly closer together, accompanied by an anisotropic distribution of HF bond lengths along the chain, with those at the centre being the shortest.

The trend for the water clusters is less clear. This is in part due to the unsystematic selection of cluster geometries, taken from DFT

calculations for clusters up to ten fragments [340], and molecular dynamics simulations for those beyond [341]. The drastic change that appears to happen between these two regimes is therefore most likely an artefact of the different levels of description. Particularly noticeable is that the frozen energy becomes highly repulsive, suggesting that the water molecules in the larger clusters are too close together. This is counteracted by a large increase in the polarisation, with the interaction therefore being controlled by the charge transfer terms. In contrast, the smaller clusters are dominated by a purely electrostatic attraction, suggesting that the water molecules are essentially polarised dipoles. Overall, the two regions suggest that there are two distinct regimes of interaction - one that is a classical, electrostatic description at slightly longer separations, and one that is charge-transfer driven at shorter separations.

5.7 COMPARISON WITH SAPT

As discussed in the introduction, not many methods exist that eliminate BSSE *a priori*, and of those that do, only SAPT additionally provides an energy decomposition. This in particular has led to its widespread use in the computational chemistry community [85, 342], although this is not without controversy. Moreover, the SAPT2 variant contains a coupled cluster style correlation term for the dispersion, so that it is a similar level of theory to ALMO+RPA. While the physical terms in the decomposition are not necessarily well separated, we would expect that a comparison of decompositions should yield qualitatively similar results. Certainly, in terms of validating relative magnitudes of for example dispersion and charge transfer, we would ideally want both methods to agree.

The components are not directly comparable, however, as we have defined our decomposition slightly differently. Moreover, there are several different ways to group SAPT terms - we choose ours to match with Psi4 [328], the grouping of which is given in Appendix E. The most significant differences with our approach are that we do not separate exchange as its own term, while SAPT does not separate charge transfer from the induction. Thus, to compare sensibly, we look instead at the SAPT induction minus the charge transfer, and term this the polarisation in analogy to the ALMO EDA. Similarly, the frozen energy of ALMO is inherently the sum of the electrostatic and exchange components of the SAPT decomposition - we cannot meaningfully separate the two in our self-consistent scheme, as exchange is included directly in the ALMO density.

To compare, I have calculated the energy decompositions at both the ALMO+RPAx and DF-SAPT2 levels for the entire S66 and X40
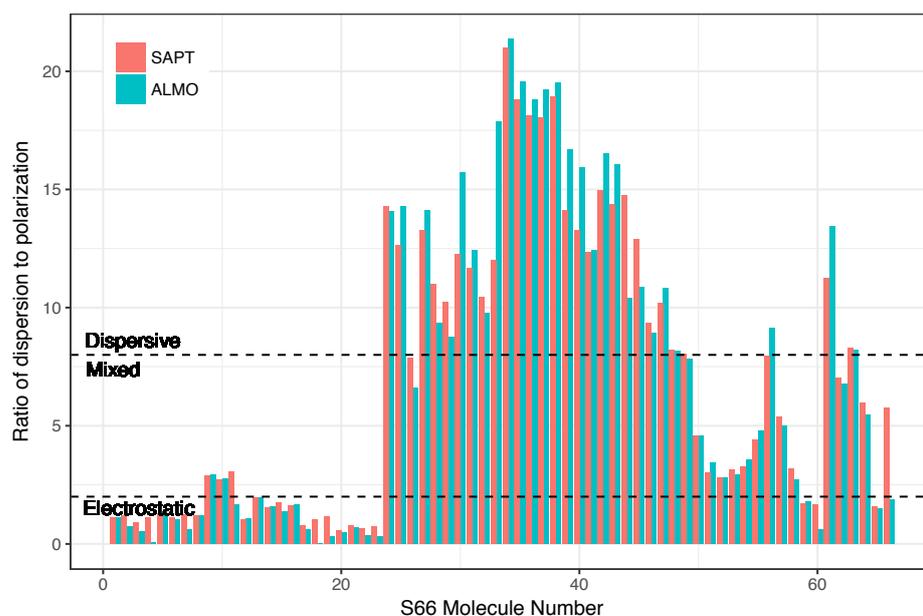
Figure 5.24: The ratio of dispersion to polarization contributions, $r_{D/P}$, in the interaction energies of the complexes in the S66 database. Results are compared for the SAPT (red, left bar) and ALMO (blue, right bar) energy decompositions, with the cutoffs for classification shown as dashed lines. The molecule number is as given in the original database[267].

datasets, using the aVTZ basis [343], or for bromine and iodine, the aVTZ-PP basis with ECP10MDF and ECP28MDF effective core potentials [344, 345], respectively, with matching auxiliary sets [151, 325, 346]. All the data from this, and from the earlier benchmark calculations, can be found in Appendix E.

The S66 set is particularly informative, as the original paper classified the complexes into three types based on the SAPT decomposition, specifically the ratio of dispersion to induction [267]. With our definition of polarisation, we can recover these categories exactly by considering instead the ratio of dispersion to polarisation, $r_{D/P}$, yielding: electrostatic if $r_{D/P} < 2$; mixed if $2 \leqslant r_{D/P} < 8$; and dispersive if $r_{D/P} \geqslant 8$.

Figure 5.24 shows a side-by-side comparison of the two decompositions. This shows that, of the 66 systems, only four are not in qualitative agreement, and these systems are shown in Figure 5.25, along with their respective ratios and energy contributions. For the first two, there is clearly no real difference in classification; rather, the ratios lie exactly on the borderline between mixed and dispersive. The dispersion energies from SAPT are larger than those from the RPA, as described earlier, but the polarisation terms are proportionately larger too. For the second two systems, MeNH$_2$ with pyridine and peptide, the SAPT dispersion energy is so much larger that there is

$r_{D/P}$: 7.94 / 9.15
Disp: -5.12 / -3.57 kcal/mol
Pol.: -0.65 / -0.39 kcal/mol

8.05 / 7.84
-5.45 / -3.46 kcal/mol
-0.68 / -0.44 kcal/mol

$r_{D/P}$: 3.05 / 1.66
Disp: -4.97 / -2.07 kcal/mol
Pol.: -1.63 / -1.25 kcal/mol

5.76 / 1.86
-5.26 / -1.10 kcal/mol
-0.91 / -0.59 kcal/mol
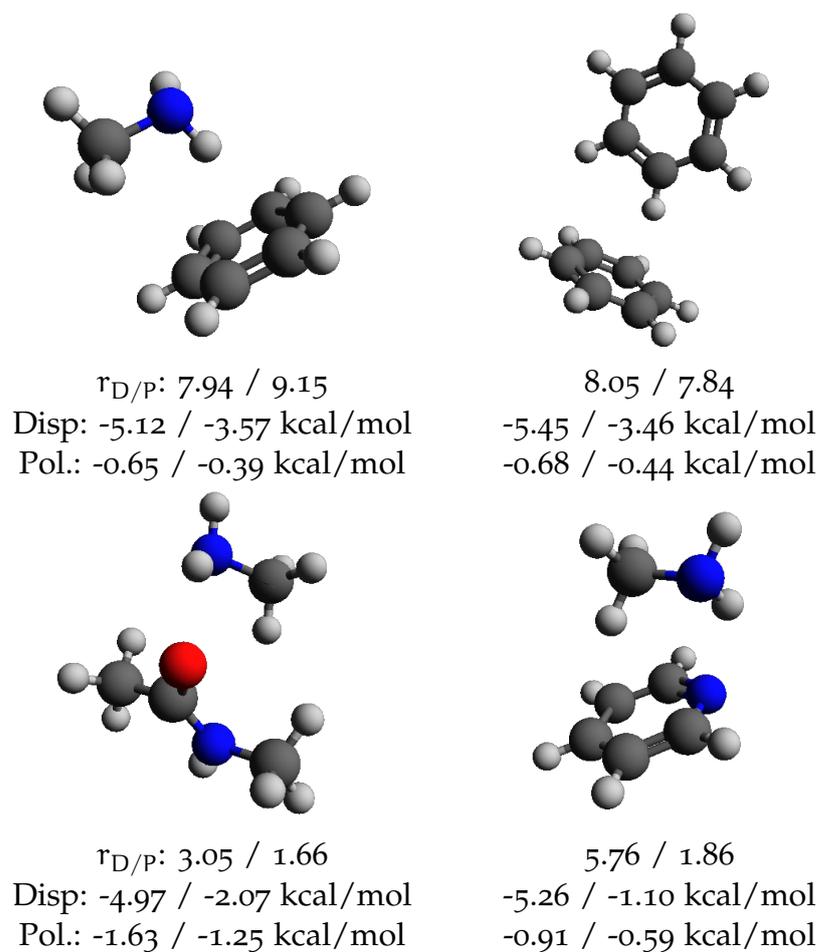
Figure 5.25: The four S66 systems for which the SAPT2 and ALMO+RPAx dispersion to polarisation ratio classifications did not agree, along with the ratios and energy values for each, given as SAPT2/ALMO+RPAx.
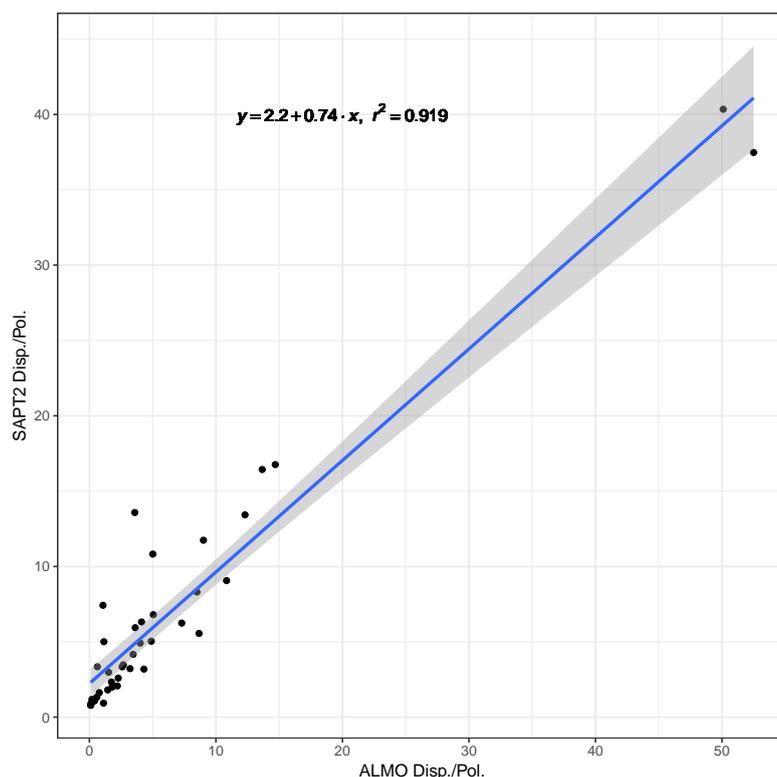
Figure 5.26: The ratio of dispersion to polarisation in the interaction energies of the X40 dataset, as calculated using both SAPT2 and ALMO+RPAxd. There is qualitative agreement between the two methods, lending confidence to the decomposition process.

a signficant difference in the ratios: ALMO classifies them as electrostatic, whereas SAPT as mixed. Comparing the total interaction energy to the complete basis set limit CCSD result, we see that SAPT overestimates both by about 0.5 kcal/mol, while ALMO+RPAx underestimates by a similar amount. From the figure, it is clear that there are cooperative interactions happening, in particular N to H hydrogen bonds in combination with H···C or H···π dispersive interactions; the balance between electrostatics and dispersion is thus very fine, and entirely dependent on the weighting of the generally weaker dispersive interaction.

The X40 database does not contain any equivalent classifications to compare, but its systems in general contain greater polarisation contributions than those of the S66 database. A similar comparison of $r_{D/P}$ ratios is given in Figure 5.26, where once again we see good qualitative agreement. The error distributions from figures 5.16 and 5.18 suggest that, of the two methods, the ALMO+RPA dispersion contributions are more accurate, and so the lack of quantitative agreement is perhaps not surprising. As most of the points in Figure 5.26 lie above the line, this agrees with the observation that SAPT2 tends to

overestimate the dispersion, leading to larger $r_{D/P}$ values than those from ALMO+RPA.

6 | ANALYTICAL DERIVATIVES

In this chapter, I derive analytical gradients for the ALMO-based methods described in previous chapters. These are necessary for the optimisation of molecular geometries, and for the calculation of properties other than the energy. By formulating the problem in terms of Lagrangians, we can systematically determine the derivatives for each separate part of the calculation, taking into account, for example, the different charge transfer and exchange terms. I then show how to adapt these expressions to include the density-fitting and local approximations, before finally demonstrating the effectiveness of the method on optimisations of large systems.

Many of the most interesting chemical properties of molecular systems - including supermolecular systems - make the assumption that the system is in a local minimum, or equilibrium. For example, to calculate the normal mode vibrational frequencies, one takes an equilibrium geometry and determines the second derivatives of the energy at that point. Moreover, many of these properties, such as multipole moments, follow directly from the forces [347]. As such, it is essential to be able to accurately and rapidly optimise molecular structures.

It is possible in principle to optimise based on the energy from any method simply by estimating the forces by numerical methods, in particular, finite differences [348]. This involves fractionally perturbing the relevant parameters - in this case the nuclear coordinates, represented here as q_i for the ith nucleus - then considering the difference in energy this results in; essentially, it is approximating the potential energy curve along that coordinate by a small line segment. It is possible to do this by forward (or backward) differences, where you add (or subtract) a small quantity h to a coordinate, then estimate the gradient at that point by, e.g. $[E(q + h) - E(q)]/h$. There are many problems associated with this, however, in particular the choice of h. A very large h will be inaccurate as the line segment is a poor approximation to the true curve, whereas a very small h will result in both a small ΔE and a large $1/h$, causing numerical problems. Moreover, the greater the curvature at the original point, the smaller h needs to be to give accurate results, such that a fixed h can lead to variable accuracy [348].

Some of these problems can be mitigated by instead using central differences:

$$\frac{\partial E}{\partial q_i} \approx \frac{E(q_i + h; \{q_{j \neq i}\}) - E(q_i - h; \{q_{j \neq i}\})}{2h} \tag{6.1}$$

The problem with this however is that each gradient then takes two additional energy calculations. Thus for a system of N atoms, 6N energies are needed on top of the original calculation. This means that numerical optimisation very rapidly becomes unfeasible as the system gets larger, or the method more expensive. As such it is essential to be able to calculate the gradients directly. The finite differences in practice are only useful for checking that the analytical gradients have been calculated correctly.

When specifically looking at intermolecular interactions, we have the additional problem that the gradient, like any quantity derived from the energy, is tainted with basis set superposition errors. Counterpoise correcting the gradients is considerably more difficult than just the energy [349], by virtue of the sheer number of terms that need to be computed. Combined with numerical differentiation as above, the problem explodes to order N^2 calculations needed at each step. The ALMO+RPA energy, and therefore gradient, is inherently BSSE-free, thus offering it an additional advantage over traditional methods. This comes at the expense of only being applicable to systems of fragments, which will necessarily have large numbers of degrees of freedom, making numerical differentiation impractical.

Analytically differentiating the energy with respect to a single parameter is difficult, however. For example, the canonical Hartree-Fock energy depends both on the integral matrices and on the MO coefficients, all of which depend on the position. The latter have been determined iteratively, with no closed functional form available, making detailed analysis close to impossible. In this particular case, we can make use of the variational nature of the method to circumvent this method, as described by Pople in the early '70s [350]. In general, though, methods are not variational. In these cases, we can instead use a Lagrangian formalism [347] as follows.

Consider an energy function which is a function of the nuclear coordinates, $\mathbf{q}$, parametrised with respect to some set of parameters $\{\tau_p\}$, and some set of constraints $\{Z_i = 0\}$ which themselves may depend on some subset of the set of coordinates and parameters. We then form a Lagrangian by introducing a set of constants $\{\lambda_i\}$ such that

$$\mathcal{L}\left(\mathbf{q}; \{\tau_p\}, \{\lambda_i\}\right) = E\left(\mathbf{q}; \{\tau_p\}\right) - \sum_i \lambda_i Z_i\left(\mathbf{q}; \{\tau_p\}\right)$$

Minimising the Lagrangian then minimises the energy but with the constraints maintained, as we have done in previous chapters.

We assume in the following that the coordinates are *uncoupled*, as would be the case in any orthogonal coordinate system. The total derivative of the energy with respect to a given nuclear coordinate q is then, by the chain rule:

$$\frac{dE}{dq} = \frac{\partial \mathcal{L}}{\partial q} + \sum_p \frac{\partial \mathcal{L}}{\partial \tau_p} \frac{d\tau_p}{dq} + \sum_i \frac{\partial \mathcal{L}}{\partial \lambda_i} \frac{d\lambda_i}{dq} + \frac{d}{dq} \sum_i \lambda_i Z_i\left(\mathbf{q}; \{\tau_p\}\right) \quad (6.2)$$

Assuming that the constraints are strictly maintained, the last two terms in the above disappear. If we then require that

$$\frac{\partial \mathcal{L}}{\partial \tau_p} = 0 \tag{6.3}$$

All constraints in this thesis are hard as opposed to soft, thus are strictly required to be everywhere zero.

for all parameters τ_p, then we simply get that $E^{(q)} = \partial\mathcal{L}/\partial q$. In the case of a variational method, this condition will necessarily and automatically be true for any variationally optimised set of parameters. In general, however, equation 6.3 leads to sets of coupled-perturbed equations that need to be solved for the multipliers, λ_i. These can then be substituted into equation 6.2 to give the analytical gradient.

In the rest of this chapter, I will use this approach to derive the gradients of the *electronic* energy with respect to a nuclear coordinate for all flavours of the ALMO+RPA method, before implementing and testing the solutions. It should be noted that the gradients of the SCF-MI variant of the ALMO SCF energy were previously derived by Famulari *et al.* [94], using a different approach to ours. Finally, the total force in that coordinate also includes a term due to the nuclear-nuclear repulsion, which is common to all methods under the clamped-nucleus approximation:

$$E_{\text{nuc.}} = \frac{1}{2} \sum_A \sum_{B \neq A} \frac{Z_A Z_B}{|\mathbf{q}_A - \mathbf{q}_B|}$$

$$\frac{dE_{\text{nuc.}}}{dq_A^{(i)}} = -\frac{1}{2} \sum_{B \neq A} \frac{Z_A Z_B \left(q_A^{(i)} - q_B^{(i)}\right)}{|\mathbf{q}_A - \mathbf{q}_B|^3} \tag{6.4}$$

This easily calculated contribution is then added to the end of the electronic gradient to give the total force.

6.1 ALMO

The Lagrangian for the ALMO SCF method is given in equation 2.46, and the only parameters it depends on are the occupied MO coefficients and the Lagrange multipliers. However, the former are varia-

tionally optimised, and as noted above, this means that no additional calculations are required. That is, $\partial E_{ALMO}/\partial T_{\mu i}$ is zero and thus

$$\frac{\partial \mathcal{L}_{ALMO}}{\partial T_{Xi}^{X\mu}} = \frac{\partial E_{ALMO}}{\partial T_{Xi}^{X\mu}} - 4 \left[\boldsymbol{\lambda} \mathbf{T}^\dagger \mathbf{S} \right]_{Xi,X\mu} = 0$$

implies that $\boldsymbol{\lambda} = 0$, removing the need to consider this term.

Similarly, the constraint that the inter-fragment coefficients be zero is maintained by the multipliers, $\boldsymbol{\gamma}$, being held to zero. In fact, none of the other terms throughout this derivation will involve inter-fragment coefficients, such that these multipliers can always be assumed to be zero; the Lagrangian constraint on them will however be relevant in Chapter 7, hence the need to include it in equation 2.46. Therefore we simply have that the gradient is given by

$$
\begin{aligned}
E_{ALMO}^{(q)} &= \frac{\partial E_{ALMO}}{\partial q} \\
&= \frac{\partial}{\partial q} \left\{ \frac{1}{2} \mathrm{Tr} \{ (\mathbf{H} + \mathbf{F}) \mathbf{P} \} \right\} \\
&= \frac{1}{2} \mathrm{Tr} \left\{ \left(\mathbf{H}^{(q)} + \mathbf{F}^{(q)} \right) \mathbf{P} + (\mathbf{H} + \mathbf{F}) \mathbf{P}^{(q)} \right\}
\end{aligned}
$$

The AO-basis matrices depend only on the nuclear position through the basis functions themselves, and algorithms for the calculation of these derivatives are well known and fairly easily calculated [351, 352]. As such, we wish to cast the gradient entirely in terms of these derivative matrices where possible. To do so, we need to see how both the Fock and density matrices depend parametrically on the nuclear coordinates. Using equations 2.45 and 2.30:

$$\mathbf{F}^{(q)} = \mathbf{H}^{(q)} + \mathcal{G}^{(q)} \cdot \mathbf{P} + \mathcal{G} \cdot \mathbf{P}^{(q)} \tag{6.5}$$

$$\mathbf{P}^{(q)} = \mathbf{T} \left[\sigma^{-1} \right]^{(q)} \mathbf{T}^\dagger = -\mathbf{T} \left[\sigma^{-1} \sigma^{(q)} \sigma^{-1} \right] \mathbf{T}^\dagger = -\mathbf{P} \mathbf{S}^{(q)} \mathbf{P} \tag{6.6}$$

where $\mathcal{G}$ is the tensor of antisymmetrised two-electron integrals, and we have used the identity $\delta\sigma^{-1} = -\sigma^{-1}(\delta\sigma)\sigma^{-1}$.

Using these and collecting all the terms in the above gives

$$E_{ALMO}^{(q)} = \frac{1}{2} \mathrm{Tr} \left\{ (2\mathbf{H}^{(q)} + \mathcal{G}^{(q)}) \mathbf{P} - (\mathbf{H} + \mathbf{F} + \mathcal{G} \cdot \mathbf{P}) \mathbf{P} \mathbf{S}^{(q)} \mathbf{P} \right\}$$

Then, collecting the derivative matrices and using the permutational invariance of traces, we get the final ALMO gradient:

$$E_{ALMO}^{(q)} = \mathrm{Tr} \left\{ \left(\mathbf{H}^{(q)} + \frac{1}{2} \mathcal{G}^{(q)} \right) \mathbf{P} - \mathbf{P} \mathbf{F} \mathbf{P} \mathbf{S}^{(q)} \right\} \tag{6.7}$$

6.2 CHARGE TRANSFER CORRECTION

The perturbative charge transfer correction to the ALMO energy is not variational, such that the Lagrangian formulation is vital. We will focus on the infinite-order correction of equation 3.14; as was demonstrated in Chapter 3, this reduces to the second-order correction under certain conditions, such that its gradient is also a subset of the infinite-order gradient.

There are two new considerations for this term: the virtual orbitals have been projected out of the occupied subspace, and the parameters $\mathbf{x}_{VO}$ have been determined iteratively. The latter means that a new constraint must be added, namely that the residual of equation 3.15 must remain zero. The former means we must expand both this residual and the energy term in such a way that the virtual projection is explicit, and thus maintained. The set of parameters in the Lagrangian, excluding the Lagrange multipliers, therefore expands from just the occupied coefficients to the following:

$$\left\{ T_i^\mu, V_a^\mu, x_{ai} \right\}$$

We introduce new Lagrange multipliers, ω_{ia}, which control the CT residual, and thus we arrive at the ALMO+CT Lagrangian:

$$
\begin{aligned}
\mathcal{L}_{\text{ALMO+CT}} =& \mathcal{L}_{\text{ALMO}}\left(\mathbf{q}; \mathbf{T}, \mathbf{V}, \boldsymbol{\lambda} \right) + E_{\text{CT}}\left(\mathbf{q}; \mathbf{T}, \mathbf{V}, \mathbf{x} \right) \\
& - \omega_{ia} \left[\mathbf{R}_{\text{CT}}\left(\mathbf{q}; \mathbf{T}, \mathbf{V}, \mathbf{x} \right) \right]_{ai}
\end{aligned}
\tag{6.8}
$$

where E_{CT} and $\mathbf{R}_{\text{CT}}$ are defined in equations 3.67 and 3.68, respectively.

6.2.1 The residual amplitudes

The original ALMO energy does not depend upon the residual amplitudes, x_{ai}. Therefore, the partial derivative of the Lagrangian with respect to these is just the partial derivative of the last two terms in equation 6.8. These are quite simply found as follows:

$$\frac{\partial E_{\text{CT}}}{\partial x_{ai}} = 2 \left[\mathbf{T}^\dagger \mathbf{F} \tilde{\mathbf{V}} \right]_{ia}$$

$$\frac{\partial \left[R_{\text{CT}} \right]_{bj}}{\partial x_{ai}} = \delta_{ij} \left[\left(\tilde{\mathbf{V}}^\dagger - \mathbf{x}\mathbf{T}^\dagger \right) \mathbf{F} \tilde{\mathbf{V}} \right]_{ba} - \delta_{ab} \left[\mathbf{T}^\dagger \mathbf{F} \left(\tilde{\mathbf{V}}\mathbf{x} + \mathbf{T} \right) \right]_{ij}$$

where

$$\tilde{\mathbf{V}} = \mathbf{Q}\mathbf{V} = \left(\mathbf{I} - \mathbf{S}\mathbf{P} \right) \mathbf{V}$$

are the projected virtual coefficients.

As per the variational condition, equation 6.3, the above combined with the relevant Lagrange multipliers yields the requirement below.

$$\frac{\partial \mathcal{L}_{\text{ALMO+CT}}}{\partial x_{ai}} = \left[2\mathbf{T}^\dagger \mathbf{F} \tilde{\mathbf{V}} + \boldsymbol{\omega} \left(\tilde{\mathbf{V}}^\dagger - \mathbf{x}\mathbf{T}^\dagger \right) \mathbf{F} \tilde{\mathbf{V}} - \mathbf{T}^\dagger \mathbf{F} \left(\tilde{\mathbf{V}}\mathbf{x} + \mathbf{T} \right) \boldsymbol{\omega} \right]_{ia} = 0$$

$$\tag{6.9}$$

for all virtual-occupied pairs, ai. This is an equation of the form

$$\boldsymbol{\omega}\mathbf{A} - \mathbf{B}\boldsymbol{\omega} = -\mathbf{C}$$

where the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, implicitly defined above, are known quantities. Equation 6.9 can thus be solved for the multipliers ω_{ia}, ensuring that the variational conditions are met.

6.2.2 The virtual coefficients

The ALMO energy necessarily only depends on the occupied co-efficients. Every other part of the Lagrangian, however, now has a dependence on the virtual coefficients that cannot be ignored. This is complicated in the CT portions by the projection, $\mathbf{Q}$, which itself depends only on the occupied coefficients, which we will deal with in the next section. The derivatives of the charge transfer terms are as follows:

$$\frac{\partial E_{CT}}{\partial V_{Xa}^{X\mu}} = 2\left[\mathbf{x}\mathbf{T}^{\dagger}\mathbf{F}\mathbf{Q}\right]_{Xa,X\mu}$$

$$\frac{\partial\left[R_{CT}\right]_{Yb,Zj}}{\partial V_{Xa}^{X\mu}} = \delta_{Xa,Yb}\left[\mathbf{Q}^{\dagger}\mathbf{F}\left(\mathbf{T} + \mathbf{Q}\mathbf{V}\mathbf{x}\right)\right]_{X\mu,Zj}$$

$$+ \left[\left(\mathbf{V}^{\dagger}\mathbf{Q}^{\dagger} - \mathbf{x}\mathbf{T}^{\dagger}\right)\mathbf{F}\mathbf{Q}\right]_{Yb,X\mu} x_{Xa,Zj}$$

For simplicity, we define the following two quantities, which will appear repeatedly throughout this chapter:

$$\begin{aligned}\boldsymbol{\Pi}_{+} &= \mathbf{T} + \mathbf{Q}\mathbf{V}\mathbf{x} \\ \boldsymbol{\Pi}_{-} &= \mathbf{V}^{\dagger}\mathbf{Q}^{\dagger} - \mathbf{x}\mathbf{T}^{\dagger}\end{aligned} \tag{6.10}$$

We note that this turns the residual equation 3.68 into

$$\mathbf{R}_{CT} = \boldsymbol{\Pi}_{-}\mathbf{F}\boldsymbol{\Pi}_{+} = 0$$

Combining the above with the orthonormality condition and the multipliers ω, found from the solution of equation 6.9, we get the following condition on the parameters λ_{ab}:

$$\begin{aligned}\frac{\partial \mathcal{L}_{ALMO+CT}}{\partial V_{Xa}^{X\mu}} &= -4\left[\boldsymbol{\lambda}\mathbf{V}^{\dagger}\mathbf{S}\right]_{Xa,X\mu} + 2\left[\mathbf{x}\mathbf{T}^{\dagger}\mathbf{F}\mathbf{Q}\right]_{Xa,X\mu} \\ &+ \left[\mathbf{Q}^{\dagger}\mathbf{F}\left[\boldsymbol{\Pi}_{+}\boldsymbol{\omega} + \boldsymbol{\Pi}_{-}^{\dagger}\boldsymbol{\omega}^{\dagger}\mathbf{x}^{\dagger}\right]\right]_{X\mu,Xa} = 0\end{aligned} \tag{6.11}$$

This then rearranges into a simple equation for the virtual orbital energies:

$$\boldsymbol{\lambda}_{VV}\mathbf{V}^{\dagger}\mathbf{S} = \frac{1}{2}\mathbf{x}\mathbf{T}^{\dagger}\mathbf{F}\mathbf{Q} + \frac{1}{4}\left[\left[\boldsymbol{\omega}^{\dagger}\boldsymbol{\Pi}_{+}^{\dagger} + \mathbf{x}\boldsymbol{\omega}\boldsymbol{\Pi}_{-}\right]\mathbf{F}\mathbf{Q}\right] = \boldsymbol{\Lambda}_{VV}^{CT} \tag{6.12}$$

which is a linear equation that can be easily be solved by standard methods.

6.2.3 The occupied coefficients

All terms in equation 6.8 depend on the occupied coefficients. However, the ALMO energy is variationally optimised with respect to them, so its partial derivative evaluates to zero. Thus, as with the virtual coefficients, the only new terms are the charge transfer ones. However, as noted earlier, the projector $\mathbf{Q}$ introduces a hidden dependence. We have that

$$\frac{\partial Q_{Yv,Z\tau}}{\partial T_{Xi}^{X\mu}} = \frac{\partial}{\partial T_{Xi}^{X\mu}} \left[\mathbf{I} - \mathbf{SP}\right]_{Yv,Z\tau} = -\sum_{W} S_{Yv,W\gamma} \frac{\partial P^{W\gamma,Z\tau}}{\partial T_{Xi}^{X\mu}}$$

Using the result from equation 2.47 for the density variation, the above then becomes

$$-\frac{1}{2}\frac{\partial Q_{Yv,Z\tau}}{\partial T_{Xi}^{X\mu}} = \left[\mathbf{SQ}^{\dagger}\right]_{Yv,X\mu} \left[\mathbf{\sigma}^{-1}\mathbf{T}\right]_{Xi,Z\tau} + Q_{X\mu,Z\tau} \left[\mathbf{ST}^{\dagger}\mathbf{\sigma}^{-1}\right]_{Yv,Xi} \quad (6.13)$$

The above can now be used to determine the energy and residual derivatives. We start with the former:

$$\frac{1}{4}\frac{\partial E_{CT}}{\partial T_{Xi}^{X\mu}} = \frac{1}{2}\left[\mathbf{FQVx}\right]_{X\mu,Xi} + \frac{1}{2}\sum_{Y,Z}\left[\mathbf{VxT}^{\dagger}\mathbf{F}\right]_{Z\tau,Yv}\frac{\partial Q_{Yv,Z\tau}}{\partial T_{Xi}^{X\mu}}$$

$$= \left[\frac{1}{2}\mathbf{FQVx} - \mathbf{QSFTx}^{\dagger}\mathbf{V}^{\dagger}\mathbf{T}^{\dagger}\mathbf{\sigma}^{-1} - \mathbf{QVxT}^{\dagger}\mathbf{FST}^{\dagger}\mathbf{\sigma}^{-1}\right]_{X\mu,Xi} = A_{X\mu,Xi}$$

where we have defined the temporary matrix, $\mathbf{A}$, for simplicity later. The equivalent for the residuals is messy, so we jump straight to the answer:

$$\frac{\partial\left[R_{CT}\right]_{Yb,Zj}}{\partial T_{Xi}^{X\mu}} = \delta_{Xi,Zj}\left[\mathbf{\Pi}_{-}\mathbf{F}\right]_{Yb,X\mu} - \left[\mathbf{F\Pi}_{+}\right]_{X\mu,Zj}x_{Yb,Xi}$$

$$+ \sum_{U,W}\left\{\left[\mathbf{\Pi}_{-}\mathbf{F}\right]_{Yb,Uv}\left[\mathbf{Vx}\right]_{W\kappa,Zj} + \left[\mathbf{F\Pi}_{+}\right]_{Uv,Zj}V_{Yb}^{W\kappa}\right\}\frac{\partial Q_{Uv,W\kappa}}{\partial T_{Xi}^{X\mu}}$$

$$= \delta_{Xi,Zj}\left[\mathbf{\Pi}_{-}\mathbf{F}\right]_{Yb,X\mu} - \left[\mathbf{F\Pi}_{+}\right]_{X\mu,Zj}x_{Yb,Xi}$$

$$- 2\left[\mathbf{\Pi}_{-}\mathbf{FSQ}^{\dagger}\right]_{Yb,X\mu}\left[\mathbf{\sigma}^{-1}\mathbf{TVx}\right]_{Xi,Zj}$$

$$- 2\left[\mathbf{V}^{\dagger}\mathbf{Q}^{\dagger}\right]_{Yb,X\mu}\left[\mathbf{\sigma}^{-1}\mathbf{TSF\Pi}_{+}\right]_{Xi,Zj}$$

Again, these are entirely in terms of known quantities. Combining all of the above with the Lagrange multipliers then gives us an equation for the occupied orbital energies:

$$\lambda_{OO}\mathbf{T}^{\dagger}\mathbf{S} = 2\mathbf{A} + \frac{1}{2}\left(\mathbf{\omega\Pi}_{-} - \mathbf{x}^{\dagger}\mathbf{\omega}^{\dagger}\mathbf{\Pi}_{+}^{\dagger}\right)\mathbf{F}$$

$$- \mathbf{\sigma}^{-1}\mathbf{T}\left(\mathbf{Vx\omega\Pi}_{-}\mathbf{FS} + \mathbf{SF\Pi}_{+}\mathbf{\omega V}^{\dagger}\right)\mathbf{Q}^{\dagger} = \mathbf{\Lambda}_{OO}^{CT} \quad (6.14)$$

6.2.4 Total derivative

Assuming that equations 6.9, 6.12, and 6.14 have been solved for the Lagrange multipliers, we can now construct the total derivative with respect to the nuclear coordinates. The derivative with respect to the ALMO Lagrangian is the same as in equation 6.7, except now the term involving $\boldsymbol{\lambda}$ does not necessarily disappear, and instead provides a contribution of

$$-2\mathrm{Tr}\left\{\left(\mathbf{T}\boldsymbol{\lambda}_{OO}\mathbf{T}^{\dagger} + \mathbf{V}\boldsymbol{\lambda}_{VV}\mathbf{V}^{\dagger}\right)\mathbf{S}^{(q)}\right\}$$

Both the charge transfer terms have non-parametric dependencies on the nuclear coordinates through the AO Fock matrix, the derivative of which was already determined in equation 6.5. However, they also depend on $\mathbf{Q}$, which itself contains both the overlap and density matrices. Its derivative is thus, using equation 6.6:

$$\mathbf{Q}^{(q)} = -\mathbf{S}^{(q)}\mathbf{P} - \mathbf{SP}^{(q)} = -\left(\mathbf{I} - \mathbf{SP}\right)\mathbf{S}^{(q)}\mathbf{P} = -\mathbf{QS}^{(q)}\mathbf{P} \qquad (6.15)$$

Using this, we can find the derivatives of the charge transfer energy and residual.

$$E_{CT}^{(q)} = 2\mathrm{Tr}\left\{\mathbf{T}^{\dagger}\left(\mathbf{F}^{(q)}\mathbf{Q} + \mathbf{FQ}^{(q)}\right)\mathbf{Vx}\right\}$$

$$\mathbf{R}_{CT}^{(q)} = \boldsymbol{\Pi}_{-}\mathbf{F}^{(q)}\boldsymbol{\Pi}_{+} + \mathbf{V}^{\dagger}\mathbf{Q}^{(q)\dagger}\mathbf{F}\boldsymbol{\Pi}_{+} + \boldsymbol{\Pi}_{-}\mathbf{FQ}^{(q)}\mathbf{Vx}$$

We can arrange the derivative into a more amenable form by first collecting all terms involving $\mathbf{F}^{(q)}$ and $\mathbf{Q}^{(q)}$ and then inserting equations 6.5 and 6.15. This gives, respectively,

$$\mathrm{Tr}\left\{\left(2\mathbf{QVxT}^{\dagger} - \boldsymbol{\Pi}_{+}\boldsymbol{\omega}\boldsymbol{\Pi}_{-}\right)\mathbf{F}^{(q)}\right\} = \mathrm{Tr}\left\{\boldsymbol{\Omega}_{1}\left(\mathbf{H}^{(q)} + \mathcal{G}^{(q)}\cdot\mathbf{P}\right)\right\}$$
$$- \mathrm{Tr}\left\{\mathbf{P}\boldsymbol{\Omega}_{1}\cdot\mathcal{G}\cdot\mathbf{PS}^{(q)}\right\}$$

$$\mathrm{Tr}\left\{\mathbf{V}\left(2\mathbf{xT}^{\dagger} - \boldsymbol{\omega}^{\dagger}\boldsymbol{\Pi}_{+}^{\dagger} - \mathbf{x}\boldsymbol{\Pi}_{-}\right)\mathbf{FQ}^{(q)}\right\} = \mathrm{Tr}\left\{\mathbf{PV}\boldsymbol{\Omega}_{2}\mathbf{FQS}^{(q)}\right\}$$

where we have repeatedly used the permutational invariance of traces, and defined the auxiliary matrices:

$$\begin{aligned}\boldsymbol{\Omega}_{1} &= 2\mathbf{QVxT}^{\dagger} - \boldsymbol{\Pi}_{+}\boldsymbol{\omega}\boldsymbol{\Pi}_{-} \\ \boldsymbol{\Omega}_{2} &= \boldsymbol{\omega}^{\dagger}\boldsymbol{\Pi}_{+}^{\dagger} + \mathbf{x}\boldsymbol{\Pi}_{-} - 2\mathbf{xT}^{\dagger}\end{aligned} \qquad (6.16)$$

All of the terms now divide into those involving three different sets of integral derivatives: $\mathbf{H}^{(q)}$, $\mathcal{G}^{(q)}$, and $\mathbf{S}^{(q)}$. These are termed the one-body, two-body, and Pulay forces [353], respectively. We can expose the form of the analytical gradients by defining three intermediate quantities associated with each category of force.

$$\mathbf{P}_{\mathrm{ALMO+CT}}^{(1)} = \mathbf{P} + \boldsymbol{\Omega}_{1} \qquad (6.17)$$

$$\mathbf{P}_{\mathrm{ALMO+CT}}^{(2)} = \mathbf{P}\left(\frac{1}{2} + \boldsymbol{\Omega}_{1}\right) \qquad (6.18)$$

$$\mathbf{F}_{\mathrm{ALMO+CT}}^{(S)} = 2\mathbf{C}\boldsymbol{\lambda}\mathbf{C}^{\dagger} + \mathbf{P}\left(\mathbf{F} + \boldsymbol{\Omega}_{1}\cdot\mathcal{G}\right)\mathbf{P} - \mathbf{PV}\boldsymbol{\Omega}_{2}\mathbf{FQ} \qquad (6.19)$$

The first two are effective one- and two-body densities, respectively, while the final quantity is an effective Fockian. In the latter, with units of energy, we have collected the multipliers $\boldsymbol{\lambda}$ into a single block-diagonal matrix for convenience. The Pulay force is then the trace of the effective Fockian with the overlap, i.e. density, derivative, $\mathbf{S}^{(q)}$.

Similarly, by inspection of equation 6.7, we can define equivalent quantities for the ALMO gradient as

$$\mathbf{P}^{(1)}_{\text{ALMO}} = 2\mathbf{P}^{(2)}_{\text{ALMO}} = \mathbf{P}$$
$$\mathbf{F}^{(S)}_{\text{ALMO}} = \mathbf{PFP}$$

Using these, the generalised gradient for both variants of the method can be written as

$$E^{(q)} = \text{Tr}\left\{\mathbf{H}^{(q)}\mathbf{P}^{(1)} + \mathcal{G}^{(q)}\cdot\mathbf{P}^{(2)} - \mathbf{S}^{(q)}\mathbf{F}^{(S)}\right\} \tag{6.20}$$

Writing the gradient in this manner not only simplifies computation, it also gives us effective densities for the non-variational methods. To see the significance of this, consider the gradient of the energy from a variational method with respect to any given parameter, α, on which the Hamiltonian depends explicitly. The wavefunction determined under this Hamiltonian will then depend *implicitly* on α and satisfy

$$\hat{H}(\alpha)\Psi_\alpha = E(\alpha)\Psi_\alpha$$

Taking the total derivative of the energy expectation with respect to the parameter gives

$$\frac{dE}{d\alpha} = \frac{d}{d\alpha}\langle\hat{H}(\alpha)\rangle = \left\langle\frac{d\hat{H}(\alpha)}{d\alpha}\right\rangle + \left\langle\frac{d\Psi_\alpha}{d\alpha}|\hat{H}(\alpha)\Psi_\alpha\right\rangle + \left\langle\hat{H}(\alpha)\Psi_\alpha|\frac{d\Psi_\alpha}{d\alpha}\right\rangle$$

$$= \langle\hat{h}^{(\alpha)}\rangle + \langle\hat{g}^{(\alpha)}\rangle + E(\alpha)\frac{d}{d\alpha}\langle\Psi_\alpha|\Psi_\alpha\rangle$$

where in the last line we have inserted the form of the molecular Hamiltonian. In the case where the wavefunction is variational, this is the Hellmann-Feynman theorem [354, 355], and reduces to the total derivative of the energy being the expectation value of the derivative of the Hamiltonian. In the discretised version, the density matrix encodes the wavefunction, and expectation values comprise traces of the Hamiltonian matrices over the density. The analogy between the above and the effective densities and Fockian of equation 6.20 then becomes clear. Most importantly, this is agnostic to the parameter, so long as the wavefunction does not depend explicitly on it. That means the same effective (sometimes called 'relaxed') density approach can be used to calculate properties in the presence of, for example, electric or magnetic fields, given a suitable adaptation of the molecular Hamiltonian [34].

6.3 RANDOM PHASE APPROXIMATION

The random phase approximation, like the charge transfer term, involves the self-consistent calculation of amplitudes satisfying a residual, as in equation 3.56. These amplitudes are then used in the calculation of the energy, via equation 3.57. Once again, the virtuals have been projected, and there is a complex dependence in both terms on both the virtual and occupied orbital coefficients. This is in addition to the amplitudes, a further parameter in the Lagrangian:

$$
\begin{aligned}
\mathcal{L}_{\text{ALMO+RPA}} =& \mathcal{L}_{\text{ALMO+CT}}\left(\mathbf{q};\mathbf{T},\mathbf{V},\mathbf{x};\boldsymbol{\lambda},\boldsymbol{\omega}\right) + E_{\text{RPA}}\left(\mathbf{q};\mathbf{T},\mathbf{V},\mathbf{t}\right) \\
& - \tau_{ij}^{ab}\left[\mathbf{R}_{\text{RPA}}\left(\mathbf{q};\mathbf{T},\mathbf{V},\mathbf{t}\right)\right]_{ij}^{ab}
\end{aligned}
\tag{6.21}
$$

where we have introduced the new Lagrange multipliers, τ_{ij}^{ab} for the residual term. Fortunately, we note that the new term does not depend on the CT amplitudes at all, so that the multipliers $\boldsymbol{\omega}$ are determined via equation 6.9 exactly as before.

There are a number of additional complications. Firstly, the exact form of both the energy and residual depend on the exchange variant chosen: direct, SOSEX, or RPAx. However, each of these have the same functional *form*, and the same dependence on the relevant parameters, all that changes is whether they use the bare Coulomb integrals, $\mathbf{K}$, or the antisymmetrised ones, $\mathbf{B}$; we will remain agnostic by using $\mathbf{Z}$ as a placeholder.

The second problem is that the method generally uses the pairwise approximation with decomposed excitations, as outlined in chapter 3. We could directly find the gradient of the *interaction energy*, but then this neglects any dynamical correlation within the fragments themselves. In terms of geometry optimisations and the calculation of molecular properties, this would be problematic. Within the RPA calculation, the intramolecular terms are determined (and identified in Figure 3.4), but excluded from the interaction energy. If we include these in our Lagrangian, we are including the intramolecular forces in the gradient, as required. However, if the pairwise approximation is used, we will be counting those forces multiple times - once for each pair they appear in, so in principle up to $F-1$ times, where F is the number of fragments.

In this section, we instead assume that the RPA calculation has been carried out on the total supermolecular system. While this is unlikely in most cases, it simplifies the derivation, and the form of the gradient found in this way is identical to when the pairwise approximation is included. The specific details of adapting it to include that approximation are deferred to the next section, as they are more a question of technical implementation than theoretical development.

6.3.1 RPA amplitudes

As we are not considering the projected coefficients at this point, it is easier to work in the orthogonalised basis of equation 3.54, indicated throughout the following by a tilde over a tensor. The results can then be translated as before into the projected ALMO basis if necessary. Under these conditions, the derivatives with respect to the amplitudes are simply:

$$\frac{\partial E_{RPA}}{\partial \tilde{t}_{ij}^{ab}} = \alpha_Z \tilde{Z}_{iajb}$$

$$\frac{\partial \left[\mathbf{R}_{dRPA} \right]_{kl}^{cd}}{\partial \tilde{t}_{ij}^{ab}} = \delta_{ki} \delta_{ac} \left[\tilde{\mathbf{A}} + \tilde{\mathbf{t}} \cdot \tilde{\mathbf{Z}} \right]_{jbld} + \delta_{lj} \delta_{bd} \left[\tilde{\mathbf{A}} + \tilde{\mathbf{Z}} \cdot \tilde{\mathbf{t}} \right]_{kcia}$$

where $\alpha_Z = 1/2$ if $\tilde{\mathbf{Z}} = \tilde{\mathbf{K}}$ and $1/4$ for $\tilde{\mathbf{Z}} = \tilde{\mathbf{B}}$. Combining these with the relevant Lagrange multipliers gives

$$\frac{\partial \mathcal{L}_{ALMO+RPA}}{\partial \tilde{t}_{ij}^{ab}} = \left[\alpha_Z \tilde{\mathbf{Z}} + \boldsymbol{\tau} \cdot \left(\tilde{\mathbf{A}} + \tilde{\mathbf{t}} \cdot \tilde{\mathbf{Z}} \right)^{\dagger} + \left(\tilde{\mathbf{A}} + \tilde{\mathbf{Z}} \cdot \tilde{\mathbf{t}} \right)^{\dagger} \cdot \boldsymbol{\tau} \right]_{iajb} = 0$$

$$(6.22)$$

which can be solved for the multipliers $\boldsymbol{\tau}$ in the same way that equation 6.9 was solved for $\boldsymbol{\omega}$. Moreover, it can be solved in the orthogonalised basis, saving considerable computational effort.

6.3.2 Orbital coefficients

The RPA equations are written in the molecular orbital basis, and transforming into the AO basis would involve three steps: projection of the virtuals, transformation of the integrals to the orthogonalised MOs, and transformation into the ALMO basis. These layers make direct differentiation in terms of the orbital coefficients exceptionally complicated. However, some simplifications can be made by working in the orthogonalised basis and then using the chain rule. That is, for any given quantity M, where a tilde overhead denotes the orthogonalised coefficients, we have that

$$\frac{\partial M}{\partial C_p^{\mu}} = \frac{\partial M}{\partial \tilde{C}_q^{\gamma}} \cdot \frac{\partial \tilde{C}_q^{\gamma}}{\partial C_p^{\mu}} + \frac{\partial M}{\partial \tilde{t}_{ij}^{ab}} \cdot \frac{\partial \tilde{t}_{ij}^{ab}}{\partial C_p^{\mu}}$$

We first find the partial derivatives with respect to the transformations. For the occupied orbitals we get

$$\tilde{\mathbf{T}} = \mathbf{T} \boldsymbol{\sigma}^{-1/2} = \mathbf{T} \left[\mathbf{T}^{\dagger} \mathbf{S} \mathbf{T} \right]^{-1/2}$$

For the virtuals, however, we have the following relation:

$$\tilde{\mathbf{V}} = \mathbf{Q} \mathbf{V} \boldsymbol{\pi}^{-1/2} = \mathbf{Q} \mathbf{V} \left[\mathbf{Q}^{\dagger} \mathbf{V}^{\dagger} \mathbf{S} \mathbf{V} \mathbf{Q} \right]^{-1/2}$$

The expansions of the inverse square-root metrics then bely the complexity of the transformation of the amplitudes:

$$\tilde{\mathbf{t}} = \boldsymbol{\sigma}^{-1/2}\boldsymbol{\pi}^{-1/2}\mathbf{t}\boldsymbol{\pi}^{-1/2}\boldsymbol{\sigma}^{-1/2}$$

In all the derivatives that follow, we therefore need to insert the partial derivatives of the above with respect to the relevant coefficient. This is complicated by the presence of the inverse square-root metrics, and additionally by the projector $\mathbf{Q}$ introducing a dependence on the occupied coefficients in the virtual transformation, so that non-zero cross terms appear. The amplitude transformation is inherently mixed. For clarity of development, we will give our results in the transformed basis, and we will denote the additional partial derivative matrices as

$$\mathcal{T}_{\mu i v j} = \frac{\partial \tilde{T}_i^\mu}{\partial T_j^v}, \; \mathcal{V}_{\mu a v p} = \frac{\partial \tilde{V}_a^\mu}{\partial C_p^v}, \text{ and } \mathcal{Z}_{ijabvp} = \frac{\partial \tilde{t}_{ij}^{ab}}{\partial C_p^v} \qquad (6.23)$$

Notice that the $\mathcal{V}$ derivative can be with respect to either an occupied or virtual coefficient in the ALMO basis. The exact forms of these are given in Appendix F.

The derivatives we need for equations 3.54 are then for the integral tensor, $\mathbf{Z}$, and for the Fock elements, $\boldsymbol{\epsilon}$. These, written explicitly in terms of the coefficients, are as follows:

$$\tilde{Z}_{pqrs} = \tilde{C}_{\cdot p}^{\mu\cdot}\tilde{C}_{\cdot q}^{v\cdot}\left[\mu v|\rho\sigma\right]\tilde{C}_{r\cdot}^{\cdot\rho}\tilde{C}_{s\cdot}^{\cdot\sigma}$$
$$\epsilon_{ia,jb} = \delta_{ij}\tilde{V}_{a\cdot}^{\cdot\mu}F_{\mu v}\tilde{V}_{\cdot b}^{v\cdot} - \delta_{ab}\tilde{T}_{i\cdot}^{\cdot\mu}F_{\mu v}\tilde{T}_{\cdot j}^{v\cdot}$$

where $[\mu v|\rho\sigma]$ are the AO-basis two-electron integrals, bare if $\tilde{\mathbf{Z}} = \tilde{\mathbf{K}}$, and antisymmetrised appropriately if $\tilde{\mathbf{Z}} = \tilde{\mathbf{B}}$. From these, the derivatives are trivial:

$$\partial_{\tau t}\tilde{Z}_{pqrs} = \sum_{u\in\{p,q,r,s\}}\delta_{ut}\hat{r}(u/\tau)[\tilde{p}q|\tilde{r}s]$$
$$\partial_{\tau k}\epsilon_{ia,jb} = -\delta_{ab}\left(\delta_{ik} + \delta_{jk}\right)\left[\mathbf{F}\tilde{\mathbf{T}}\right]_{\tau k}$$
$$\partial_{\tau c}\epsilon_{ia,jb} = \delta_{ij}\left(\delta_{ac} + \delta_{bc}\right)\left[\mathbf{F}\tilde{\mathbf{V}}\right]_{\tau c}$$

where the operator $\hat{r}(u/\tau)$ replaces the index u with τ in the following quantity. In this and the following, $\partial_{\tau t}$ is a shorthand for the partial derivative with respect to the *transformed* coefficient, $\tilde{C}_t^\tau$.

We define the following intermediate for clarity:

$$\beta_{iajb} = \delta_{ij}\delta_{ab} + \tilde{t}_{ij}^{ab} \qquad (6.24)$$

From the above, the energy and residual derivatives are

$$
\begin{aligned}
\partial_{\tau t} E_{RPA} &= \alpha_Z \left(\partial_{\tau t} \tilde{Z}_{iajb} \right) \tilde{t}_{ij}^{ab} \\
&= \alpha_Z \sum_{u \in \{i,a,j,b\}} \left(\hat{r}(u/\tau)[i\tilde{a}|j\tilde{b}] \right) \left(\hat{r}(u/t)\tilde{t}_{ij}^{ab} \right) = Y_{\tau t}
\end{aligned}
$$

$$
\tag{6.25}
$$

$$
\begin{aligned}
\partial_{\tau t} \left[\mathbf{R}_{RPA} \right]_{iajb} &= \beta_{iakc} \left(\partial_{\tau t} Z_{kcld} \right) \beta_{ldjb} \\
&\quad + \partial_{\tau t} \epsilon_{iakc} \tilde{t}_{kj}^{cb} + \tilde{t}_{ik}^{ac} \partial_{\tau t} \epsilon_{kcjb} = \mathcal{Y}_{iajb\tau t}
\end{aligned}
\tag{6.26}
$$

Clearly, the second of these, as a rank-six tensor, will be exceptionally computationally intensive to determine. However, in the final expression, it is contracted with the multipliers $\boldsymbol{\tau}$, reducing to a rank-two tensor; thus it should never need to be stored in its six-index form.

Finally, we need the derivatives with respect to the transformed amplitudes. These are much simpler:

$$
\frac{\partial E_{RPA}}{\partial \tilde{t}_{ij}^{ab}} = \alpha_Z \tilde{Z}_{iajb}
\tag{6.27}
$$

$$
\frac{\partial \left[\mathbf{R}_{RPA} \right]_{kcld}}{\partial \tilde{t}_{ij}^{ab}} = \left[\tilde{\mathbf{A}} + \tilde{\mathbf{t}} \cdot \tilde{\mathbf{Z}} \right]_{kcia} \delta_{lj} \delta_{db} + \delta_{ik} \delta_{ac} \left[\tilde{\mathbf{A}} + \tilde{\mathbf{Z}} \cdot \tilde{\mathbf{t}} \right]_{jbld}
\tag{6.28}
$$

We can now combine these results with those of equations 6.12 and 6.14, to give equations for the multipliers $\boldsymbol{\lambda}$. First we define an auxiliary term common to both the occupied and virtual expressions:

$$
\eta_{\mu t} = \left(\alpha_Z \tilde{Z}_{kcld} - \left[\tilde{\mathbf{A}} + \tilde{\mathbf{t}} \cdot \tilde{\mathbf{Z}} \right]_{mekc} \tau_{ml}^{ed} - \tau_{kn}^{cf} \left[\tilde{\mathbf{A}} + \tilde{\mathbf{Z}} \cdot \tilde{\mathbf{t}} \right]_{meld} \right) \mathcal{Z}_{klcd\mu t}
\tag{6.29}
$$

which leads to the expressions for $\boldsymbol{\lambda}$:

$$
4 \left[\boldsymbol{\lambda}_{VV} \mathbf{V}^\dagger \mathbf{S} \right]_{a\mu} = 4 \left[\boldsymbol{\Lambda}_{VV}^{CT} \right]_{a\mu} + \left(Y_{\tau p} - \tau_{kl}^{cd} \mathcal{Y}_{kcld\tau p} \right) \mathcal{V}_{\mu a \tau p} + \eta_{\mu a}
\tag{6.30}
$$

$$
4 \left[\boldsymbol{\lambda}_{OO} \mathbf{T}^\dagger \mathbf{S} \right]_{i\mu} = 4 \left[\boldsymbol{\Lambda}_{OO}^{CT} \right]_{i\mu} + \left(Y_{\tau m} - \tau_{kl}^{cd} \mathcal{Y}_{kcld\tau m} \right) \mathcal{T}_{\mu i \tau m} + \eta_{\mu i}
\tag{6.31}
$$

Note that while these are the same in form, the virtual expression sums over *all* indices, p, due to the projection, while the occupied expression only sums over all occupied indices, m.

6.3.3 Total derivative

Following the formulation of equation 6.20, we now need to collect terms involving $\mathbf{H}^{(q)}$, $\mathcal{G}^{(q)}$, and $\mathbf{S}^{(q)}$ from the RPA energy and residual into new effective densities. The complication here arises from the fact that $\tilde{\mathbf{Z}}$ is different depending on the choice of exchange. As such we introduce an antisymmetriser, $\hat{A}_Z$, which reduces to the identiy when

$\tilde{\mathbf{Z}} = \tilde{\mathbf{K}}$, and is the standard antisymmetriser when $\tilde{\mathbf{Z}} = \tilde{\mathbf{B}}$. As such, we have that

$$\tilde{Z}_{iajb} = \tilde{T}_{i\cdot}^{\cdot\mu}\tilde{V}_{a\cdot}^{\cdot\nu}\hat{A}_Z \langle\mu\nu|\rho\sigma\rangle \tilde{T}_{\cdot j}^{\rho\cdot}\tilde{V}_{\cdot b}^{\sigma\cdot}$$

Of course the transformed coefficients also introduce terms involving $\mathbf{S}^{(q)}$ due to the presence of the inverse metrics. We follow the same strategy as in the previous section:

$$\tilde{\mathbf{Z}}^{(q)} = \tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{V}}^{\dagger} \cdot \hat{A}_Z \mathcal{G}^{(q)} \cdot \tilde{\mathbf{T}}\tilde{\mathbf{V}} + \frac{\partial\tilde{\mathbf{Z}}}{\partial\tilde{\mathbf{C}}} \cdot \frac{\partial\tilde{\mathbf{C}}}{\partial\mathbf{S}} \cdot \mathbf{S}^{(q)}$$

where we have slightly abused notation by writing the partial derivatives in terms of the emboldened matrices to simplify matters; in reality, these are derivatives with respect to elements of the tensor, and the dots imply contractions over pairs of indices. We write

$$\begin{aligned} \mathcal{S}_{\mu p\nu\tau}^{(1)} &= \frac{\partial\tilde{C}_p^{\mu}}{\partial S_{\nu\tau}} \\ \mathcal{S}_{iajb\nu\tau}^{(2)} &= \frac{\partial\tilde{t}_{ij}^{ab}}{\partial S_{\nu\tau}} \end{aligned} \tag{6.32}$$

the derivation of which is given in Appendix F. The energy contribution to the derivative is then simply

$$E_{\text{RPA}}^{(q)} = \alpha_Z\tilde{Z}_{iajb}^{(q)}\tilde{t}_{ij}^{ab} + \alpha_Z\tilde{Z}_{iajb} \cdot \mathcal{S}_{iajb\nu\tau}^{(2)}S_{\tau\nu}^{(q)}$$

The residual introduces terms in the effective densities and effective Fock matrix, due to the presence of the Fockian in the $\boldsymbol{\epsilon}$ tensor:

$$\begin{aligned} \left[\mathbf{R}_{\text{RPA}}^{(q)}\right]_{iajb} =& \beta_{iakc}\tilde{Z}_{kcld}^{(q)}\beta_{ldjb} + \epsilon_{iakc}^{(q)}\tilde{t}_{kj}^{cb} + \tilde{t}_{ik}^{ac}\epsilon_{kcjb}^{(q)} \\ &+ \left(\left[\tilde{\mathbf{A}} + \tilde{\mathbf{Z}}\cdot\tilde{\mathbf{t}}\right]_{ldjb} \mathcal{S}_{ialdv\tau}^{(2)} + \left[\tilde{\mathbf{A}} + \tilde{\mathbf{t}}\cdot\tilde{\mathbf{Z}}\right]_{iakc} \mathcal{S}_{kcjb\nu\tau}^{(2)}\right) S_{\tau\nu}^{(q)} \end{aligned}$$

where we have that

$$\begin{aligned} \epsilon_{iajb}^{(q)} &= \mathcal{E}_{ijab\mu\nu}^{(1)}F_{\mu\nu}^{(q)} + \left[\delta_{ij} - \delta_{ab}\hat{P}(i/a)\hat{P}(j/b)\right]\mathcal{E}_{iajb\mu\nu}^{(2)}S_{\mu\nu}^{(q)} \\ \mathcal{E}_{iajb\mu\nu}^{(1)} &= \delta_{ij}\tilde{V}_{a\cdot}^{\cdot\mu}\tilde{V}_{\cdot b}^{\nu\cdot} - \delta_{ab}\tilde{T}_{i\cdot}^{\cdot\mu}\tilde{T}_{\cdot j}^{\nu\cdot} \\ \mathcal{E}_{iajb\mu\nu}^{(2)} &= \left[1 + \hat{P}(a/b)\right]\left[\mathbf{F}\tilde{\mathbf{C}}\right]_{\tau a}\mathcal{S}_{\tau b\mu\nu}^{(1)} \end{aligned} \tag{6.33}$$

The permutation operators, $\hat{P}$, act on the indices, but are not indexed quantities, therefore the summation convention does not apply to them.

where we recall that $\hat{P}(p/q)$ interchanges the indices p and q. A useful compound of this is

$$\hat{\mathcal{P}}(ijk|abc) = 1 + \hat{P}(i/k)\hat{P}(a/c)\hat{P}(k/j)\hat{P}(c/b) \tag{6.34}$$

which will be used repeatedly below to simplify expressions.

Collecting all like terms, the effective density and Fock matrices are thus, in comparison to equations 6.17 through 6.19:

$$\left[\mathbf{P}_{\text{ALMO+RPA}}^{(1)}\right]_{\mu\nu} = \left[\mathbf{P}_{\text{ALMO+CT}}^{(1)}\right]_{\mu\nu} - \tau_{ij}^{ab}\hat{\mathcal{P}}(ijk|abc)\left[\mathcal{E}_{ikac\mu\nu}^{(1)}\tilde{t}_{kj}^{cb}\right] \tag{6.35}$$

$$\left[\mathbf{P}^{(2)}_{\text{ALMO+RPA}}\right]_{\mu\nu\rho\sigma} = \left[\mathbf{P}^{(2)}_{\text{ALMO+CT}}\right]_{\mu\nu\rho\sigma} + \alpha_Z \tilde{t}^{ab}_{ij} \tilde{T}^{\cdot\mu}_{i\cdot} \tilde{V}^{\cdot\nu}_{a\cdot} \tilde{T}^{\rho\cdot}_{\cdot j} \tilde{V}^{\sigma\cdot}_{\cdot b} \cdot \hat{A}_Z$$
$$- \tau^{ab}_{ij} \beta_{iakc} \beta_{ldjb} \tilde{T}^{\cdot\mu}_{k\cdot} \tilde{V}^{\cdot\nu}_{c\cdot} \tilde{T}^{\rho\cdot}_{\cdot l} \tilde{V}^{\sigma\cdot}_{\cdot d} \cdot \hat{A}_Z \qquad (6.36)$$
$$- \tau^{ab}_{ij} \hat{\mathcal{P}}(ijk|abc) \left[\mathcal{E}^{(1)}_{ikac\mu\nu} \tilde{t}^{cb}_{kj}\right] P_{\sigma\rho}$$

$$\left[\mathbf{F}^{(S)}_{\text{ALMO+RPA}}\right]_{\mu\nu} = \left[\mathbf{F}^{(S)}_{\text{ALMO+CT}}\right]_{\mu\nu} - \alpha_Z \tilde{Z}_{iajb} \cdot \mathcal{S}^{(2)}_{iajb\mu\nu}$$
$$- \alpha_Z \tilde{t}^{ab}_{ij} \sum_{q\in\{i,a,j,b\}} \left(\hat{r}(q/\tau)\left[\tilde{ia}|\tilde{jb}\right]\right) \mathcal{S}^{(1)}_{\tau q\mu\nu}$$
$$+ \tau^{ab}_{ij} \beta_{iakc} \beta_{ldjb} \sum_{q\in\{k,c,l,d\}} \left(\hat{r}(q/\tau)\left[\tilde{kc}|\tilde{ld}\right]\right) \mathcal{S}^{(1)}_{\tau q\mu\nu}$$
$$+ \tau^{ab}_{ij} \left(\left[\tilde{A}+\tilde{Z}\cdot\tilde{t}\right]_{ldjb} \mathcal{S}^{(2)}_{iald\mu\nu} + \left[\tilde{A}+\tilde{t}\cdot\tilde{Z}\right]_{iakc} \mathcal{S}^{(2)}_{kcjb\mu\nu}\right)$$
$$+ \tau^{ab}_{ij} \left[1+\hat{p}(ij|k)\hat{p}(ab|c)\right] \left(\tilde{t}^{cb}_{kj}\left[\delta_{ik}-\delta_{ac}\hat{P}(i/a)\hat{P}(k/c)\right]\mathcal{E}^{(2)}_{iakc\mu\nu}\right)$$
$$- \tau^{ab}_{ij} \left[1+\hat{p}(ij|k)\hat{p}(ab|c)\right] \left(\tilde{t}^{cb}_{kj}\mathcal{E}^{(1)}_{ikac\sigma\tau}\mathcal{G}_{\sigma\tau\gamma\eta}P_{\eta\mu}P_{\gamma\nu}\right)$$
$$(6.37)$$

where the conjugating-permutation operators, $\hat{p}(pq|r)$, have the following effect:

$$\hat{p}(pq|r)\left(M_{pr}N_{rq}\right) = N_{pr}M_{rq}$$

The final ALMO+RPA gradient is then found by inserting the effective densities and Fockian of equations 6.35 through 6.37 into equation 6.20.

6.4 IMPLEMENTATION

The basic implementation of the gradients is fairly simple. The derivative integral matrices are either calculated on the fly or stored in memory, then combined with the effective densities relevant to the level of theory being used. To calculate these effective densities, we need to solve sets of coupled-perturbed equations for the multipliers ω and τ; both of these are in the form of Riccatti equations, the solution of which was detailed in earlier chapters. With these, and the orbital energies λ, the effective densities are built according to their energy expressions.

The evaluation of the various expressions in the preceding sections is somewhat technically involved, as there are several high-rank intermediate tensors, particularly in the RPA gradient. It is impractical and ill-advised to calculate and store these directly, as they would require up to $\mathcal{O}(O^4V^4)$ storage. The gradient for a given coordinate is a single-valued, scalar quantity, such that in principle no intermediates should need to be stored. In practice, a great deal of time can be saved by storing the two-index intermediates, in particular the derivative in-

tegral matrices $\mathbf{H}^{(q)}$ and $\mathbf{S}^{(q)}$, as these require negligible storage. All other quantities can be computed on the fly through a series of at most $\mathcal{O}(N^5)$ contractions. In particular, the derivatives of the two-electron integrals can be built in the same way as the two-electron integrals themselves, by providing the four-index quantities they will contract with. As such, the two-body density and the two-electron integral derivatives should be computed in tandem and combined directly.

However, all of the above assumes that we have used none of the approximations that make the method linear-scaling. Of these, the ones that affect the gradient calculation are: the density-fitting of the two-electron integrals; the local approximation in the Fock build; and the pairwise RPA decomposition. In this section, I will outline how the first and last of these change the results from the beginning of this chapter, and how they can be used to give analytic gradients that scale roughly on the same order as the energy calculations.

The second point listed above, the local Fock build, is an approximation to the exact Fockian. If the derivatives are with respect to the true Fockian, this is not a problem, although it will disagree with numerical gradients. In fact, it removes the choice of domains from the gradient, which would help avoid possible discontinuities. In the derivative routine itself, the only place where a local approximation could be employed is in the contraction of the two-electron integral derivatives with the two-body effective density. We have not considered this yet due to the sheer complexity of the effective density in the case of the RPA variants of the method. I will, however, comment briefly on this contraction and how it can be made negligible for the ALMO and ALMO+CT gradients at the end of the next section.

6.4.1 Density fitting

The density fitting of the two-electron integrals affects the calculation of $\mathcal{G}^{(q)}$, only insofar as it is an approximation to the true integral tensor $\mathcal{G}$. In principle, one could use density fitting in the energy calculation but not in the gradient [356–358], and the gradients would be *more* accurate, although they would not necessarily agree with numerical gradients due to the approximation. However, the gradient calculation would then lose all of the savings and improved scaling of the method, and would take an order of magnitude more effort. There would then be no real computational advantage in using analytical versus numerical gradients.

The derivative is fairly simple to determine, starting from equation 1.12:

$$\langle\mu\nu|\rho\sigma\rangle^{(q)} \approx (\mu\nu|P)^{(q)}\left[\mathbf{G}^{-1}\right]_{PQ}(Q|\rho\sigma) + (\mu\nu|P)\left[\mathbf{G}^{-1}\right]_{PQ}(Q|\rho\sigma)^{(q)}$$
$$+ (\mu\nu|P)\left[\mathbf{G}^{-1}\right]_{PQ}^{(q)}(Q|\rho\sigma)$$

As per Algorithm 4.1, we Cholesky-decompose $\mathbf{G} = \mathbf{L}\mathbf{L}^{\mathsf{T}}$ and write

$$b_{\mu\nu P} = (\mu\nu|R)L_{RP}^{-1}$$

We can then similarly define

$$b_{\mu\nu P}^{(q)} = (\mu\nu|R)^{(q)}L_{RP}^{-1}$$

where we note $\mathbf{b}^{(q)}$ is *not* the total derivative of $\mathbf{b}$. The ERI derivative can then be written as

$$\langle\mu\nu|\rho\sigma\rangle^{(q)} \approx b_{\mu\nu P}^{(q)}b_{P\rho\sigma}^{\mathsf{T}} + b_{\mu\nu P}b_{P\rho\sigma}^{(q)\mathsf{T}} + (\mu\nu|P)\left[\mathbf{G}^{-1}\right]_{PQ}^{(q)}(Q|\rho\sigma)$$

The tensors $\mathbf{b}$ and $\mathbf{L}^{-1}$ are already computed and stored in the course of the energy calculation, and can thus be reused here. All that then needs computing for the first two terms is the three-index derivative tensor elements $(\mu\nu|P)^{(q)}$, which is an order of magnitude cheaper to compute and store than the full four-index derivative. The final term in the above, involving the derivative of the inverse fitting metric, appears to complicate things slightly. However, we can expand this as

$$\left[\mathbf{G}^{-1}\right]^{(q)} = -\mathbf{G}^{-1}\mathbf{G}^{(q)}\mathbf{G}^{-1} = -\mathbf{L}^{-1}\left(\left[\mathbf{L}^{-1}\right]^{\mathsf{T}}\mathbf{G}^{(q)}\mathbf{L}^{-1}\right)\left[\mathbf{L}^{-1}\right]^{\mathsf{T}}$$

If we then define

$$\overline{\mathbf{G}}^{(q)} = \left[\mathbf{L}^{-1}\right]^{\mathsf{T}}\mathbf{G}^{(q)}\mathbf{L}^{-1} \tag{6.38}$$

the ERI derivative simplifies to give

$$\langle\mu\nu|\rho\sigma\rangle^{(q)} \approx b_{\mu\nu P}^{(q)}b_{P\rho\sigma}^{\mathsf{T}} + b_{\mu\nu P}b_{P\rho\sigma}^{(q)\mathsf{T}} - b_{\mu\nu P}\overline{G}_{PQ}^{(q)}b_{Q\rho\sigma}^{\mathsf{T}} \tag{6.39}$$

which only requires computation of the new two-index quantity from equation 6.38.

A further consideration at this point is the integral blocking technique used to reduce the storage and computation cost in the energy calculation, Algorithm 4.2. A negligible integral value does *not* imply a negligible value of the integral derivative, such that we cannot simply apply the same procedure. Again, this may seem like it will result in erroneous gradients, however this is not the case - the integral blocking is an approximation to the true DF-ERI tensor, such that

the true DF-ERI gradient will give correct results, although not necessarily in agreement with numerical differentiation of the approximate tensor.

If we wish to exploit the potential for linear scaling in the integral calculation as per the original integrals, we need to block screen the gradient integrals, $\mathbf{b}^{(q)}$, directly; as before, there is no real value in doing the same for the metric, as it is a negligible overall cost. Unfortunately, the inequality of equation 4.7 no longer applies, so we need to derive a new bound. The same algorithm as in Algorithm 4.2 can then be applied but with this new bound replacing line 15.

Consider the derivative of the electron repulsion integral of equation 4.3 with respect to a coordinate, q_P, on the atomic centre P. Using the chain rule, we get

$$\frac{d}{dq_P}V_{pq}^n = \sqrt{\frac{4\alpha}{\pi}}F'_n\left(\alpha R_{PQ}^2\right) \cdot \frac{d}{dq_P}\left(\alpha R_{PQ}^2\right) = \frac{4\alpha^{3/2}}{\sqrt{\pi}}\left(q_P - q_Q\right)F'_n\left(\alpha R_{PQ}^2\right)$$

The Boys function derivative is, using equation 4.4:

$$F'_n(q) = \frac{d}{dq}\int_0^1 dt\; t^{2n}\exp\left(-qt^2\right) = -\int_0^1 dt\; t^{2n+2}\exp\left(-qt^2\right) = -F_{n+1}(q)$$

Therefore the derivative of the ERI is

$$\frac{d}{dq_P}V_{pq}^n = -\frac{4\alpha^{3/2}}{\sqrt{\pi}}\left(q_P - q_Q\right)F_{n+1}\left(\alpha R_{PQ}^2\right) = -2\alpha(q_P - q_Q)V_{pq}^{n+1}$$

$$(6.40)$$

From equation 6.40, we see that the derivative of the integral is itself an ERI integral, but with higher angular momentum. As such, all but the distance dependence of equation 4.7 can be reused, with the derivatives of the 'diagonal' ERIs, and the two-particle effective density instead of the one-particle reduced density. To determine the distance dependence asymptotically, we go back to equation 4.5 but with a higher order Boys function. This gives

$$\frac{d}{dq_P}V_{pq}^n \sim -2\alpha R_{PQ} \cdot \frac{2}{\alpha^{n+1}R_{PQ}^{2n+3}} = -\frac{4}{\alpha^n R_{PQ}^{2n+2}} \qquad (6.41)$$

Thus the worst case is again for spherical Gaussians, but now the distance dependence is R^{-2} instead of R^{-1}. As such, we can write the screening criterion, using the notation of equation 4.7, for the ERI derivatives as

$$\left\|\mathbf{B}_{XYZ}^{(q)}\mathbf{P}_{XY}^{(2)}\right\|_\infty \lesssim 4\left\|\mathbf{g}_{XYXY}^{(q)}\right\|_\infty^{1/2}\left\|\mathbf{G}_{ZZ}^{(q)}\right\|_\infty^{1/2}\left\|\mathbf{P}_{XY}^{(2)}\right\|_\infty R_{(XY)Z}^{-2} \qquad (6.42)$$

As with the density fitting of the original ERIs, the only new quantities that need to be calculated here are the diagonal blocks of the ERI derivative tensor and the fitting metric derivative matrix. The latter

is needed anyway in the density fitting procedure of equation 6.39, while the cost of the former is only $\mathcal{O}(N^2)$.

Further savings in the computational cost, but not the scaling, can be found through density fitting the gradient if we considered in detail the contraction of equation 6.39 with the effective two-particle density, $\mathbf{P}^{(2)}$. For the ALMO+RPA expression of equation 6.36, this is complicated, and would require considerable analysis and fine tuning that is yet to be done. Similar considerations were needed in the implementation of canonical RPA derivatives [359], and those for the RPA variants used in range-separated DFT [360, 361], which we could build from in future. For the ALMO gradient, however, it is much simpler. The two-electron term in the gradient is simply

$$\frac{1}{2}\mathrm{Tr}\left\{\mathcal{G}^{(q)} \cdot \mathbf{P}\right\} \approx \frac{1}{2}\mathrm{Tr}\left\{\mathbf{b}^{(q)}\,[\mathbf{b}\mathbf{P}] + \mathbf{b}\left[\mathbf{P}\mathbf{b}^{(q)}\right]^{\mathsf{T}} - \mathbf{b}\overline{\mathbf{G}}^{(q)}\,[\mathbf{b}\mathbf{P}]^{\mathsf{T}}\right\}$$
$$= \frac{1}{2}\left\{\left(\mathbf{b}^{(q)} - \mathbf{b}\overline{\mathbf{G}}^{(q)}\right) \cdot [\mathbf{b}\mathbf{P}] + \mathbf{b}\left[\mathbf{P}\mathbf{b}^{(q)}\right]^{\mathsf{T}}\right\}$$

The contraction of $\mathbf{b}$ or $\mathbf{b}^{(q)}$ with $\mathbf{P}$ is an $\mathcal{O}(N^2)$ term resulting in a rank-one tensor of dimension M, where M is the size of the auxiliary basis. The two-body force calculation is thus transformed from an $\mathcal{O}(N^3)$ trace over the four-index $\mathcal{G}$ to a series of two $\mathcal{O}(N^2)$ and two $\mathcal{O}(NM)$ traces, by partitioning the contractions as shown by the brackets.

The ALMO+CT effective density of equation 6.18 can be handled in exactly the same manner, with the only additional consideration being the contraction of $\mathbf{P}$ with $\mathbf{\Omega}_1$. As can be seen from equation 6.16, this corresponds to a series of matrix multiplications of order at most NV. Thus the overall scaling is still reduced an order of magnitude overall.

6.4.2 Pairwise RPA

In the RPA portion of the calculation, a list, $\mathcal{P}$, of fragment pairs is created, culled by some cutoff distance between the fragments. The total *interaction energy* is then determined as the sum of all the dispersion and exchange-dispersion terms for the pairs in this list, as defined by equations 3.70 and 3.71.

However, in a geometry optimisation, using this as the contribution to the forces is not sufficient, as it neglects the on-fragment correlation energies, which will strongly affect the local geometry of each fragment. One possibility is to include the total RPA contributions from every pair, but there are two significant problems there: it would re-introduce BSSE into the gradients, and we would count each fragment's correlation energy multiple times, in general. Instead, we compute the averaged fragment correlation energies.

There is one excitation diagram not shown in Figure 3.4, which is where both excitations happen on the same fragment. Naturally, this corresponds to on-fragment correlation, and is an important part of the total energy of that fragment. Moreover, it is necessarily untainted by BSSE. Of the other diagrams, it is impossible to say what part of the ionic (or charge-transfer) contribution is necessary in the total energy of the system, but there is a plausibility argument for excluding them. Such diagrams are necessarily not possible in a calculation on a single fragment. Their availability does, however, expand the virtual space available and thus a higher quality fragment correlation energy could be computed. On the other hand, their removal does not particularly affect the quality of the interaction energy calculated; in fact, it seems to improve it. The correlated part of the interaction energy is equivalent to the total correlation energies of the system minus that of the individual fragments. As such, we would expect minimal error by reversing this and approximating the total correlation energy of the system as the interaction energy plus the fragment energies.

In terms of implementation, this is also the simplest approach to take, as it does not affect the overall form of the gradient. In particular, it does not change the individual residual equations for each pair, which must be satisfied. Thus all that changes is how the Lagrangian is written, c.f. equation 6.21:

$$
\begin{aligned}
\mathcal{L}_{\text{ALMO+RPAd}} =& \mathcal{L}_{\text{ALMO+CT}} + \sum_{X,Y \in \mathcal{P}} \left[1 + \hat{P}(a/b) \right] \sum_{(ia) \in X} \sum_{(jb) \in Y} \tilde{t}_{ij}^{ab} \tilde{Z}_{iajb} \\
&+ \sum_{X,Y \in \mathcal{P}} \left\{ n_X^{-1} \sum_{(iajb) \in X} + n_Y^{-1} \sum_{(iajb) \in Y} \right\} \tilde{t}_{ij}^{ab} \tilde{Z}_{iajb} \\
&- \sum_{X,Y \in \mathcal{P}} \left[\boldsymbol{\tau}^{XY} \right]_{ij}^{ab} \left[\mathbf{R}_{\text{RPA}}^{XY} \right]_{ij}^{ab}
\end{aligned}
$$

(6.43)

where n_X is the multiplicity of fragment X in $\mathcal{P}$; that is, the fragment X appears in n_X non-negligible fragment pairs in the pairlist, $\mathcal{P}$.

The effective densities of equations 6.35 through 6.37 only change in that each of the RPA-based contributions must now be calculated for each relevant pair, and their energy-derivative contributions summed in the same way as equation 6.43. These energy contributions are only found in the two-body and Pulay densities, equations 6.36 and 6.37, and can be identified as they always include the prefactor α_Z.

At first, it appears that having to perform $\dim(\mathcal{P})$ RPA gradient calculations (and thus the solution of as many coupled-perturbed RPA equations) will introduce an extreme overhead. However, the gradient calculation scales at roughly the same complexity as the energy calculation, albeit with a larger prefactor. Therefore, we immediately inherit asymptotic linear scaling of the RPA gradient calculations with

number of fragments. That is to say $\dim(\mathcal{P})$ gradient calculations with roughly $2N_F$ basis functions is orders of magnitude less complex than one such calculation with FN_F basis functions.

6.5 TESTS

In this section, we validate the derivation and implementation of the analytical gradients. We do this by comparing to both CCSD results, and to those from numerical differentiation. The latter were all calculated using the central differences method with a step length of 0.01 bohr in either direction, on the full 3N coordinates; Hessian calculations used the same method, but on the analytical gradients. The geometry optimisations were performed using a conjugate gradients procedure [124], with a heuristic line search [362], again on the full 3N coordinates. Considerable time savings, and improved robustness, could be achieved by instead using a quasi-Newton Raphson method on the $3N - 6$ independent degrees of freedom, in an internal coordinate representation [363]; this has not yet been implemented in GAMMA, nor have effective symmetry constraints. All calculations were performed using the aug-cc-pVDZ basis sets [324, 343, 364], and where applicable, the corresponding JKFit [151, 162, 325] and MP2Fit [325, 327] sets. ALMO calculations used the RPAx exchange variant, with the full infinite-order charge transfer correction. We did not use the local Fock build in the dimer calculations, as it is unnecessary, but did in the larger cluster calculations.

6.5.1 Water dimer

We have already presented interaction energy curves for the water dimer as a function of intermolecular separation in Figure 5.13. While this showed good agreement, it does not demonstrate the efficacy of the method on the overall geometry, where there are 12 degrees of freedom in total. Figure 6.1 shows a comparison of the ALMO+RPAx geometry with that calculated with CCSD, demonstrating that the agreement is in fact good across all degrees. The largest errors are in the HOH bond angles of the monomers, with a difference of around 1.8 degrees; the intermolecular separation on the other hand agrees to within 0.01 angstrom. It is to be expected that a method focused on providing accurate interaction energies would perform best for the forces between monomers, and slightly less well for the forces within monomers. Overall, though, these differences are small and of the order of the differences between CCSD and MP2 geometries.

Figure 6.2 shows the forces on each atom as calculated numerically versus analytically. This clearly demonstrates that the analytical gra-
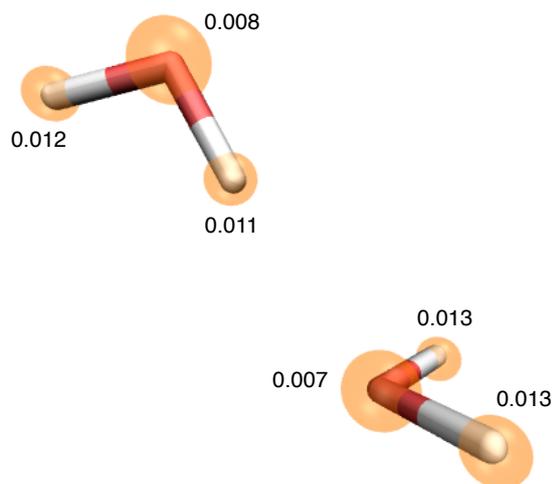
Figure 6.1: Optimised geometry of the water dimer at the ALMO+RPAxd level (solid bonds) compared to the CCSD level (orange spheres), both using the aug-cc-pVDZ basis set. The structures agree well, with only minor deviations, mainly in the hydrogen positions. Deviations of the centroids are given in bohr.
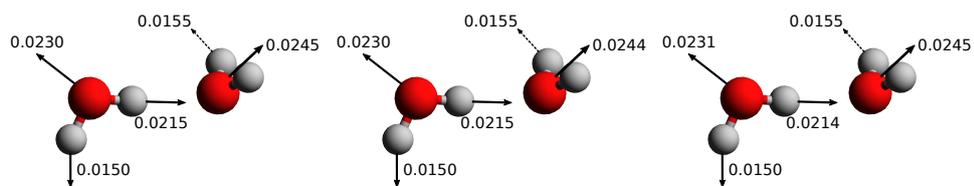
Figure 6.2: Comparison of the total forces on each atom in the water dimer, as calculated, from left to right: numerically using central differences; analytically with the full RPA calculation on the dimer; analytically with the pairwise RPA approximation, where only the dispersion and intramolecular terms are included. The dashed line indicates going into the page, and there is a totally symmetric line coming out of the page, reflected in the xz-plane, in all cases. The preservation of this symmetry is an important indicator that the method is robust.
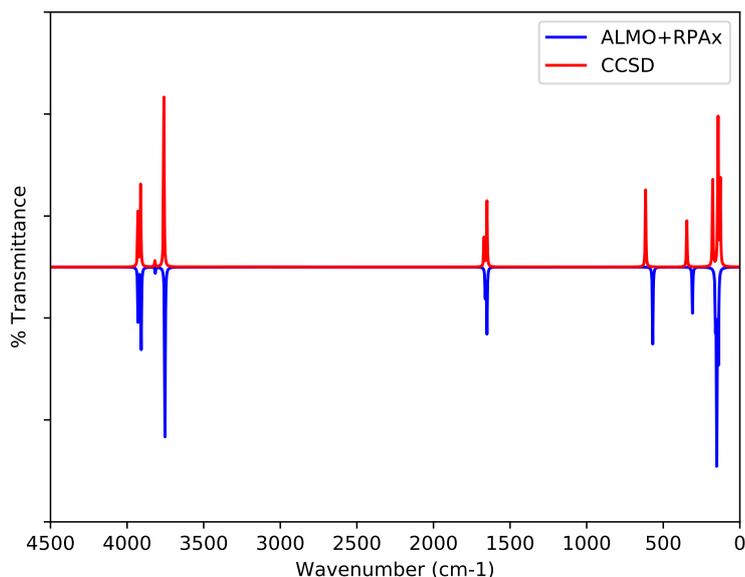
Figure 6.3: Harmonic frequency calculations for the water dimer at the ALMO+RPAx (bottom) and CCSD (top) levels using the aug-cc-pVDZ basis set. Intensities were calculated using electric dipole moments. Overall, the two spectra show very good agreement, with only slight deviations in the fingerprint region.

dients are correct, differing from the numerical results only in the fourth decimal place, which given the step size of 0.01, is the limit of precision for the central differences results. More interesting, however, is the difference between using the full RPA term in the total energy - i.e. the undecomposed energy - versus just including the dispersion and intramolecular terms. There is again essentially no difference, suggesting that we are valid in excluding the ionic and BSSE terms from the RPA energy. We see the same order of deviations for the water trimer and pentamer, i.e. negligible. It is difficult to validate this for even larger complexes, however, as it becomes impossible to include the full RPA on the total system, and including the ionic terms in the pairwise approximation raises a difficult question about how to weight them in the total energy. Our ad-hoc weighting of the intramolecular terms by how many times they appear, however, seems to not affect the accuracy of the gradients.

Finally, we can further validate both the method and the accuracy of the gradients by calculating the harmonic frequencies of the water dimer. These are essentially the eigenvalues of the mass-weighted Hessian, which can be calculated from the gradients using central differences. Figure 6.3 shows the spectra from ALMO+RPAx versus CCSD. Notably, the shape and position of the peaks is largely indistinguishable between the two, with the largest differences at around 615

$$345\ \text{cm}^{-1} \qquad\qquad 615\ \text{cm}^{-1}$$
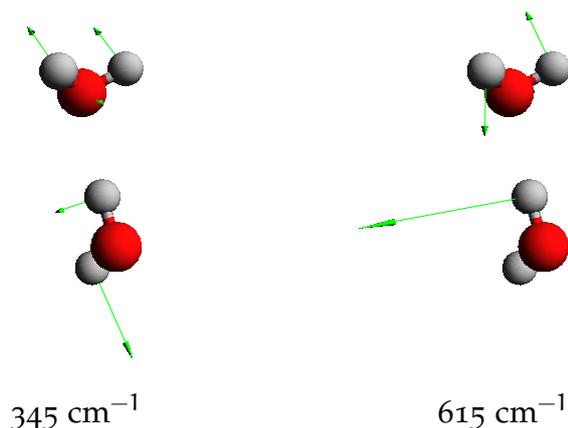
Figure 6.4: Forces corresponding to two of the normal modes of the water dimer calculated at the CCSD/aVDZ level.

and 346 wavenumbers, where the RPAx results are roughly 10 wavenumbers lower than their CCSD counterparts. These two normal modes can be seen in Figure 6.4, where it is clear that these are motions involving rocking of the two water molecules, i.e. the predominantly intermolecular vibrations. This is perhaps to be expected, as the method is providing an approximate treatment of the intermolecular interaction. However, we note that the CCSD frequencies will contain some BSSE, whereas the ALMO ones will not; as BSSE leads to an overestimation of force constants (i.e. a higher frequency mode), it is difficult to untangle the errors due to the approximation from those due to BSSE.

6.5.2 S66 benchmark set

As with the energy calculations, it is not sufficient to consider only select model complexes such as the water dimer. Therefore, we have optimised geometries for the entire S66 database [267], and compared these to the CCSD benchmarks. The root-mean-square deviations in the total structure, and maximum absolute deviations in individual parameters, are shown in Figure 6.5.

Overall, the structures show good agreement, performing particularly well for bond lengths. The largest absolute deviations in bond lengths are consistently in the intermolecular separation, but these are also the longest bond lengths overall - thus the percentage deviation is consistently low. The bigger differences are found in both the bond and dihedral angles, with maximum deviations of up to four degrees. This is not unlike the deviations that would be found comparing MP2 to CCSD [365], and suggest that our description of the monomers is not as high quality as our description of the interaction. Again, this
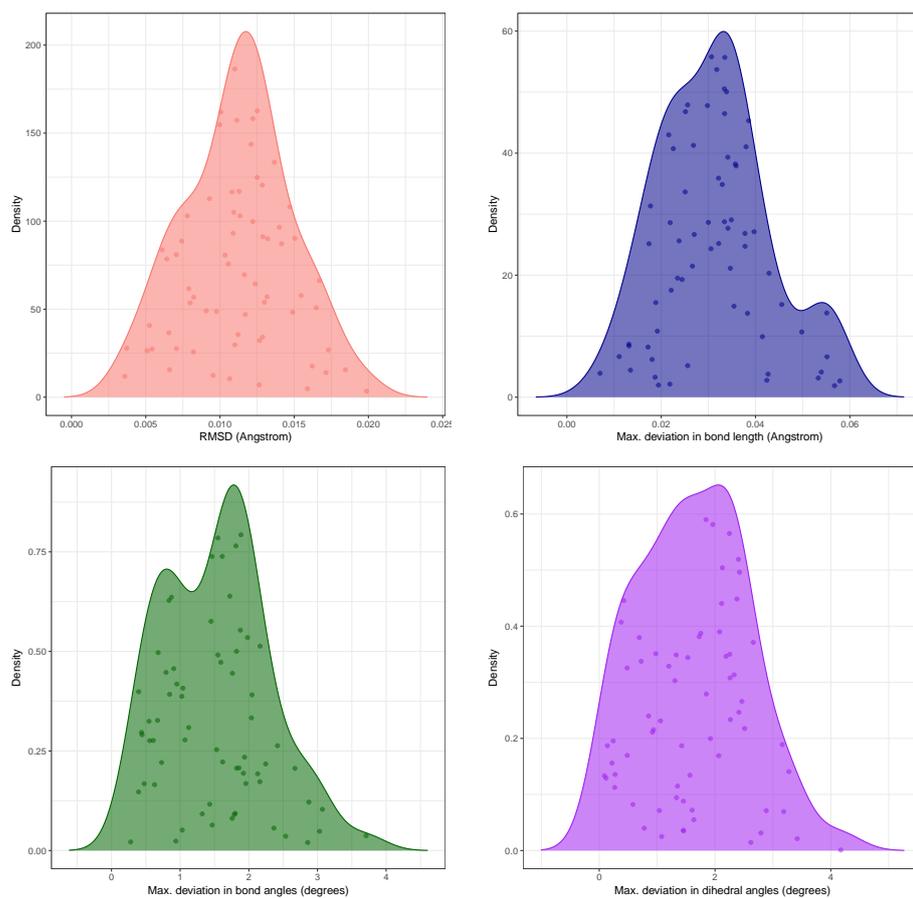
Figure 6.5: Geometric deviations in the ALMO+RPAxd/aVDZ optimised geometries compared to CCSD/aVDZ geometries, for the S66 benchmark dataset.

is to be expected, as the monomer description in the ALMO procedure is effectively a monomer-centred basis quality, as opposed to the dimer-centred basis quality of the full HF calculation. As noted earlier, the errors do not seem to be due to how we handle the RPA portion of the gradients.

There is no clear answer as to how to deal with the inherently poorer description of the intramolecular forces. The charge transfer correction to the interaction energy does achieve counterpoise-corrected HF agreement, but this does not solve the problem. To see why, consider the relation between the two:

$$E_{\text{ALMO+CT}}(AB) - E_A(A) - E_B(B) \approx E_{\text{HF}}(AB) - E_A(A\overline{B}) - E_B(\overline{A}B)$$

such that we rearrange to get

$$E_{\text{ALMO+CT}} \approx E_{\text{HF}}(AB) - \left[E_A(A\overline{B}) - E_A(A)\right] - \left[E_B(\overline{A}B) - E_B(B)\right]$$
$$= E_{\text{HF}} - E_{\text{BSSE}}$$

That is, by removing the BSSE *a priori* from the interaction energy, we are effectively reintroducing it into the total energy, and thus the gradients. Specifically, as E_{BSSE} is necessarily negative, we are raising the total energy with respect to the HF result, which appears to result in the change in bond angles within the monomers in particular, and concomitantly shorter intermolecular separations.

Pragmatically, however, the geometrical deviations are small enough as to be justifiable, especially when we consider the computational savings. Figure 6.6 compares the timings using numerical differentiation to using the analytical gradients with the implementation given in the previous section. For the larger complexes, such as the benzene dimer, we see decreases of two orders of magnitude; a single optimisation step took around 50 hours numerically, compared to 20 minutes analytically. With effective parallelisation of the gradients, as done for the energy calculations, we would expect this to reduce even further.

6.5.3 Larger clusters

Ideally, the gradients should scale on the same order as the energy calculations, so that geometry optimisations are feasible. Numerically, we need to do roughly $6N$ displacements, and thus for a method that scales as $\mathcal{O}(N)$, central differences scales as $\mathcal{O}(N^2)$ with a very large prefactor. This can be seen in Figure 6.7 for a series of water clusters.

Strictly speaking, in the analytical gradients, we still need to calculate $\mathcal{O}(N)$ integral derivative matrices, so even with the linear density-fitting and integral blocking routine, we will asymptotically reach $\mathcal{O}(N^2)$ scaling. However, the prefactor is expected to be much smaller. Additionally, the calculation of the integral derivatives with respect
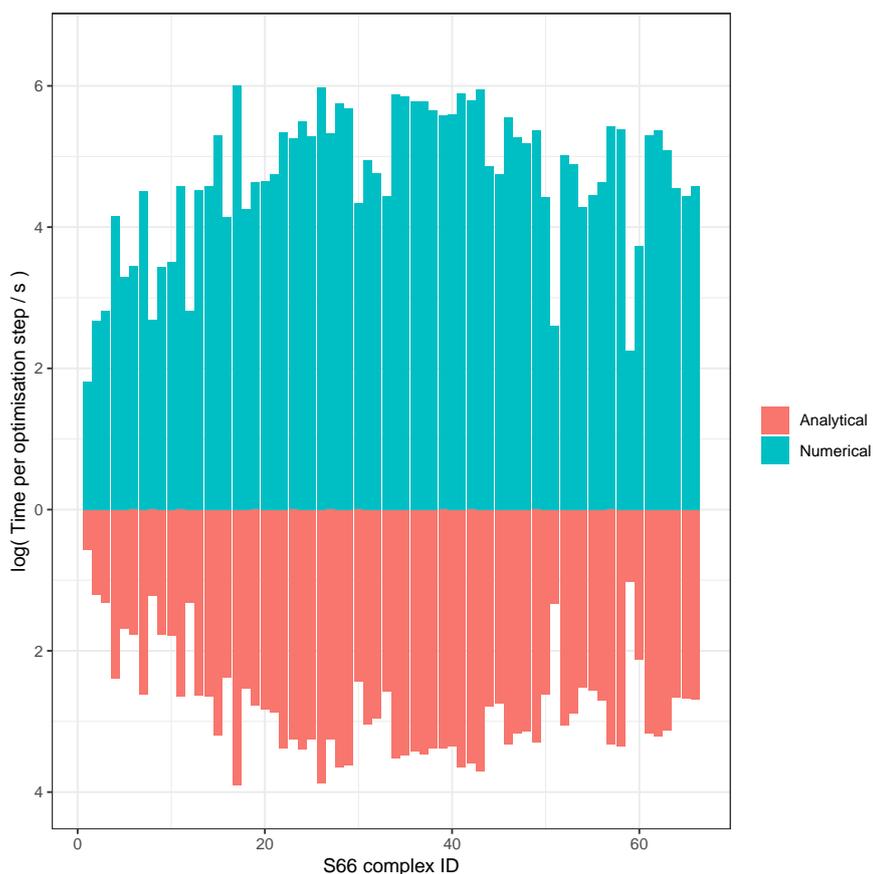
Figure 6.6: Timings in $\log_{10}$ seconds for a single optimisation step where the gradient is calculated numerically (top, blue) versus analytically (bottom, red), for the S66 benchmark dataset. The analytical derivatives are 1 to 2 orders of magnitude faster in all cases.
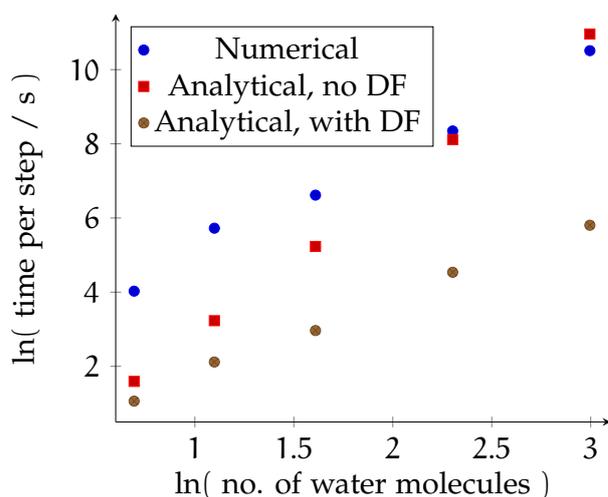
Figure 6.7: Scaling in the time taken to perform a single optimisation step, comparing numerical differentiation, analytical with the full 4-index ERI derivatives, and analytical with density-fitted integrals and screening. The scaling exponents are 2.5, 3.8, and 1.9, respectively.
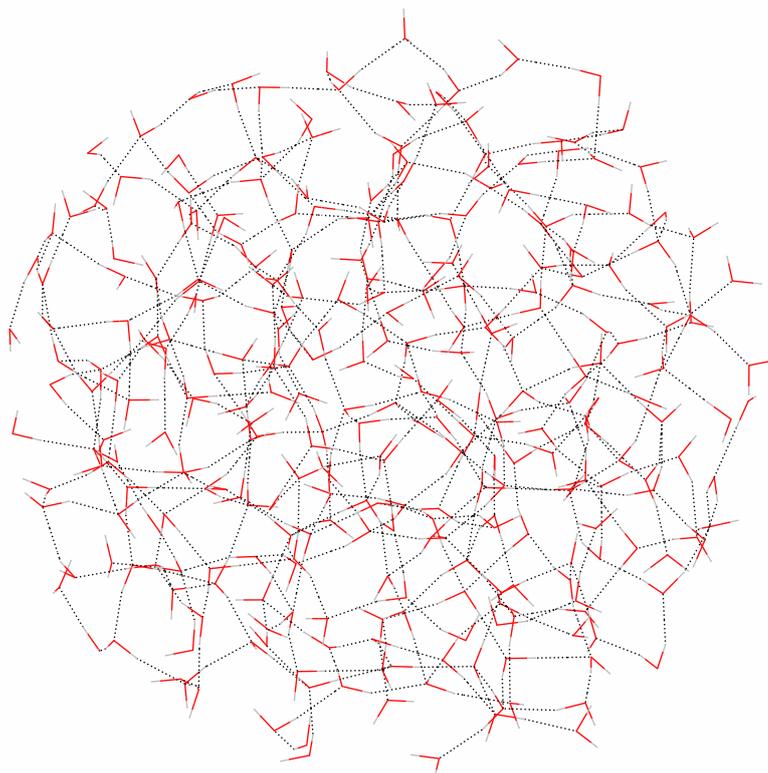
Figure 6.8: A cluster of 202 water molecules, geometry optimised at the ALMO+RPAxd/aVDZ level. The unoptimised cluster, taken from MD simulations, can be seen in Figure 5.10.

to each coordinate is not strictly order N, as the gradients between different coordinates are coupled [363]. As such, we ideally expect sub-quadratic (but still super-linear) scaling of the analytical gradients with the density fitting procedure. If we use the full ERI tensor, however, the prefactor will still be much lower, but the asymptotic $\mathcal{O}(N^4)$ scaling will eventually make numerical differentiation cheaper. Both of these scaling behaviours, and their scaling exponents, can be found in Figure 6.7.

We finally demonstrate the efficacy of the analytical approach by optimising the cluster of 202 water molecules from Figure 5.10. The new, optimised structure can be found in Figure 6.8, where we can see a clear move towards a more crystalline structure. As these calculations are effectively at absolute zero, this is to be expected. Further evidence of this can be seen in the radial distribution function of a water molecule at the centre of the cluster, as shown in Figure 6.9. The geometry from the MD simulation is essentially a stochastic distribution, whereas there are clear peaks - solvation shells - in the optimised structure.

It is essentially impossible to validate this cluster structure against a known method, due to the sheer size. However, on the smaller clus-
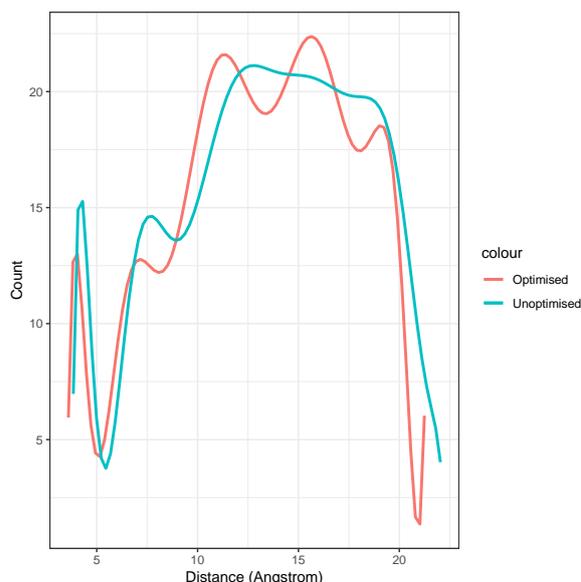
Figure 6.9: The change in the radial distribution function of a water molecule in the centre of a cluster of 202 water molecules, going from unoptimised geometry, taken from MD simulations, to the optimised, ALMO+RPAxd geometry. The latter shows distinct solvation shells appearing as peaks in the distribution.

ters considered in Figure 6.7, with up to 10 water molecules, we can compare against known CCSD-level geometries [329]. The RMSDs for these are on average 0.02 Angstrom, suggesting good agreement, and on the same order of the dimer calculations in Figure 6.5. Moreover, a single optimisation step on the 202-water cluster using the analytical gradients derived here took slightly under 20 hours on a single CPU core (a single-point energy took around 3 hours). This opens up the possibility of accurate, quantum-level optimisations of large molecular systems at a reasonable cost.

7 | MULTI-CONFIGURATIONAL ALMO SCF

The previous chapters have specialised in giving accurate descriptions of non-covalent interactions in large systems, culminating in analytic gradients for optimising their geometries. In this chapter, I extend this idea to classes of molecules for which a single Slater determinant does not provide an adequate description, while maintaining the absolute localisation of the molecular orbitals. This will allow us to treat, in particular, chemical reactions in explicit solvent. I begin by describing how the configuration interaction method can be applied to ALMOs, before giving a second-order method for the optimisation of the orbitals themselves. These are then combined into a multiconfigurational ALMO SCF procedure. Finally, I describe how this can be used to create a theory where a multiconfigurational subsystem is embedded in a much larger system of small molecules.

In a large proportion of chemical systems, a single Slater determinant - and thus any HF and ALMO SCF-based method - gives a good baseline description of the wavefunction [34]. This is equivalent to saying that, in the wider Fock space, a single 'point' or configuration contains by far the largest coefficient in the overall expansion of the wavefunction in the basis of determinants. In principle, however, we can only achieve the full solution in the limit of the full basis of configurations. In particular, if there are strong couplings between points in Fock space - that is strong static correlation [366] - ignoring these other configurations can lead to qualitatively incorrect results. This 'full expansion' approach is the principle behind the full configuration interaction (FCI) method.

Static correlation can be considered to be the coupling between points in Fock space, as opposed to dynamical correlation due to the electron-electron couplings.

Used alone, FCI typically still uses a single reference determinant - normally the HF solution - as the starting point for an expansion in terms of excitations [210]. Much as with the coupled-cluster type approaches described in chapter 3, this expansion is written in terms of excitations between the occupied and virtual molecular orbitals. In a truly multi-determinental wavefunction, however, the orbitals should be able to relax within that expansion. The simultaneous optimisation of both the FCI expansion and the orbitals would then yield a greatly improved solution with the Fock space. Unfortunately, this is impractical for all but the smallest systems due to the factorial scaling of the number of configurations.

Nonetheless, so-called multiconfigurational (MC) SCF methods are frequently used for any chemical problem with strong static correlation [367]. Such problems typically include anything where a bond is broken - so chemical reactions - or where a system has a number of energetically close states, such as in transition metal chemistry [368, 369]. Moreover, MCSCF wavefunctions can then be used as systematically generated multi-*reference* wavefunctions for the purposes of calculating dynamical correlation. Using each individual configuration as a reference for, for example, a perturbation theory calculation, we can in principle reach the exact solution to the quantum eigenvalue problem.

As per equation 2.16, each Slater determinant can be written in terms of a unitary transformation, described by the exponential of the antisymmetrised operator $\hat{\kappa}$, given in equation 2.17. This then dictates the form of the molecular orbitals. We can then construct 'excited' determinants from this by replacing what in HF is an occupied orbital with a virtual orbital, and expanding the wavefunction as a linear combination of determinants, via linear parameters c_I; note that in this chapter, uppercase latinate indices indicate determinants. The MCSCF wavefunction can thus be written as

$$|\Psi_{\text{MCSCF}}\rangle = \exp(-\hat{\kappa}) \sum_I c_I |\mathbf{I}\rangle \tag{7.1}$$

We then variationally optimise the energy expectation with respect to both $\boldsymbol{\kappa}$ and $\mathbf{c}$.

Generally, with canonical orthogonal orbitals, the two parameter sets are strongly coupled, resulting in large CI expansions being necessary, and a very difficult nonlinear optimisation problem [370, 371]. As a result, many restricted variants of MCSCF have been suggested, the most popular of which is the complete active space (CAS) method [372, 373]. In this, we partition the orbitals into three domains: *inactive*, doubly-occupied orbitals; *active*, 'fractionally' occupied orbitals; and *secondary*, unoccupied orbitals. A further restriction of this is the restricted active space approach [374, 375], where we further partition the active space into domains with fixed total occupation.

In reality, the active orbitals are singly occupied within some determinants, but are 'fractionally' occupied overall due to not being occupied in all determinants in the CI expansion.

Interestingly, there has been a considerable amount of evidence that using suitably optimised nonorthogonal orbitals can greatly reduce the lengths of the CI expansions [376–378]. Essentially, we are changing the representation of Fock space in such a way that connections between determinants are minimised, reducing the complexity and nonlinearity of the procedure. However, this is at the cost of a much more complicated orbital optimisation procedure due to the nonorthogonality. Therefore, using *a priori* localised orbitals, wherein we get the best of both worlds, promises to increase the scope of such calculations. Specifically by using ALMOs, we hope to provide a more effective multiconfigurational method for systems involving

many small molecules; this is particularly useful for chemical reactions, where solvent can play a significant role. In the following sections, I will derive an MC-ALMO SCF method, and discuss how it can be applied to study multiconfigurational systems *embedded* within a larger environment.

7.1 NON–ORTHOGONAL CONFIGURATION INTERACTION

In Chapters 2 and 3, we used a system of separately orthogonalised occupied and virtual orbital subspaces, utilising the redundancy within those subspaces. Now, however, we have a third subspace, namely the active space. In principle we could also separate this and orthogonalise independently, but it is easier to instead use a global space of bi-orthogonal orbitals. In this, we transform only the creation operators by the inverse metric:

$$a_{p\sigma}^{\dagger} = \left[\mathbf{s}^{-1}\right]_{pq} b_{q\sigma}^{\dagger} \tag{7.2}$$

These, together with the original annihilation operators, then satisfy the usual anticommutation relations, in particular that

$$\left\{a_{p\sigma}^{\dagger}, b_{q\gamma}\right\} = \delta_{pq}\delta_{\sigma\gamma}$$

To distinguish quantities in the bi-orthogonal basis, we will use an overbar. For example, the singlet excitation operators are given by

$$\overline{E}_{pq} = \sum_{\sigma} a_{p\sigma}^{\dagger} b_{q\sigma}$$

$$\overline{\Sigma}_{pqrs} = \overline{E}_{pq}\overline{E}_{rs} - \delta_{qr}\overline{E}_{ps}$$

and the molecular electronic Hamiltonian is simply

$$\hat{H} = \overline{h}_{pq}\overline{E}_{pq} + \frac{1}{2}\overline{g}_{pqrs}\overline{\Sigma}_{pqrs}$$

We can thus construct determinants in the Hilbert space in two ways:

$$|\mathbf{I}\rangle = \prod_{p \in I_{occ.}} b_{p\sigma_p}^{\dagger} |0\rangle$$

$$|\overline{\mathbf{I}}\rangle = \prod_{p \in I_{occ.}} a_{p\sigma_p}^{\dagger} |0\rangle \tag{7.3}$$

where $I_{occ.}$ denotes the occupations of determinant $|\mathbf{I}\rangle$. The two representations are then necessarily bi-orthogonal; that is, $\langle\mathbf{I}|\overline{\mathbf{J}}\rangle = \delta_{IJ}$.

The fundamental premise of configuration interaction (CI) is very simple. We expand the wavefunction as a linear combination of the

determinants, then variationally optimise the expansion coefficients, C_I, by minimising the energy. In the bi-orthogonal representation, this energy expression is given by

$$E = \frac{C_I \langle \mathbf{I}|\hat{H}|\overline{\mathbf{J}}\rangle \, \overline{C}_J}{C_I \langle \mathbf{I}|\overline{\mathbf{J}}\rangle \, \overline{C}_J} = \frac{C_I \overline{\sigma}_I}{C_I \overline{C}_I} \qquad (7.4)$$

where we have used the bi-orthogonality of the determinants in the denominator. In a naïve approach, one would calculate the full Hamiltonian matrix, then simply minimise the above expression. This is very costly, however, as the number of such matrix elements scales as the square of the number of determinants, which itself necessarily scales factorially with the number of electrons. Instead, we have defined in the second equality of equation 7.4 a sigma-vector, $\overline{\sigma}_I$, where the matrix elements are directly contracted with one set of coefficients. This immediately reduces the storage requirements to the square root of $N_{det.}$, but it also greatly improves the computational complexity, as I will demonstrate presently.

7.1.1 Alpha and beta strings

To preface the following developments, we need to introduce string-based CI, as first introduced in the '80s by Handy [379, 380], then later developed by Olsen and coworkers [374]. The reasons for this are two-fold: it informs how the calculations are performed in general, in such a manner as to be efficient; it will allow us to easily use ALMOs to reduce the scaling. The essence of the idea is to split the determinants into separate "strings" for the alpha and beta (up and down spin) electrons. Assuming that we wish to conserve the overall spin quantum number, excitations are only allowed within the alpha and beta subspaces, never between them. Then, we need only consider interactions between string pairs.

Let I_α be the list of occupied alpha-spin orbital indices, $\{i, j, \ldots, k\}$, in a given determinant, $|\mathbf{I}\rangle$. The alpha string, $\alpha(I_\alpha)$ is then the ordered product of creation operators corresponding to these spin orbitals; similarly, a beta string is such an ordered list but for the beta subspace. The determinant is then represented by

$$|\mathbf{I}\rangle = |\alpha(I_\alpha)\beta(I_\beta)\rangle \qquad (7.5)$$

In direct CI methods, using the sigma-vector approach, we need an efficient way to determine which determinants interact with each other through the Hamiltonian. If we were to consider each determinant pair directly, this would entail $N_{det.}^2$ comparisons. With strings, however, we only need consider which string pairs interact, reducing the indexing to effectively at most $2N_{det.}$. We can then use these string
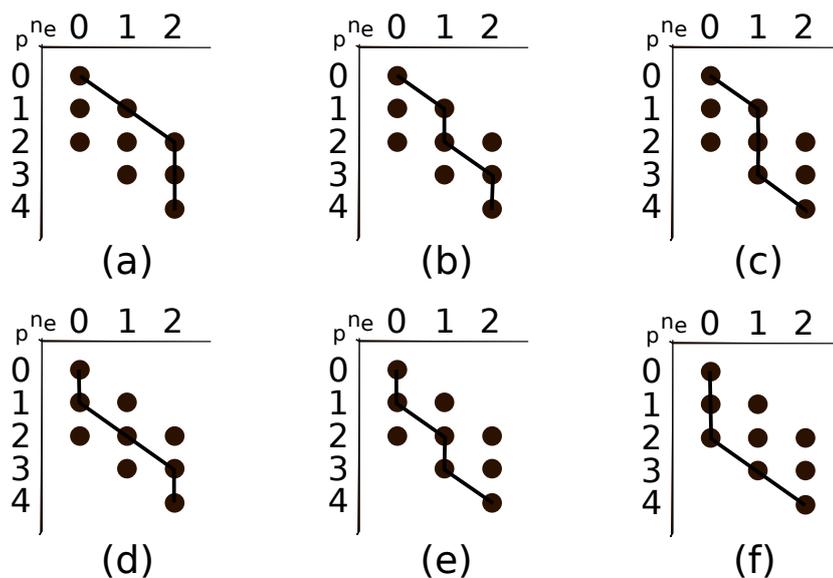
Figure 7.1: All possible alpha/beta strings in the full CI space of $n_e = 2$ electrons in 4 orbitals, p. Vertical and diagonal lines indicate unoccupied and occupied orbitals, respectively.

indices to avoid calculating negligible matrix elements, and the coefficients are simply a phase factor of $+1$ or -1, depending on whether there are an even or odd number of permutations in normal ordering.

This scheme requires a simple and rapid way of addressing the strings. The suggestion of Handy [379] was the graph-based approach, shown in Figure 7.1 for a simple system where we have $n_e = 2$ electrons (horizontal axis) over $n_p = 4$ orbitals (vertical axis). All strings are then enumerated, as shown, by finding all valid paths from the $(0,0)$ vertex to the (n_e, n_p) vertex. A further computational advantage of this is that if we wish to restrict our active space, we simply remove vertices or paths from this generation procedure. For example, if in the above problem we wish to perform a restricted active space (RAS) calculation where there must be at least one electron in either orbital 1 or 2, we would remove the $(0, 2)$ vertex, corresponding to zero electrons in the first two orbitals. In this case, that would only eliminate string (f) from Figure 7.1, but in general will greatly simplify a problem. We will also use this when exploiting locality later on.

The addresses of the strings are then determined using reverse-lexical ordering [34] in the following manner. The vertices are assigned weights equal to the number of valid routes from the origin to that vertex; the diagonal edge (occupation) weights are then assigned such that they join the vertex weights, with the vertical edges weighted as zero. This is shown in Figure 7.2. For example, we can

Other orderings are possible, but we choose the convention of Olsen.
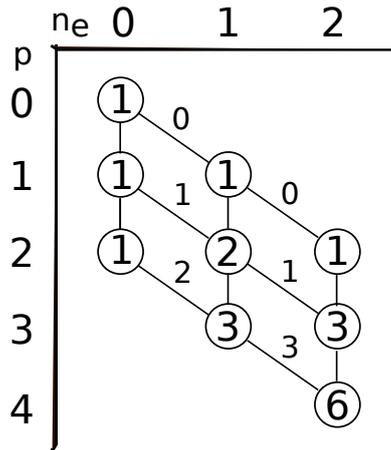
Figure 7.2: The network representation of the full CI space of $n_e = 2$ electrons in 4 orbitals, p, using reverse-lexical ordering. The vertex weights indicate the number of valid paths that pass through that vertex, and the arc weights connect these vertex weights; vertical arcs are always zero. The address of a string (determinant) is then calculated by summing the arc weights in the string path.

Table 7.1: The corresponding determinant and reverse-lexical address of each string given in Figure 7.1, as determined using Figure 7.2. The addresses of the connecting determinants - i.e. those that differ through a single excitation - are given in the final column.

String	Determinant	Address	Connections
(a)	$\lvert \Psi_0 \rangle$	0	$1, 2, 3, 4$
(b)	$\lvert \Psi_2^3 \rangle$	1	$0, 2, 3, 5$
(c)	$\lvert \Psi_2^4 \rangle$	3	$0, 1, 4, 5$
(d)	$\lvert \Psi_1^3 \rangle$	2	$0, 1, 4, 5$
(e)	$\lvert \Psi_1^4 \rangle$	4	$0, 2, 3, 5$
(f)	$\lvert \Psi_{12}^{34} \rangle \, / \, \lvert \Psi_{12}^{43} \rangle$	5	$1, 2, 3, 4$

reach $(1,2)$ via $(0,1)$ or $(1,1)$, so its vertex weight is 2. The diagonal edge joining it to $(0,1)$, which there is only one path to, is therefore $2 - 1 = 1$. Finally, the address of a string is calculated as the sum of all edge weights in the graph of that string. For example, string (e) in Figure 7.1 has diagonal edges with weights 1 and 3, so is assigned the address 4.

The addresses for each string in Figure 7.1 are given in Table 7.1, along with the addresses of the strings it connects to through a single excitation. These can easily be read off the graphs by the positions of the diagonal edges - e.g. string (b) does not connect to (e), as the former has diagonals in rows 1 and 3, while the latter has them in rows 2 and 4. In this toy problem, all of the strings are connected by

double excitation, so we do not show them here. In general, however, these connections can be enumerated in the same manner.

Further savings are then possible by considering the structure of the sigma vector in terms of these strings. We rewrite it as follows:

$$\overline{\sigma}_I = \langle \alpha(I_\alpha)\beta(I_\beta)|\hat{H}|\alpha(\overline{J}_\alpha)\beta(\overline{J}_\beta)\rangle \, \overline{C}_J \tag{7.6}$$

The Hamiltonian is written in terms of singlet excitation operators, either single excitations or doubles, which then allows us to split the sigma vector into three parts: the first and second involving only alpha and beta strings, respectively, and the third comprising the two-electron terms where one excitation is of alpha and one of beta spin. Thus, if each string is connected to on average k other strings, then for N_σ strings each term needs approximately $N_\sigma k$ comparisons, as opposed to $N_{det.}^2$, which in larger active spaces is a considerable saving.

7.1.2 Evaluating the ALMO-CI sigma vector

Firstly, we follow the usual convention of creating an effective one-electron Hamiltonian from the usual core Hamiltonian and the exchange term, which has matrix elements

$$o_{pq} = h_{pq} - \frac{1}{2} g_{prrq} \tag{7.7}$$

such that the Hamiltonian in the bi-orthogonal basis becomes

$$\hat{H} = \overline{o}_{pq}\overline{E}_{pq} + \overline{g}_{pqrs}\overline{E}_{pq}\overline{E}_{rs}$$

Equation 7.6 can thus be split into one- and two-excitation terms, with the former only appearing in the like-spin case:

$$[\sigma_I]_\alpha^{(1)} = \langle \alpha(I_\alpha)|\overline{E}_{pq}^\alpha|\alpha(\overline{J}_\alpha)\rangle \, \overline{o}_{pq}\overline{C}_J \tag{7.8}$$

$$[\sigma_I]_{\alpha\alpha}^{(2)} = \frac{1}{2} \langle \alpha(I_\alpha)|\overline{E}_{pq}^\alpha\overline{E}_{rs}^\alpha|\alpha(\overline{J}_\alpha)\rangle \, \overline{g}_{pqrs}\overline{C}_J$$

$$= \frac{1}{2} \langle \alpha(I_\alpha)|\overline{E}_{pq}^\alpha|\alpha(\overline{K}_\alpha)\rangle \langle \alpha(K_\alpha)|\overline{E}_{rs}^\alpha|\beta(\overline{J}_\alpha)\rangle \, \overline{g}_{pqrs}\overline{C}_J \tag{7.9}$$

$$[\sigma_I]_{\alpha\beta}^{(2)} = \langle \alpha(I_\alpha)|\overline{E}_{pq}^\alpha|\alpha(\overline{J}_\alpha)\rangle \langle \beta(I_\beta)|\overline{E}_{rs}^\beta|\beta(\overline{J}_\beta)\rangle \, \overline{g}_{pqrs}\overline{C}_J \tag{7.10}$$

where in $\sigma_{\alpha\alpha}^{(2)}$ we have inserted the resolution of the identity, assuming the full CI space is being used for the problem. From these equations, we clearly need only consider the effect of the excitation operators of a given spin on strings of that same spin. We can considerably simplify our calculations by doing so in turn for the inactive, active, and secondary subspaces.

For the inactive orbitals, all strings necessarily contain the creation operator for that orbital. Thus, $b_{j\alpha}$ acting on any alpha string, by

bi-orthogonality, gives a determinant with j now unoccupied plus a phase determined by the parity of that orbital. If $a_{i\alpha}^{\dagger}$ then acts on it, the determinant will annihilate from double occupation of the spin-orbital, unless $i = j$. The one-electron element involving the inactive-inactive terms thus reduces to

$$\langle \alpha(I_\alpha) | \alpha(\bar{J}_\alpha) \rangle \, \overline{o}_{ii} \overline{C}_J = \overline{o}_{ii} \overline{C}_I$$

The overlap element evaluates to zero unless $I = J$, as the bra and ket are in bi-orthogonal representations. Thus, we can define a core energy as the combination of the (equivalent) alpha and beta one-electron sigma vector terms in the inactive-inactive subspace.

$$E_{core} = 2\overline{h}_{ii} + 2\overline{g}_{iijj} - \overline{g}_{ijji} \tag{7.11}$$

Note that this does not include exchange contributions involving active (or secondary) orbitals, despite these appearing in inactive-inactive terms. These are not neglected in the final result, it is merely convenient to group them with the other terms involving these orbitals.

Similarly, for the secondary orbitals, these are unoccupied in every string, and thus automatically annihilate when in the de-excitation index of the excitation operator. If the secondary orbital is in the excitation index, then this generates a new string which is not in the CI space, which will thus be orthogonal to whichever string is in the bra. Thus, no secondary interactions are included in the sigma vector.

Therefore, the only non-constant and non-zero contribution to the sigma vector is from inactive-active and active-active interactions. The former appear only through the inactive-inactive excitations earlier: if we create or destroy an inactive electron, without replacement, the string will not be one in our CI space. An active-active excitation operator will in general connect exactly one string at J to the given string at I. Thus, in the string list where the addresses of all strings connected to I are stored, we also store the active orbital pair that connects them. When computing the sigma vector, we then only need to match up the matrix element and phase to each term. We collect the active-active terms from equation 7.8 and the inactive-active terms from equations 7.9 and 7.10 into an effective Fock matrix as follows:

$$\overline{f}_{rs} = \overline{h}_{rs} + 2\overline{g}_{rsii} - \overline{g}_{riis} \tag{7.12}$$

Note that E_{core} can then effectively be written as the trace over inactive orbitals of equation 7.12; as only diagonal terms are needed though, it is more convenient to only compute $\overline{f}$ for the active orbitals.

Putting all the above together, the final sigma vector, with the splitting by spin left out for simplicity, is then given by

$$\overline{\sigma}_I = E_{core}\overline{C}_I + \sum_{r,s\in active} \left\{ \overline{f}_{rs} \langle \mathbf{I} | \overline{E}_{rs} | \bar{\mathbf{J}} \rangle + \frac{1}{2} \sum_{t,u\in active} \overline{g}_{rstu} \langle \mathbf{I} | \overline{\Sigma}_{rstu} | \bar{\mathbf{J}} \rangle \right\} \overline{C}_J \tag{7.13}$$

7.1.3 Solving the CI equations

The core energy and effective Fock matrix of equations 7.11 and 7.12 can readily be evaluated in the atomic orbital basis. The transformation can be written entirely in terms of the absolutely localised orbital coefficients by first defining an effective density, in analogy to that of equation 2.45:

$$[\mathbf{P}_{\text{eff.}}]_{\mu\nu} = T^{\mu\cdot}_{\cdot i} \left[\mathbf{s}^{-1}\right]_{ip} C^{\cdot\nu}_{p\cdot} \qquad (7.14)$$

Note that this differs from the original density by the extension of the right-hand term over *all* coefficients, resulting in a non-symmetric density. This is a side-effect of using the bi-orthogonal orbitals as opposed to the symmetrically orthogonalised ones used in Chapter 2. The effective Fock matrix (and thus also the core energy) can then be written as

$$\bar{f}_{rs} = C^{r\cdot}_{\cdot\mu} \{H_{\mu\nu} + G_{\mu\nu}(\mathbf{P}_{\text{eff.}})\} \left[\mathbf{Cs}^{-1}\right]_{\nu s} = C^{r\cdot}_{\cdot\mu} F_{\mu\nu}(\mathbf{P}_{\text{eff.}}) \left[\mathbf{Cs}^{-1}\right]_{\nu s} \qquad (7.15)$$

That is, the AO Fock matrix can be constructed exactly as it was in the standard ALMO SCF routine, using the same linear-scaling procedures, but with a different density. In particular, this means the density-fitting and local exchange procedures can be used.

The most time consuming term in equation 7.13 then becomes the transformation of the two-electron integrals for the active-active interactions. However, the number of active electrons is in general considerably less than the number of total orbitals, such that an integral-direct procedure with the usual AO integral screening will be rapid. This could also utilise density fitting in the usual manner. In addition, we only need to calculate the transformed integrals for which there are associated non-zero CI density elements to contract them with, further reducing the cost. The formal scaling will still be $\mathcal{O}(N^4)$ (or N^3 for density fitting), but with a very small pre-factor. In future, we could try to reduce this scaling by more severe screening of the AO integrals, as is done in multipole-based integration procedures [113, 116].

Solving the CI equations then follows the standard procedure for canonical string-based CI. Instead of directly diagonalising the Hamiltonian, which would be very expensive, we instead form the sigma vector directly and then iteratively find the lowest eigenvalues and eigenvectors. The most popular approach is the Davidson-Liu iterative subspace procedure, details of which can be found elsewhere [370, 381, 382].

Table 7.2: The corresponding determinant and reverse-lexical address of each string given in Figure 7.1, with the pruned excitations according to localisation. In this example, each orbital is on a different fragment, with fragments 1 and 3 well separated from 2 and 4.

String	Determinant	Address	Connections
(a)	$\lvert \Psi_0 \rangle$	0	$2, 3$
(b)	$\lvert \Psi_2^3 \rangle$	1	—
(c)	$\lvert \Psi_2^4 \rangle$	3	$0, 5$
(d)	$\lvert \Psi_1^3 \rangle$	2	$0, 5$
(e)	$\lvert \Psi_1^4 \rangle$	4	—
(f)	$\lvert \Psi_{12}^{34} \rangle \,/\, \lvert \Psi_{12}^{43} \rangle$	5	$2, 3$

7.1.4 String pruning using localisation

Apart from the construction of the effective Fock matrix in equation 7.15, we have not exploited the localisation of the original molecular orbitals. The refactorisation of the non-orthogonal CI equations we have presented is very efficient, but does not reduce the overall size of the full CI space. In the canonical case, we cannot *a priori* determine which orbitals will have negligible excitation amplitudes, due to the delocalised nature of the orbitals. ALMOs on distant fragments, on the other hand, are generally well separated, such that excitations between them are likely to not contribute.

The string-based algorithm offers an efficient opportunity to exploit this, reducing the size of the CI expansion. If the Hamiltonian element between two strings can be neglected, then we can immediately use this by pruning connections between strings with excitations between these orbitals. Suppose that we have an active space comprising two electrons between four orbitals, as in Figure 7.1, but this time each orbital is located on a different fragment. Further, suppose fragments 1 and 3 are well separated from 2 and 4. So while strings (a) and (b) only differ in one occupation line (rows 1 and 2, versus 1 and 3, respectively), the difference consitutes an excitation from 2 to 3, which is neglected; this connection can therefore also be ignored. In this manner, Table 7.1 reduces to Table 7.2. The total number of connections has reduced from 20 to 8.

An alternative, computationally simple, way of generating the strings that connect to any given string, is to use a pruned graph according to whether certain orbitals are to be neglected. For example, all the strings connecting to (a) in Figure 7.1 are generated from the grid given in Figure 7.3. Following the rules outlined earlier, it is only possible to create strings (c) and (d) from this graph, which is in agreement with Table 7.2. The addresses are then looked up by summing
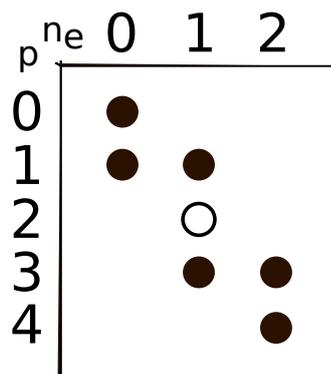
Figure 7.3: The pruned graph for generating strings connecting to string (a) in Figure 7.1. The hollow vertex is unidirectional; that is, an edge must leave in the same direction it enters.

the edge weights of the *complete* network, i.e. those in Figure 7.2. The pruned graphs can be rapidly generated according to a hash table of screened orbitals, in the same manner we would screen by acceptable excitations described earlier.

All that remains is to determine the screening criterion. This is effectively an extension of the two-electron integral screening studied in Chapter 4.1, but over the full Hamiltonian. To study the asymptotics of the matrix elements, it is convenient to return to the determinental representation and use the Slater-Condon rules [383, 384]. These are, given some reference determinant $|\Phi\rangle$ and assuming p, q, r, s are all active orbitals:

$$\langle\Phi|\hat{H}|\Phi_p^q\rangle = \bar{f}_{pq}$$
$$\langle\Phi|\hat{H}|\Phi_{pq}^{rs}\rangle = \bar{g}_{pqrs} - \bar{g}_{prqs}$$

From equation 4.5, we know that the two electron integrals are bounded by $2R^{-1}$, where R is the distance between the vectors from centre one to centre two, and centre three to centre four. Similarly, the effective Fockian of equation 7.12 contains these two-electron integrals. If desired, the net contribution of this density can be estimated by the infinity norm of the relevant block of the density, as per the screening in equation 4.7. It is simpler however to just use the distance dependence.

The only remaining term then is the core Hamiltonian in the effective Fock term, and seeing how this depends on the orbital separation. This contains two parts: the kinetic energy, and the nuclear-electronic attraction. The latter term is of the same form as the two-electron integral, but with the distance being to the nuclear centre. The resulting

asymptotic behaviour is thus as R^{-1} but where R is now the distance between the atomic centres. The kinetic energy term is of the form

$$T_{pq} = -\frac{1}{2} \langle \nabla \phi_p | \nabla \phi_q \rangle \sim -\frac{1}{2} K s_{pq}$$

where K is some unknown constant. This is due to the derivative of a Gaussian distribution being a linear combination of Gaussians [34], such that the integral reduces to a linear combination of overlap integrals. The product of two Gaussians results in a transformed Gaussian as follows

$$\phi_p \phi_q = \exp\left(-\mu R_{PQ}^2\right) \exp(-\alpha r^2)$$

where μ and α are determined as combinations of the exponents of the original distributions. The r dependence then integrates out to a constant depending only on α, leaving the distance dependence as $\exp(-\mu R_{PQ}^2)$.

The exponential dependence of the kinetic energy will tail off much more quickly than the $1/R$ dependence of the Coulombic terms, and so can effectively be ignored, unless the atomic orbitals are *very* diffuse. Thus, we screen connections by simply considering whether the separation between centres is larger than some threshold value, which we could estimate using the above, or determine by trial and error. The extents of the atomic orbitals are determined as part of the domain selection algorithm in the local exchange calculation, so these could be checked against on a fragment to see if they are larger than the distance threshold. If they are, the overlap element may be non-negligible. Alternatively, as the overlap metric has already been calculated, we could compare values directly, or more coarsely by considering the norm of fragment-fragment blocks of the overlap.

7.2   ORBITAL OPTIMISATION

In MCSCF, we need to simultaneously optimise both the CI expansion, and the orbitals. The latter is achieved through considering the orbital rotations, using an antisymmetrised rotation operator

$$\hat{\kappa}^a = \kappa_{pq}^a E_{pq}$$

with $\kappa_{qp}^a = -\kappa_{pq}^a$. The rotation is then applied to a given creation operator as

$$c_{p\sigma}^\dagger = \exp\left(-\hat{\kappa}^a\right) b_{p\sigma}^\dagger \exp\left(\hat{\kappa}^a\right)$$

such that the overlap metric is conserved:

$$\left\{ c_{p\sigma}^\dagger, c_{q\sigma} \right\} = \exp\left(-\hat{\kappa}^a\right) \left\{ b_{p\sigma}^\dagger, b_{p\sigma} \right\} \exp\left(\hat{\kappa}^a\right)$$

$$= \exp\left(-\hat{\kappa}^a\right) s_{pq} \delta_{pq} \exp\left(\hat{\kappa}^a\right) = s_{pq} \delta_{pq}$$

where in the transformation of the anticommutator we note that successive exponentials cancel, leaving the form given.

However, in the case of absolutely localised molecular orbitals, where between-fragment redundancies will be removed later by constraining the AO coefficients, considering only metric-conserving transformations would not lead to the best result. In the context of other non-orthogonal orbital optimisations, Olsen [371] introduced the idea of combining the above with a symmetric rotation operator, $\hat{\kappa}^s$, with the same form as $\hat{\kappa}^a$ but with symmetric matrix elements. We may still set diagonal elements to zero, though, as these change only the norm of the orbitals. Our transformed creation operators are thus now

$$c_{p\sigma}^\dagger = \exp\left(-\hat{\kappa}^a\right)\exp\left(-\hat{\kappa}^s\right) b_{p\sigma}^\dagger \exp\left(\hat{\kappa}^s\right)\exp\left(\hat{\kappa}^a\right) \qquad (7.16)$$

Let us consider each transformation in turn. The exponentiation can be rewritten using the Baker-Campbell-Hausdorff (BCH) expansion in terms of commutators:

$$\exp(-\hat{\kappa})b_{p\sigma}^\dagger\exp(\hat{\kappa}) = b_{p\sigma}^\dagger + \left[b_{p\sigma}^\dagger, \hat{\kappa}\right] + \frac{1}{2}\left[\left[b_{p\sigma}^\dagger, \hat{\kappa}\right], \hat{\kappa}\right] + \cdots$$

The first commutator is evaluated as

$$\begin{aligned}
\left[b_{p\sigma}^\dagger, \hat{\kappa}\right] &= \kappa_{rs}\sum_\gamma \left[b_{p\sigma}^\dagger, b_{r\gamma}^\dagger b_{s\gamma}\right] \\
&= \kappa_{rs}\left(\left\{b_{p\sigma}^\dagger, b_{r\sigma}^\dagger\right\} b_{s\sigma} - b_{r\sigma}^\dagger\left\{b_{p\sigma}^\dagger, b_{s\sigma}\right\}\right) \\
&= -\left(\boldsymbol{\kappa s}\right)_{rp} b_{r\sigma}^\dagger
\end{aligned}$$

Inserting this result into each term of the expansion then yields

$$\begin{aligned}
\exp(-\hat{\kappa})b_{p\sigma}^\dagger\exp(\hat{\kappa}) &= b_{p\sigma}^\dagger - \left(\boldsymbol{\kappa s}\right)_{rp} b_{r\sigma}^\dagger - \frac{1}{2}\left(\boldsymbol{\kappa s}\right)_{rp}\left[b_{r\sigma}^\dagger, \hat{\kappa}\right] + \cdots \\
&= b_{p\sigma}^\dagger - \left(\boldsymbol{\kappa s}\right)_{rp} b_{r\sigma}^\dagger + \frac{1}{2}\left(\boldsymbol{\kappa s}\right)_{rp}^2 b_{r\sigma}^\dagger \\
&= b_{r\sigma}^\dagger\left[\exp\left(-\boldsymbol{\kappa s}\right)\right]_{rp}
\end{aligned}$$

Doing this transformation twice then gives a more useful form of the operators in equation 7.16:

$$c_{p\sigma}^\dagger = b_{r\sigma}^\dagger\left[\exp\left(-\boldsymbol{\kappa}^a\boldsymbol{s}\right)\exp\left(-\boldsymbol{\kappa}^s\boldsymbol{s}\right)\right]_{rp} \qquad (7.17)$$

7.2.1 Redundancies

In other non-orthogonal orbital approaches [376–378], the inactive and secondary orbitals can be constrained to be orthogonal without affecting the resulting wavefunction. This results in the inactive-inactive and secondary-secondary antisymmetric parameters being redundant. In our case, however, this only applies to the *on-fragment* subspaces.
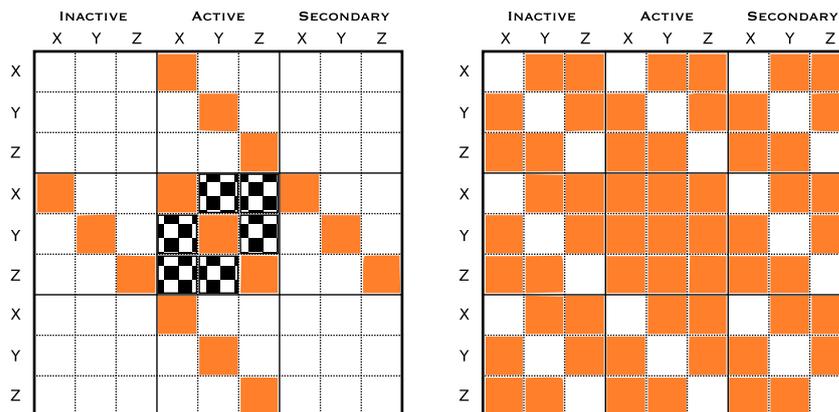
Figure 7.4: The parameter matrices $\boldsymbol{\kappa}^a$ (left) and $\boldsymbol{\kappa}^s$ (right), partitioned first by space (inactive, active, secondary), then by each of the three fragments (X, Y, Z). Non-redundant blocks are orange, while redundant blocks are white. The checkerboard blocks in the antisymmetric matrix are redundant if the active orbitals remain localised, but non-redundant if we wish to fully mix the active orbitals as described in the main text.

If we wish to apply the absolutely localised constraint in the atomic-orbital basis, we necessarily need *all* mixed-fragment blocks of both sets of rotation parameters, all active-active on-fragment symmetric parameters, and all on-fragment antisymmetric parameters involving active orbitals, to be non-redundant. This is an inexorably large number of parameters.

The significance of the number of non-redundant parameters is that the symmetric parameters lead to terms in the gradients that are of higher rank than two - that is, higher than the rank of the Hamiltonian. These terms are what lead to non-orthogonal orbital optimisations being much more computationally intensive than canonical ones [371]. We can alleviate the problem for the antisymmetric parameters by moving the localisation constraint directly to the rotation parameters. There is necessarily no mixing between orbitals on different fragments - this automatically translates to the mixed-fragment blocks of the antisymmetric parameters being fixed to be zero. The overlap between these mixed-fragment orbitals will in general change, however; we will instead have to try and use the localised orbitals to find savings in the evaluations of the gradients. Figure 7.4 shows an example of the redundancy structure of both sets of parameters for a system with three fragments.

One additional possible consideration is that a more flexible, but not absolutely localised throughout, solution could be found by allowing the active orbitals between fragments to mix. In the context of an embedding procedure, to be described in the next section, this

is equivalent to considering a full calculation on whatever subspace of fragments is defined to be 'active', in the mean field of other fragments treated at the ALMO-SCF level. Doing this would only complicate the problem insofar as adding the active-active parameters back to $\hat{\kappa}^a$, which will typically be minimal additional cost, in comparison to the aforementioned symmetric terms.

7.2.2 Second–order optimisation

The overall MC-ALMO wavefunction is given by equation 7.1, but with the orbital rotation parameters as above, and the CI coefficients in the bi-orthogonal basis as per the previous section. For convenience, we rewrite the CI parameterisation as differences, $\boldsymbol{\delta}$, from an initial guess, Ψ_0. Thus, the wavefunction is written as

$$|\Psi\left(\boldsymbol{\kappa}^a, \boldsymbol{\kappa}^s, \boldsymbol{\delta}\right)\rangle = \exp\left(-\hat{\kappa}^a\right)\exp\left(-\hat{\kappa}^s\right)\{|\Psi_0\rangle + \delta_I\left(|\mathbf{I}\rangle - |\Psi_0\rangle\langle\Psi_0|\mathbf{I}\rangle\right)\}$$

(7.18)

with energy

$$E\left(\boldsymbol{\kappa}^a, \boldsymbol{\kappa}^s, \boldsymbol{\delta}\right) = \frac{\langle\Psi(\boldsymbol{\delta})|\exp(-\hat{\kappa}^s)\exp(\hat{\kappa}^a)\hat{H}\exp(-\hat{\kappa}^s)\exp(-\hat{\kappa}^a)|\Psi(\boldsymbol{\delta})\rangle}{\langle\Psi(\boldsymbol{\delta})|\exp(-2\hat{\kappa}^s)|\Psi(\boldsymbol{\delta})\rangle}$$

(7.19)

Note that the symmetric transformation, which is not metric-conserving, changes the overall norm of the wavefunction.

We assume that the guess, $\Psi_0 = \Psi(\mathbf{0}, \mathbf{0}, \mathbf{0})$ is normalised, for convenience. The orbital optimisation procedure is typically performed via a second-order expansion of the energy, requiring us to know the orbital gradient and Hessian, which we will determine in the next subsection. The MCSCF procedure then comprises uncoupled macro-iterations, alternating between solving for the orbital parameters via a Newton-Raphson procedure, and solving for the CI vectors using the Davidson-Liu algorithm (each of which are termed the micro-iterations). The assumption that the different parameters are uncoupled is not necessarily accurate, and can result in convergence issues [385], but greatly simplifies the problem; specifically, it removes any cross-terms from the gradients, and allows us to separate the two micro-iteration procedures.

The orbital gradient

The gradient with respect to the antisymmetric parameters is simpler than that of the symmetric parameters, for two reasons. Firstly, it only appears in the numerator of equation 7.19; secondly, it can be expanded using the BCH formula, as was used earlier for the creation operators. The gradient is then necessarily the first-order term, i.e. the first commutator, in this expansion, evaluated at the reference point -

i.e. where the wavefunction is normalised such that the denominator is irrelevant. Note that, as $\kappa_{qp} = -\kappa_{pq}$, there are two terms in the gradient. Exchange of the indices in the excitation operator necessarily changes the sign though, by the anticommutation relations, such that the effect is simply to introduce a factor of two:

$$\frac{\partial E}{\partial \kappa_{pq}^a} = 2 \langle \Psi_0 | \left[E_{pq}, \hat{H} \right] | \Psi_0 \rangle \tag{7.20}$$

For the symmetric parameters, we must directly expand the exponential operators as follows (obfuscating the other parameters for clarity):

$$E = \langle \Psi_0 | \left(1 - \hat{\kappa}^s + \frac{1}{2} [\hat{\kappa}^s]^2 + \cdots \right) \hat{H} \left(1 - \hat{\kappa}^s + \frac{1}{2} [\hat{\kappa}^s]^2 + \cdots \right) | \Psi_0 \rangle$$

$$\times \left[\langle \Psi_0 | \left(1 - \hat{\kappa}^s + \frac{1}{2} [\hat{\kappa}^s]^2 + \cdots \right)^2 | \Psi_0 \rangle \right]^{-1}$$

$$= \left[E_0 - \langle \Psi_0 | \{ \hat{\kappa}^s, \hat{H} \} | \Psi_0 \rangle + \frac{1}{2} \langle \Psi_0 | \{ \hat{\kappa}^s, \{ \hat{\kappa}^s, \hat{H} \} \} | \Psi_0 \rangle + \mathcal{O}(\kappa^s)^3 \right]$$

$$\times \left[1 - 2 \langle \Psi_0 | \hat{\kappa}^s (1 + \hat{\kappa}^s) | \Psi_0 \rangle + \mathcal{O}(\kappa^s)^3 \right]^{-1}$$

$$= E_0 - \langle \Psi_0 | \{ \hat{\kappa}^s, \hat{H} \} - 2E_0 \hat{\kappa}^s | \Psi_0 \rangle + \frac{1}{2} \langle \Psi_0 | \{ \hat{\kappa}^s, \{ \hat{\kappa}^s, \hat{H} \} \} + 4E_0 [\hat{\kappa}^s]^2 | \Psi_0 \rangle$$

$$+ 2 \langle \Psi_0 | \{ \hat{\kappa}^s, \hat{H} \} | \Psi_0 \rangle \langle \Psi_0 | \hat{\kappa}^s | \Psi_0 \rangle + 4E_0 \langle \Psi_0 | \hat{\kappa}^s | \Psi_0 \rangle^2 + \mathcal{O}(\kappa^s)^3$$

where in the last line we have used the series $[1 - x]^{-1} = 1 + x + x^2 + \cdots$. As the parameters are symmetric, and assuming the reference wavefunction is real, the anticommutator reduces to $2\hat{\kappa}^s \hat{H}$. The symmetry also means that the gradient must contain the reverse-index term; to encompass this, we define

$$E_{pq}^+ = E_{pq} + E_{qp}$$

such that the gradient is given by

$$\frac{\partial E}{\partial \kappa_{pq}^s} = -2 \langle \Psi_0 | E_{pq}^+ \left(\hat{H} - E_0 \hat{I} \right) | \Psi_0 \rangle \tag{7.21}$$

We now return to the bi-orthogonal basis, where the antisymmetric orbital gradient is simply the usual expression, as given previously in equation 2.20. To fully exploit this, we thus define one- and two-particle density matrices as

$$\overline{D}_{pq} = \langle \Psi_0 | \overline{E}_{pq} | \Psi_0 \rangle$$
$$\overline{d}_{pqrs} = \langle \Psi_0 | \overline{\Sigma}_{pqrs} | \Psi_0 \rangle \tag{7.22}$$

Using the commutation relations

$$\left[\overline{E}_{pq}, \overline{E}_{rs} \right] = \delta_{qr} \overline{E}_{ps} - \delta_{ps} \overline{E}_{rq}$$
$$\left[\overline{E}_{pq}, \overline{\Sigma}_{rstu} \right] = \delta_{qr} \overline{\Sigma}_{pstu} + \delta_{qt} \overline{\Sigma}_{rspu} - \delta_{ps} \overline{\Sigma}_{rqtu} - \delta_{pu} \overline{\Sigma}_{rstq}$$

we can thus rewrite the antisymmetric gradient as

$$
\begin{aligned}
\frac{\partial E}{\partial \kappa_{pq}^a} &= 2s_{pr} \langle \Psi_0 | \left[\overline{E}_{rq}, \hat{H} \right] | \Psi_0 \rangle \\
&= 2s_{pr} \overline{h}_{tu} \left[\delta_{qt} \overline{D}_{ru} - \delta_{ru} \overline{D}_{tq} \right] \\
&\quad + s_{pr} \overline{g}_{tuvw} \left[\delta_{qt} \overline{d}_{ruvw} + \delta_{qv} \overline{d}_{turw} - \delta_{ru} \overline{d}_{tqvw} - \delta_{rw} \overline{d}_{tuvq} \right] \\
&= 2\overline{F}_{pq} - 2\overline{F}_{qp}
\end{aligned}
$$

In the above, we have defined a second effective Fockian as

$$
\overline{F}_{pq} = \left[\mathbf{h}\mathbf{D}\mathbf{s} \right]_{qp} + s_{pr} \overline{g}_{quvw} \overline{d}_{ruvw} \tag{7.23}
$$

where we have exploited the permutational symmetry of the two-electron integrals and two-body density. By writing it in this form, we can once again apply all the machinery developed for the standard ALMO-SCF procedure, by expanding back into the ALMO basis and exploiting the block constraints. This is a reflection of the constraints being applied directly to the parameter matrix.

The symmetric gradient, equation 7.21, in principle looks much more daunting. However, we rewrite it as follows:

$$
\begin{aligned}
\frac{\partial E}{\partial \kappa_{pq}^s} &= 2E_0 s_{pr} \left(\overline{D}_{rq} + \overline{D}_{qr} \right) - 2 \langle \Psi_0 | E_{pq}^+ \hat{H} | \Psi_0 | \\
&= 2E_0 \left\{ \mathbf{s}, \overline{\mathbf{D}} \right\}_{pq} - 2 \langle \Psi_0 | E_{pq}^+ | \boldsymbol{\sigma} \rangle
\end{aligned}
$$

The order of the micro-iterations thus becomes important. Specifically, by performing the CI iterations first, we will have already generated the sigma vector, and thus we only need to consider its connection through single excitations to the reference wavefunction. This is still somewhat problematic, however, as it extends the connections from double to triple excitations in total. These connections are rapidly generated using the string-based approaches noted earlier, with the resulting matrix elements being readily screened using the locality conditions outlined in the previous section. The transition density matrix, $\langle \Psi_0 | E_{pq}^+ | \boldsymbol{\sigma} \rangle$, effectively corresponds to a three-body density matrix, though, and so even with heavy screening will likely be the bottleneck in the calculation. This quantity needs to be considered in more detail in the future to try and alleviate this cost.

The orbital Hessian

The antisymmetric-antisymmetric Hessian terms are simply given by the second-order commutator in the BCH expansion. This term is

$$
E^{(2)} = \frac{1}{2} \kappa_{pq}^a \kappa_{rs}^a \langle \Psi_0 | \left[E_{pq}^-, \left[E_{rs}^-, \hat{H} \right] \right] | \Psi_0 \rangle
$$

Recall that the permutation operator $\hat{P}(a, b)$ swaps the index groups a and b.

such that the derivatives are given by

$$
\begin{aligned}
\frac{\partial^2 E}{\partial \kappa_{pq}^a \partial \kappa_{rs}^a} =& \frac{1}{2} \left(1 + \hat{P}(pq, rs)\right) \langle \Psi_0 | \left[E_{pq}^-, \left[E_{rs}^-, \hat{H}\right]\right] |\Psi_0\rangle \\
=& \frac{1}{2} \left(1 + \hat{P}(pq, rs)\right) s_{pt} s_{ru} \langle \Psi_0 | \left[\overline{E}_{tq}^-, \left[\overline{E}_{us}^-, \hat{H}\right]\right] |\Psi_0\rangle \\
=& \frac{1}{2} \left(1 - \hat{P}(pt, qu)\right) \left(1 - \hat{P}(rv, sw)\right) \left(1 + \hat{P}(ptq, rvs)\right) \\
& \times s_{pt} s_{ru} \langle \Psi_0 | \left[\overline{E}_{tq}, \left[\overline{E}_{us}, \hat{H}\right]\right] |\Psi_0\rangle
\end{aligned}
$$

We then need to make repeated use of the commutation relations for the bi-orthogonal singlet excitation operators, as

$$
\begin{aligned}
s_{ru} \left[\overline{E}_{us}, \hat{H}\right] =& s_{ru} \overline{h}_{vw} \left[\overline{E}_{us}, \overline{E}_{vw}\right] + \frac{1}{2} s_{ru} \overline{g}_{vwxy} \left[\overline{E}_{us}, \overline{\Sigma}_{vwxy}\right] \\
=& s_{ru} \left(\overline{h}_{sw} \overline{E}_{uw} + \overline{g}_{swxy} \overline{\Sigma}_{uwxy} - \overline{h}_{vu} \overline{E}_{sv} - \overline{g}_{vuxy} \overline{\Sigma}_{vsxy}\right)
\end{aligned}
$$

Substituting this into the original expression then gives, after much simplification,

$$
\begin{aligned}
\frac{\partial^2 E}{\partial \kappa_{pq}^a \partial \kappa_{rs}^a} =& \left[\left(1 - \hat{P}(p, q)\right) \left(1 - \hat{P}(r, s)\right) \left(1 - \hat{P}(p, q)\right)\right]^2 \\
& \times s_{ru} s_{pt} \left[2\overline{h}_{tu} \overline{D}_{qs} - s_{tu} \left(\overline{F}_{qs} + \overline{F}_{sq}\right) + \overline{\mathcal{G}}_{tq,us}^-\right]
\end{aligned}
$$
(7.24)

$$
\overline{\mathcal{G}}_{pq,rs}^{\pm} = \mp 2\overline{g}_{prtu} \overline{d}_{qstu} \mp 2\overline{g}_{trpu} \overline{d}_{tsqu} + 2\overline{g}_{purt} \overline{d}_{qust}
$$
(7.25)

The only new quantities required are thus $\overline{\mathcal{G}}^{\pm}$, which will also be used in the symmetric Hessian elements. These necessarily scale as at least fourth-order in the number of orbitals, as they are contractions between two-electron integrals and the two-body density. We can absorb the metric elements into these, transforming all but one index back to the ALMO basis, then screen over the three ALMO indices in the same manner as Chapter 4, effectively reducing the scaling to quadratic. However, the effectiveness of this will depend on the sparsity of the two-body density.

To calculate the mixed derivative, we directly use the BCH-type expansion of the rotations in commutators and anticommutators, for the antisymmetric and symmetric cases, respectively. Transferring the transformation directly to the Hamiltonian, this is most simply written as

$$
\hat{H}(\boldsymbol{\kappa}^a, \boldsymbol{\kappa}^s) = \hat{H}_0 - \left\{\hat{\kappa}^s, \hat{H}_0\right\} + \left[\hat{\kappa}^a, \hat{H}_0\right] + \left\{\hat{\kappa}^s, \left[\hat{H}_0, \hat{\kappa}^a\right]\right\} + \mathcal{O}(\kappa^2)
$$

However, the energy denominator from equation 7.19 contains a first-order term in $\hat{\kappa}^s$ that needs to be combined with the first-order term in $\hat{\kappa}^a$ in the transformed Hamiltonian. The mixed second-order term in the energy is thus in total

$$
E^{(11)} = \langle \Psi_0 | \left\{\hat{\kappa}^s, \left[\hat{H}_0, \hat{\kappa}^a\right]\right\} |\Psi_0\rangle + 2 \langle \Psi_0 | \left[\hat{\kappa}^a, \hat{H}_0\right] |\Psi_0\rangle \langle \Psi_0 | \hat{\kappa}^s | \Psi_0\rangle
$$

Using the same considerations as per the orbital gradient with respect to the (anti)symmetries of the rotation parameters, the derivatives can then be written in terms of the symmetric and antisymmetric singlet excitation operators. Moreover, the anticommutator evaluates the same forwards and backwards, due to the matrix elements being real, simply introducing a factor of two. The Hessian elements are therefore

$$
\frac{\partial^2 E}{\partial \kappa_{pq}^a \partial \kappa_{rs}^s} = 2 \langle \Psi_0 | E_{rs}^+ \left[\hat{H}, E_{pq}^- \right] | \Psi_0 \rangle + 2 \langle \Psi_0 | \left[E_{pq}^-, \hat{H} \right] | \Psi_0 \rangle \langle \Psi_0 | \hat{E}_{rs}^+ | \Psi_0 \rangle
$$

$$
= 2 s_{pt} s_{ru} \left[\langle \Psi_0 | \overline{E}_{us}^+ \overline{E}_{tq}^- | \boldsymbol{\sigma} \rangle - \overline{\mathcal{F}}_{tq,us} \right] + 2 s_{ru} \frac{\partial E}{\partial \kappa_{pq}^a} \left(\overline{D}_{us} + \overline{D}_{su} \right)
$$

$$
\text{(7.26)}
$$

where we have defined the intermediate quantity

$$
\overline{\mathcal{F}}_{pqrs} = \langle \Psi_0 | \overline{E}_{rs}^- \hat{H} \overline{E}_{pq} | \Psi_0 \rangle
$$

$$
= \overline{h}_{tu} \langle \Psi_0 | \overline{E}_{rs}^- \overline{E}_{tu} \overline{E}_{pq}^+ | \Psi_0 \rangle + \frac{1}{2} \overline{g}_{tuvw} \langle \Psi_0 | \overline{E}_{rs}^- \overline{\Sigma}_{tuvw} \overline{E}_{pq}^+ | \Psi_0 \rangle
$$

$$
\text{(7.27)}
$$

In equation 7.26, we again need to evaluate a transition density through the sigma vector, this time up to fourth-order excitations. Similarly, equation 7.27 contains terms in the three- and four-body density matrices. While the transition densities, as already noted, can be heavily screened based on localisation, I can devise no such scheme for the pure densities. One possibility would be to try to rewrite all of these in terms of the single- and double-excitation transition densities, but it is not immediately obvious how to do so. Certainly, as it stands, this step will be prohibitively restrictive on the size of active space that can be considered. To make the method feasible, future efforts need to focus on reducing this cost.

Finally, the symmetric-symmetric Hessian terms can be read from the expansion that we performed when calculating the gradient. These are, directly:

$$
\frac{\partial^2 E}{\partial \kappa_{pq}^s \partial \kappa_{rs}^s} = 2 \langle \Psi_0 | E_{pq}^+ E_{rs}^+ \left(\hat{H} + 2 E_0 \hat{I} \right) | \Psi_0 \rangle + 2 \langle \Psi_0 | \left\{ E_{pq}^+, \hat{H} \right\} | \Psi_0 \rangle \langle \Psi_0 | E_{rs}^+ | \Psi_0 \rangle
$$

$$
+ 4 E_0 \langle \Psi_0 | E_{pq}^+ | \Psi_0 \rangle \langle \Psi_0 | E_{rs}^+ | \Psi_0 \rangle
$$

This can then be rewritten in the same manner as the other terms, to give

$$
\frac{\partial^2 E}{\partial \kappa_{pq}^s \partial \kappa_{rs}^s} = 2 s_{pt} s_{ru} \left[\langle \Psi_0 | \overline{E}_{tq}^+ \overline{E}_{us}^+ | \boldsymbol{\sigma} \rangle + E_0 \langle \Psi_0 | \overline{E}_{tq}^+ \overline{E}_{us}^+ | \Psi_0 \rangle \right]
$$

$$
+ 4 s_{pt} s_{ru} \langle \Psi_0 | \overline{E}_{tq}^+ | \boldsymbol{\sigma} \rangle \left[\overline{D}_{us} + \overline{D}_{su} \right]
$$

$$
+ 4 E_0 s_{pt} s_{ru} \left[\overline{D}_{tq} + \overline{D}_{qt} \right] \left[\overline{D}_{us} + \overline{D}_{su} \right]
$$

$$
\text{(7.28)}
$$

Note that equation 7.28 is written entirely in terms of the one- and two-body densities, and the first- and second-order transition densities. As a result, it is considerably simpler than equation 7.26 to evaluate, by virtue of the transition densities being easier to calculate than the higher-body densities. This is particularly important, as the symmetric parameters are in general dense, as shown in Figure 7.4. The symmetric-antisymmetric Hessian terms are much reduced in comparison, due to the sparsity of the antisymmetric parameters. Thus, while the three- and four-body densities of equation 7.27 will be a considerable bottleneck, the terms that need to be evaluated are at least limited to only *on-block* terms in the two antisymmetric indices.

7.2.3 MC-ALMO SCF

We now have all the components required for a multiconfigurational ALMO SCF procedure. In this section, we present in summary the high-level overview of the algorithm. As already noted, we have set up the problem in such a manner as to be able to borrow a great deal of the machinery of the standard ALMO SCF procedure; in particular, the density fitting and Fock build algorithms. The computational details of building transition and higher-body densities require thorough consideration in future, however, in order to make this algorithm feasible for anything but small active spaces. These appear necessarily due to the non-orthogonal nature of ALMOs, but we cannot sacrifice this localisation without losing the benefits - benefits which we will exploit to great effect in the embedding to be described in the next section.

The optimisation procedure follows a second-order expansion of the energy about some reference guess, $\boldsymbol{\lambda}_0$, in terms of the combined parameter set

$$\boldsymbol{\lambda} = \{\boldsymbol{\kappa}^a, \boldsymbol{\kappa}^s, \boldsymbol{\delta}\}$$

given as

$$E(\boldsymbol{\lambda}_0 + \Delta\boldsymbol{\lambda}) \approx E_0 + \Delta\boldsymbol{\lambda} \cdot \frac{\partial E}{\partial \boldsymbol{\lambda}} + \frac{1}{2}\Delta\boldsymbol{\lambda} \cdot \frac{\partial^2 E}{\partial \boldsymbol{\lambda}^2} \cdot \Delta\boldsymbol{\lambda}$$

$\boldsymbol{\kappa}^a$ *and* $\boldsymbol{\kappa}^s$ *are by construction uncoupled from one another, but in general these can be strongly coupled to the CI deltas, so this is not an assumption made lightly.*

We wish to variationally optimise the energy with respect to the parameters. Assuming that the parameters are uncoupled, this is equivalent to finding $\Delta\boldsymbol{\lambda}$ such that the gradients with respect to each parameter are zero. We note from equation 7.19 that

$$\frac{\partial E}{\partial \delta_I} = 2 \langle \Psi_0 | \hat{H} | \mathbf{I} \rangle - 2E_0 \langle \Psi_0 | \mathbf{I} \rangle \tag{7.29}$$

Setting this to zero is equivalent to solving the CI eigenproblem, and is approached as internal microiterations, as described in section 7.1.

The orbital rotation parameters are then determined as a Newton step in the usual manner. We solve for $E - E_0 = 0$, yielding

$$\Delta\boldsymbol{\lambda} = -2\mathbf{H}^{-1}\mathbf{g}$$

where $\mathbf{H}$ and $\mathbf{g}$ are the orbital Hessian and gradient, respectively. The problem with this is two-fold [362]: we need the step to remain in a region where the quadratic expansion is valid, i.e. fairly close to the reference point; and we need the Hessian to be positive definite. These problems can be solved by using a line-search and level-shift, respectively, for which we introduce two new parameters, α and μ. The step, in terms of only non-redundant parameters, is then

$$\Delta\boldsymbol{\lambda} = -\alpha\left[\mathbf{H} + \mu\mathbf{I}\right]^{-1}\mathbf{g} \tag{7.30}$$

The parameter μ is determined such that the resulting Hessian is positive definite, which can be done by finding or approximating the lowest eigenvalue, μ_0, of $\mathbf{H}$; shifting by this will guarantee a semi-positive definite, while shifting by $\mu = (1 + \delta)\mu_0$ with $\delta > 0$, guarantees positive definiteness.

The line-search parameter is found in the usual way by minimising the energy with respect to it. Given that $\mathbf{z} = \left[\mathbf{H} + \mu\mathbf{I}\right]^{-1}\mathbf{g}$ is a known quantity, the quadratic energy expansion in terms of α is

$$E(\alpha) \approx E_0 - \alpha\mathbf{z} \cdot \mathbf{g} + \frac{1}{2}\alpha^2 \mathbf{z} \cdot \mathbf{H} \cdot \mathbf{z}$$

such that the optimum α is given by

$$\alpha^* = \mathbf{z} \cdot \mathbf{g}\left[\mathbf{z} \cdot \mathbf{H} \cdot \mathbf{z}\right]^{-1} \tag{7.31}$$

which can be calculated at very little extra cost. This is then finally adjusted with a hard cutoff by maximum step size; that is, we choose α closest to α^* such that $\|\Delta\boldsymbol{\lambda}\| = \alpha\|\mathbf{z}\|$ is less than or equal to some maximum allowed step size. This maximum is then updated according to the ratio of the expected energy to the actual energy from the last step - if we underestimate the energy change, the maximum step is reduced, while if we overestimate, it is increased accordingly.

Finally, this is all summarised in Algorithm 7.1 below. We note that Kreplin and coworkers [386] very recently gave improvements to the standard second-order algorithm of Werner and Knowles [387], which ours is based on; it is likely that these improvements can also be applied here in future. Note that the convergence of the macroiterations is highly dependent on the microiterations converging to a sufficiently tight threshold [370]. Given sensible thresholds, there are therefore relatively few macroiterations, such that the construction of the higher-order densities is minimised - they will still represent the bottleneck, however.

Algorithm 7.1 The multiconfigurational (MC) ALMO SCF routine.

1: Construct all valid strings for the active space
2: **Macroiterations**
3: **while** Energy and gradient norm are not below threshold **do**
4:     Construct the one- and two-body densities
5:     Transform the integrals from the AO to the ALMO basis
6:     Use the transformed integrals to build the core energy (equation 7.11) and the effective Fock matrices with at least one active orbital index (equations 7.12 and 7.23), half-transformed into the bi-orthogonal basis
7:     **for** Each alpha and beta string **do**
8:         Prune the generating graph according to the ALMO distance criterion
9:         Find all strings connecting through up to four excitations, store addresses and phases
10:    **end for**
11:    **CI microiterations** (Davidson-Liu algorithm, ref. [381, 382])
12:    Choose a set of L guess vectors $\{c_I\}$, where L is the number of desired roots
13:    **while** CI gradient is higher than threshold **do**
14:        Calculate the sigma vector in the active space for this guess, via equation 7.6
15:        Solve the reduced eigenproblem for the matrix with elements $c_I \cdot \sigma_J$, for eigenvalues and vectors μ^k and v^k
16:        Calculate the shifts, δ_I^k for each root k, according to

$$\delta_I^k = \left(\mu^k - H_{II}\right)^{-1} \left[\sum_{J=1}^{L} v_J^k \cdot \left(\sigma_J - \mu^k c_J\right)\right]_I$$

17:        Orthogonalise the shifts against the guess, appending them to the guess if their norm is greater than some threshold
18:    **end while**
19:    **Orbital microiterations** (Werner-Knowles algorithm, ref. [387])
20:    **while** Orbital gradient norm is higher than threshold **do**
21:        Use the sigma vector to generate first- and second-order transition densities
22:        Build the three- and four-body densities, and from this the intermediates $\mathcal{G}^{\pm}$ (equation 7.25) and $\mathcal{F}$ (equation 7.27)
23:        Build the non-redundant orbital gradient (equations 7.20 and 7.21) and Hessian (equations 7.24, 7.26, and 7.28)
24:        Diagonalise the Hessian or estimate the lowest eigenvalue, to determine the shift μ
25:        Calculate α^* via equation 7.31
26:        Apply the Newton step of equation 7.30
27:    **end while**
28: **end while**

7.3 EMBEDDED ALMO THEORY

The design of the methods in this thesis has up until this point focused primarily on interaction energies between well-separated molecular fragments. In principle, however, there is no need to make such a distinction - the absolutely localised 'fragments' can be defined to be any collection of atomic centres. The lack of charge transfer means that defining fragments across covalent bonds will result in a large error, which is the main reason we have focused on non-covalent interactions. However, there is no reason why we cannot go in the other direction and choose a fragment of *more than one* molecule.

For example, non-covalent interactions in systems with strong static correlation are of interest, particularly in the case of exciplexes found in chromophoric systems [388, 389], where electrons are excited between molecules. However, reactive systems typically involve the transfer of atoms between molecules, where a consistent description essentially requires that the reactive molecules are considered together. Another way of thinking about this is that the energies of these molecules are strongly coupled, as opposed to the generally weaker coupling between, for example, a solvent molecule and a substrate.

If we define these 'embedded' subsystems as fragments, however, the definition of an interaction or binding energy becomes less clear. We are then almost arbitrarily decomposing the interaction energies as being between supermolecular systems. This problem has actually appeared in a different form earlier, when determining the analytic gradient within the pairwise RPA approximation. The decomposition into pairs of subsystems so as to determine interaction energies directly results in each monomer energy gaining a multiplicity, which we then need to average over. As was demonstrated in Chapter 6, this ad hoc procedure appears to be quite effective, but it alludes to a greater underlying problem.

It would be more natural to determine the total energy and then calculate a given property by calculating differences. The ALMO SCF procedure does in fact give a total energy, not an interaction energy directly; we already do monomer-basis fragment calculations in order to determine the interaction. This generates a consistent ALMO density for the entire system, which then allows us to perform our pairwise correlated calculations. By extension, we could instead perform correlated calculations using this density on *any* subsystem defined as a collection of one or more fragments. If we are concerned with total energies, the correlation energy for these subsystems does not need to be decomposed as in the RPAxd approach developed in Chapter 3, thus removing an approximation. It also opens up the possibility of treating different subsystems with different correlated methods, depending on their importance to the property of interest. Specifically

in the context of MC-ALMO-SCF, this could be where the relevant active space is in a reactive subsystem surrounded by spectators and solvent molecules. These latter molecules can individually be treated with a lower-level correlation if desired, or left at the mean-field level by putting all their orbitals in the inactive and secondary spaces. Similarly, if we wanted to calculate spectra of a molecule in solution, we can treat the solvent with monomer RPA correlation, and the molecule of interest with as high a level method as desired: CCSD(T) for example, or for excitation energies, equation-of-motion CCSD; we will use this approach in the next chapter.

The primary caveat, however, is that the ALMO energy is missing all 'charge transfer' terms, and this is a representation of the ALMO density (or wavefunction) not being the optimal Hartree-Fock wavefunction. Adding the perturbative charge transfer correction does not fix this problem. We have shown in Chapter 5 that the ALMO+CT interaction energy is fairly consistently equivalent to the counterpoise-corrected HF interaction energy, thus, as noted at the end of Chapter 6, the ALMO total energy is the HF energy minus the basis set superposition error. While the latter is an unphysical *energy*, it is a physically relevant contribution to the density, due to the more complete, unconstrained basis used in the HF procedure.

We do know, however, from the consideration of the perturbative charge transfer, that we can essentially recover the HF density by diagonalising the full Fock matrix rather than the fragment Fock matrices separately. This is not exact, as we have not converged this full density, but is a good approximation. Thus, if we wish for more accurate total energies for a given subsystem or set of subsystems, we can take the minor of the total ALMO Fock matrix corresponding to that subsystem, diagonalise it, and from this generate a pseudo-HF density for that subsystem. The advantages are then two-fold: the density is improved, containing at least the charge transfer between the fragments in the subsystem; the orbitals for that subsystem are now canonical, and *any* canonical correlation method can be used. We therefore do not need to develop non-orthogonal correlated higher-order methods. Analytical gradients then follow naturally by using the ALMO+CT gradients defined in Chapter 6, then adding in the canonical gradients for whichever correlated method was used on each subsystem. Assuming that none of the subsystems overlap in definition, there is no need to average gradient terms, thus avoiding any possibility of discontinuities.

We have demonstrated proof-of-concept of this approach in Figure 7.5, by calculating the harmonic vibrational spectra of phenol, chosen because it is the subject of the next chapter. It is not possible to do this using the pair-based method, as phenol is only a single molecule. With the embedded approach, we can instead surround
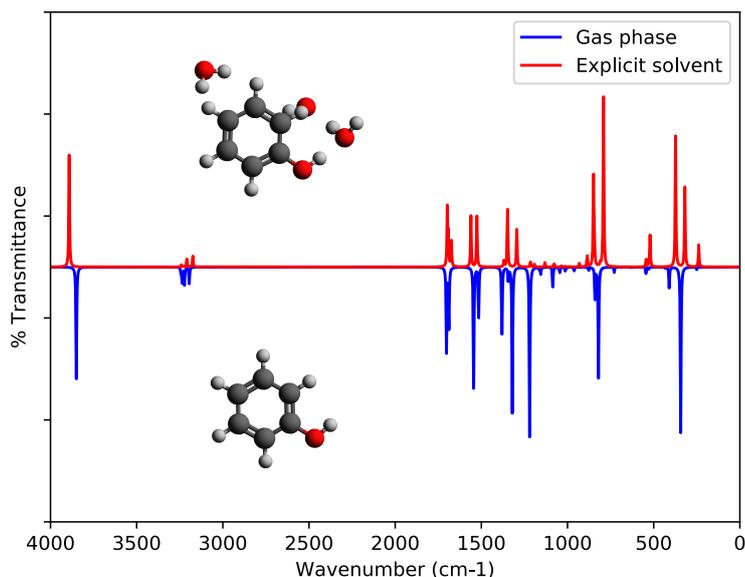
Figure 7.5: Infrared spectrum of phenol in the gas phase (bottom, blue lines) and solvated in a shell of 19 water molecules (top, red lines). The optimised structure in each instance is insetted, with the three closest waters shown in the latter. Calculations were performed at the ALMO+CT//CCSD(T)/aVDZ level; this notation reads as bath//embedded subsystem (phenol here)/basis set.

the phenol molecule with water solvent, perform an ALMO+CT calculation, then canonicalise just the phenol density, before calculating the CCSD(T) harmonic frequencies. In the figure, we show the results of the canonical gas-phase CCSD(T) calculation, compared with that from ALMO+CT/[CCSD(T)] with a 6 Å radius spherical shell of 19 water molecules. The geometries are slightly different, as both were relaxed at their given level, as is necessary when calculating frequencies from the electronic Hessian. Both calculations were performed in GAMMA with the aVDZ basis and matching auxiliary sets, on a single node with 8 threads. For reference, the gas-phase calculation took 3.5 hours, compared to 4.2 hours for the explicitly solvated calculation; the difference was almost entirely due to the latter requiring two more optimisation steps than the former. Interestingly, one can clearly see that the solvent shifts the OH peak at around 3300 cm^{-1}. I could not find an experimental water-solvated IR spectrum of phenol to confirm this, but it makes intuitive sense: the hydroxyl bond is weakened by hydrogen bonds with water, lowering the force constant of the vibrational mode.

The method of canonicalising the subsytem will not work, however, for multiconfigurational (and therefore usually also multi-reference) methods, wherein the density is determined as part of the proce-

dure. If we take the canonicalised subsystem density as described above, and use this to seed a standard MCSCF procedure, the total density will change, and will no longer be embedded correctly. In other words, there will be an incongruous, discontinuous description of the electron density between layers in the embedding. Instead, we perform the MC-ALMO-SCF described in this chapter on the whole system, but constrain the active space to be in the subsystem of interest. As noted earlier, we can then still relax the ALMO approximation for that subsystem by allowing mixing between the active orbitals in the antisymmetric orbital rotation parameters. Further dynamic correlation can then also be added in the usual manner (e.g. CASPT2 [390, 391]) for just that subsystem, by virtue of the definition of the active space.

The main concern with this approach, however, is sheer cost. In the exposition of the theory, I have tried to make as much use of the inherent localisation as possible, but as was already noted, the higher-order density matrices are a serious bottleneck. Even extending only over the inactive-active and active-active spaces, this can be very large when the inactive space contains many fragments, even with all the usual speed-ups when building the effective Fock matrices. Future work in this area therefore needs to focus not on improving the description of dynamic correlation, but rather on finding new ways to either evaluate or approximate the costly density matrices. One advantage of this is that the transition densities are very useful [392, 393], and used to determine properties of excited states [394], which would also extend the applicability of the embedded ALMO method.

8 | THE PHOTOELECTRON SPECTRUM OF PHENOL

In this chapter, I apply the culmination of the methods developed so far in this thesis to the task of simulating the photoelectron spectrum of phenol in water. Solvent has been shown experimentally to have a significant effect on the excited states of molecules, and this poses a difficult problem for quantum chemistry to address accurately. I briefly outline the theory behind excited state calculations, and implement them as a direct continuation of the embedding theory outlined at the end of the previous chapter. Using molecular dynamics simulations, we can sample the conformational distribution of phenol in water. These samples are then used to calculate the ultra-violet absorption and photoelectron spectra. Through these calculations, I demonstrate the utility of the new methods via comparison with experiment, and in particular the need for the inclusion of a large shell of explicit solvent to give accurate results.

Most of the work so far in this thesis has been concerned with the ground state electronic structure and properties of molecules. The ground state is often the focus of such work as typically, at room temperature, the electronic ground state is the predominant form of a molecule. When determining, for example, reaction dynamics, we thus want a detailed description of that state. However, as was touched on in the previous chapter, many interesting systems are multi-reference in character, whereby multiple states lying close to one another collectively determine the behaviour of a molecular system.

On a more fundamental level, even energetically well-separated excited states are of interest. Experiments ranging from Förster resonance energy transfer (used to investigate the dynamics of large biomolecules [395, 396]) to ultrafast laser spectroscopy (used in the study of, for example, light-harvesting proteins [397]) use the excitations between electronic states in chromophores. Using computation to guide and interpret such experiments therefore requires methods to treat these excited electronic states.

The separation between electronic energy levels in chromophores is typically on the order of 1 to 100 eV, placing it firmly in the UV-visible portion of the electromagnetic spectrum. The most common experiments to probe electronic structure are thus UV-visible absorption

and emission spectroscopy, and photoelectron spectroscopy (PES). In the former, light excites an electron, usually from the ground to some excited state. By scanning over wavelengths and monitoring either the intensity of light absorbed or emitted, we can see the energy separations between electronic states; within this, there is also finer, rovibronic structure, but we ignore this here for simplicity. PES, in contrast, determines the binding, or ionisation, energies of electrons. The kinetic energy of emitted electrons is measured in comparison to the energy of incoming light, with the difference being the energy required to detach that electron.

A recent study by Riley and coworkers considered the photoelectron spectra of phenol in water using liquid-microjet PES [398]. As they state, 'electronically excited states can be very sensitive to their microenvironment, and the extent to which dynamical insights obtained from gas-phase studies can be used to inform our understanding of the dynamics in chemically biologically relevant environments is a subject of considerable discussion'. Polar solvents in particular have been shown to greatly influence these dynamics [399]. Water, unarguably the most prevalent solvent, is highly polar, so that understanding its effect on electronically excited states is vital. They chose phenol as a small molecule representative of the aromatic moieties commonly found in biologically-relevant chromophores. In the present chapter, I will reproduce the experimental shifts from the gas phase to solvated UV absorption and photoelectron spectra computationally, demonstrating the utility of the embedded ALMO approach.

8.1 THEORY OF EXCITED STATES

The principle behind determining excited states is identical to that of the ground state: we wish to solve the Hamiltonian eigenvalue problem, but for the k-th eigenvalue instead of the zeroth. This eigenvalue, E_k, has a wavefunction Ψ_k associated with it, satisfying the usual Schrödinger equation:

$$\hat{H}\Psi_k = E_k\Psi_k \tag{8.1}$$

We have already seen one typical treatment for this problem, the configuration interaction method, in the previous chapter, directly in the context of absolutely localised molecular orbitals. Truncated CI has considerable limitations, however, as it scales factorially with the truncation space, and is not size consistent [34, 210].

In the ground-state problem, we typically circumvent this by using an exponential ansätz, leading to the coupled-cluster (CC) type approaches described in Chapter 3. As already noted, though, excited states are usually open-shell systems dominated by more than a sin-

gle determinant, such that single-reference CC will not suffice. Multireference CC is possible [400], but computationally very demanding. Instead, we propose to use equation-of-motion (EOM) CC [401], embedded in an unrestricted ALMO-SCF density, as described at the end of the last chapter. In this section, I will briefly outline the method, before discussing how it can be related to the ring and GW approximations previously described for the ground state case. The primary advantage of this is that we can start from the already-implemented single reference rCCD wavefunction and obtain the excited states from there.

8.1.1 Equation–of–motion coupled cluster

By considering the change from the reference state, Ψ_0, we can rewrite the excited state problem directly in terms of the excitation energies

$$\omega_k = E_k - E_0$$

However, we have not solved for the reference state, but rather for the reference determinant, $|\mathbf{n}\rangle$, with energy $E_{\text{ref.}}$. As per the CI equations, the reference state will be expanded in terms of excited determinants, $|\mathbf{I}\rangle$. We can project out the reference determinant from the eigenproblem by writing a normal-ordered Hamiltonian, $\hat{H}_N$, as

$$\hat{H}_N = \hat{F} + \hat{W} - \langle \mathbf{n} | \hat{H} | \mathbf{n} \rangle = \hat{H} - E_{\text{ref.}} \tag{8.2}$$

where $\hat{F}$ is the usual Fock operator, and $\hat{W}$ is the fluctuation potential described in Chapter 3.3. The action of this on a state is then

$$\hat{H}_N \Psi_k = (E_k - E_{\text{ref.}}) \Psi_k = \Delta E_k \Psi_k$$

Note that $\Delta E_k - \Delta E_0 = \omega_k$, which is directly the excitation energy.

The initial CC wavefunction is expanded as usual as

$$|\Psi_0\rangle = \exp(\hat{T}) |\mathbf{n}\rangle$$

where the cluster operator $\hat{T}$ is typically truncated; if not truncated, this is equivalent to the full CI expansion. From this reference state, we can generate the excited states through an exictation operator, $\hat{R}_k$, which in analogy to CI is a string of singlet excitation operators:

$$\hat{R}_k = r_0 + r_i^a E_{ai} + \frac{1}{2} r_{ij}^{ab} E_{ai} E_{bj} + \cdots \tag{8.3}$$

where the r values are the amplitudes, effectively equivalent to the CI coefficients. Note that this necessarily commutes with the cluster operator, which follows the same expansion with different amplitudes, but whereas the latter is exponential, $\hat{R}_k$ is linear.

Combining equations 8.2 and 8.3, and left-multiplying by $\exp(-\hat{T})$, the eigenvalue problem to be solved can be written as

$$\overline{H}\hat{R}_k \ket{\mathbf{n}} = \exp(-\hat{T})\hat{H}_N \exp(\hat{T})\hat{R}_k \ket{\mathbf{n}} = \Delta E_k \hat{R}_k \ket{\mathbf{n}} \qquad (8.4)$$

where we have defined a similarity-transformed Hamiltonian, $\overline{H}$. Note that equation 8.4 is now in the same form as the CI eigenvalue equation, except $\overline{H}$ is non-Hermitian. This means that there are both right eigenfunctions, $\hat{R}_k \ket{\mathbf{n}}$, and left eigenfunctions, $\hat{L}_k \ket{\mathbf{n}}$, with the same (real) eigenvalues. The operator $\hat{L}_k$ is a de-excitation operator:

$$\hat{L}_k = l_0 + l_a^i E_{ia} + \frac{1}{2}l_{ab}^{ij} E_{ia}E_{jb} + \cdots \qquad (8.5)$$

The left and right eigenfunctions are necessarily bi-orthogonal, and thus provide a resolution of the identity. This will allow us to remove terms from the excited state common to the reference state, simplifying the problem.

For the reference state, $\hat{R}_0$ is easily seen to be the identity, such that equation 8.4 simply becomes

$$\overline{H}\ket{\mathbf{n}} = \Delta E_0 \ket{\mathbf{n}}$$

Subtracting this from the same equation for state k yields

$$\overline{H}\left(\hat{R}_k - \hat{I}\right)\ket{\mathbf{n}} = \left(\Delta E_k \hat{R}_k - \Delta E_0\right)\ket{\mathbf{n}}$$

Inserting the excitation operator to the left of the reference problem, this can be rewritten to give the core EOM-CC equation:

$$\left[\overline{H}, \hat{R}_k\right]\ket{\mathbf{n}} = \omega_k \hat{R}_k \ket{\mathbf{n}} \qquad (8.6)$$

The computational solution of these equations is somewhat complicated, due to the non-Hermiticity of the similarity-transformed Hamiltonian, but is described in detail elsewhere [222]. The starting point is constructing a density from the ground-state amplitude equations solved as described in Chapters 3 and 4, then iteratively solving the linear-response (or lambda) equations for the left and right eigenfunctions [402]. As with CI, this is typically only done for a select few of the lowest excited states using a Davidson-type procedure.

8.1.2 Diagrammatic EOM–CC, ring diagrams, and the GW approximation

As with the ground-state CC theory, we can gain a great deal of insight into the EOM equations, and in particular their link to other theories, by writing them diagrammatically. The approach is exactly the same as that given in Figure 3.2, but instead of describing the

Singles

Doubles

Figure 8.1: The diagrammatic amplitude equations for the right eigenfunctions in EOM-CCSD. The interaction lines are in terms of the similarity-transformed Hamiltonian. Ring terms are highlighted with a grey background.

cluster operator, $\hat{T}$, the amplitudes are describing the excitation operator, $\hat{R}_k$. We distinguish the latter by giving the amplitude diagrams a thick vacuum line. The dashed interaction lines are still taken to be the fluctuation potential $\hat{W}$, which as per the RPA theory can be taken to be the bare Coulomb potential, or the antisymmetrised variant. We then generate all amplitude diagrams up to second order in exactly the same manner as Figure 3.2, and these are shown in Figure 8.1. As before, these can be generated up to arbitrary orders and the algebraic equations read off directly from the diagrams. For example, the single excitation equations in the EOM-CCSD truncation are

$$\chi_{ad} r_i^d - \chi_{li} r_l^a + \chi_{ladi} r_l^d + \chi_{ld} r_{il}^{ad} + \frac{1}{2}\chi_{alde} r_{il}^{de} - \frac{1}{2}\chi_{lmid} r_{lm}^{ad} = \omega r_i^a$$

where χ_{pq} and χ_{pqrs} are the one- and two-electron matrix elements of the similarity-transformed Hamiltonian, respectively.

We note that for the right eigenfunctions, only the *connected* diagrams need to be included, as per canonical CC [222]. In EOM methods, however, we have the added complication of needing to solve for the left eigenfunctions as well. These de-excitation amplitudes, given in Figure 8.2, have no such connectivity condition, and are somewhat more problematic. The simultaneous solution of these two problems is the primary source of complexity in the EOM-CC method. Our embedding approach greatly reduces the cost of performing the calculation in the mean field of some macro-environment, but does nothing to reduce the cost of the core EOM calculation. It is therefore useful to try and consider approximations that simplify the calculation.

Singles

Doubles

Figure 8.2: The diagrams for the de-excitation (or Λ) amplitudes, required to solve for the left eigenfunctions in the EOM-CCSD problem. Interactions are in terms of the similarity-transformed Hamiltonian, with ring terms highlighted in grey.

In Figures 8.1 and 8.2, I have highlighted the terms that generate ring diagrams. Note, however, that in the Λ-equations, there are a pair of disconnected ring diagrams that must be included. The right-side amplitude equations in the EOM-rCCD approximation described by only including the highlighted diagrams, corresponds to a very similar Riccatti-type equation to that of equation 3.53. The left-side amplitude equations are much more similar to a full CCSD calculation, and will in general result in the largest expense. Both sets of amplitude equations can readily be generated from the diagrams directly, or equivalently by excluding the neglected terms from the standard EOM-CCSD equations, which can be found elsewhere [222]. We do not present these here, as there are no special considerations given we will be canonicalising the ALMO density on the embedded fragment.

Finally, we note that the EOM-rCCD approximation described is essentially an extension of the solution of the RPA excitation problem, as described in a paper by Berkelbach [403]. This approach is thus the natural extension from our RPA-based ground state theory, and most closely related to the RPAx variant for the correlation energy. In a similar manner to that discussed in Chapter 3.4, this is in turn related to the GW approximation [279, 404]. It is worth being somewhat more explicit here, as this could lead to potential expansion of the approach to higher-order approximations, and to applications in the solid state, where GW-type methods are routinely used.

As noted in the detailed analysis by Lange and Berkelbach [404], the primary difference in the EOM-CCSD and GW Green's functions is that the former is exact through second-order in the exchange diagrams, while the latter is not. However, at higher orders, the GW

Green's function actually includes more terms than the truncated coupled cluster Green's function; to include these exactly through higher orders, we would need to truncate the CC excitation operator at the relevant excitation level. Thus, through iterative expansion of the vertex function, as described in Chapter 3.4, we can once again include approximate triples (and higher) corrections to either the EOM-CCSD or EOM-rCCD equations by collecting these terms from the GW approximation. These effectively perturbative corrections are considerably cheaper to calculate than the full triples and higher CC amplitudes. More details can be found in refs. [403, 404].

8.1.3 Dyson orbitals

For the calculation of photoelectron spectra, we need the N-electron neutral ground state and the $(N-1)$-electron cationic ground and excited states. These can then be used to compute the ionisation potentials that correspond to the peaks in the spectra [405]. The typical representation of the orbitals involved in these excitations are thus the overlaps between the neutral and cationic states, known as Dyson orbitals [406]. These are written as

$$\phi_D^k(\mathbf{r}) = \sqrt{N} \int d\mathbf{r}_2 \ldots d\mathbf{r}_N \; \Psi_0^N(\mathbf{r}_1, \ldots, \mathbf{r}_N) \Psi_k^{N-1}(\mathbf{r}_2, \ldots, \mathbf{r}_N) \qquad (8.7)$$

and are thus a function of the ionised electron.

The orbitals of equation 8.7 can then be used to give the photoelectron matrix element - or transition dipole moment - between the initial and final states as

$$\mathbf{D}_k = \langle \phi_D^k | \hat{\mu} | \Psi_k \rangle_r$$

where $\hat{\mu}$ is the electric dipole operator, and the subscript r denotes that the expectation is spherically averaged. Calculation of these moments requires the densities of the initial and final states, and the dipole moment matrix elements between orbitals. Details of how these are then used to calculate the photoionisation cross sections can be found elsewhere [405].

8.2 MOLECULAR DYNAMICS SIMULATIONS

Molecules in solution cannot typically be assumed to exist in a single, energetically-optimal configuration. Instead, to get meaningful results, we need to sample several different conformations, as dictated by the interaction of the small molecule with the surrounding solvent, and the general thermal variations in the molecule itself. To
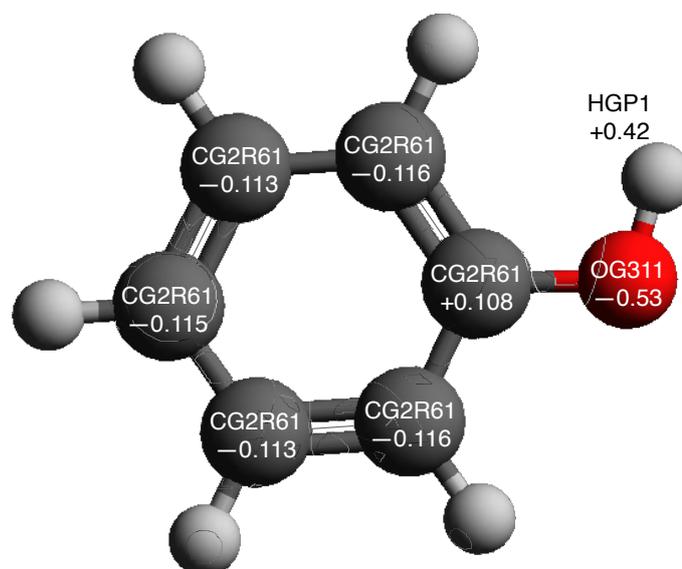
Figure 8.3: The CGenFF atom types and partial charges (in reduced units) for phenol used in the MD simulations described herein. All unlabelled hydrogens are of the aromatic hydrogen (HGR61) type, with charge +0.115 reduced units, as prescribed by the CGenFF standard procedure [331].

do this, we use classical molecular dynamics simulations to generate an ensemble of configurations. While these will not give an accurate representation of the true dynamics of the molecule, in the limit of infinite such samples (assuming ergodicity), we would expect to tend towards the correct distribution for any given ensemble property [407]. In this case, the spectroscopy we are interested in is not performed on single molecules, but rather on low-concentration samples in solution, such that the spectra are indeed such an ensemble property.

There are several different forcefields that we could use, ranging in accuracy for different applications. We choose the CHARMM general forcefield (CGenFF [331]), because its atomic charges are determined so as to reproduce quantum-mechanical interaction curves with water. The most important element of the simulation for our conformational generation is the interactions with the solvent, such that this seems a sensible choice. As with any point-charge-based forcefield, the molecules are parametrised in terms of harmonic bond-stretching, bond-angle, and torsional terms (the latter typically expanded as a truncated Fourier series), coupled with non-bonded Lennard-Jones and Coulomb terms. For phenol, which is a standard organic moiety, we can accurately assign all these parameters by analogy with

existing parameters in CGenFF. We have thus used the ParamChem server [408, 409] to generate these parameters, noting that the heuristic measure of suitability in all cases was approximately zero, indicating a low likelihood of inaccuracy. The atom types and charges are shown in Figure 8.3.

Simulations were performed in GROMACS version 2016.4 [330] with the CHARMM 36 (June 2019) forcefield [331], including the relevant TIP4P four-point water model parameters [332]. The latter was selected for its balance between accuracy and efficiency when compared to the TIP3P [332] and TIP5P [410] models, respectively. The temperature was held to 298.15 K using a Nosé-Hoover thermostat on the whole system, with a damping parameter of 2.0 reduced units. Pressure was fixed at 1 bar using a Parrinello-Rahman barostat with time constant 2.0 ps and compressibility 4.46×10^{-5} bar^{-1}. A standard time-step of 2 fs is used throughout, along with a short-range van der Waals cutoff of 1.0 nm, and particle-mesh Ewald for long-range electrostatic interactions, also with a cutoff of 1.0 nm. We performed five independent 25 ns simulations of phenol in a cubic box of water, with side-length 2.3 nm. In each simulation, we energy-minimised the phenol, then equilibrated the temperature and pressure, in that order, with short simulations of 500 ps.

Figure 8.4 shows two validation plots for a representative simulation - the same plots for the other simulations are essentially indistinguishable. In the first of these, the root-mean-square deviations of the atomic positions in the phenol molecule are given as a function of simulation time. In a truly ergodic simulation, we would expect this to show no correlation, with mean equal to the average conformational flexibility of the molecule. That is exactly what we see in the plot, with mean RMSD of around 3 Å, typical for small aromatic molecules [411]. The velocity autocorrelation is particularly important, as it dictates the time intervals on which we can take independent samples. The autocorrelation function tells us the probability that a property at time t_0 can be used to predict the same property at time $t_0 + t$. Truly independent observations should show no correlation. We see that the velocities of the phenol atoms become essentially uncorrelated within very short timescales. This is to be expected for a small molecule free to diffuse in water, where there is relatively little rigidity in the structure, when compared to, for example, large proteins.

The plots in figure 8.4 suggest that we can sample independent snapshots of the solvated phenol from any point in the simulations (there is no additional equilibration period in the RMSD plot) with essentially any time separation. We are fundamentally limited, however, by the frequency at which the positions and velocities were outputted, and in the time it would take to run the subsequent quantum-
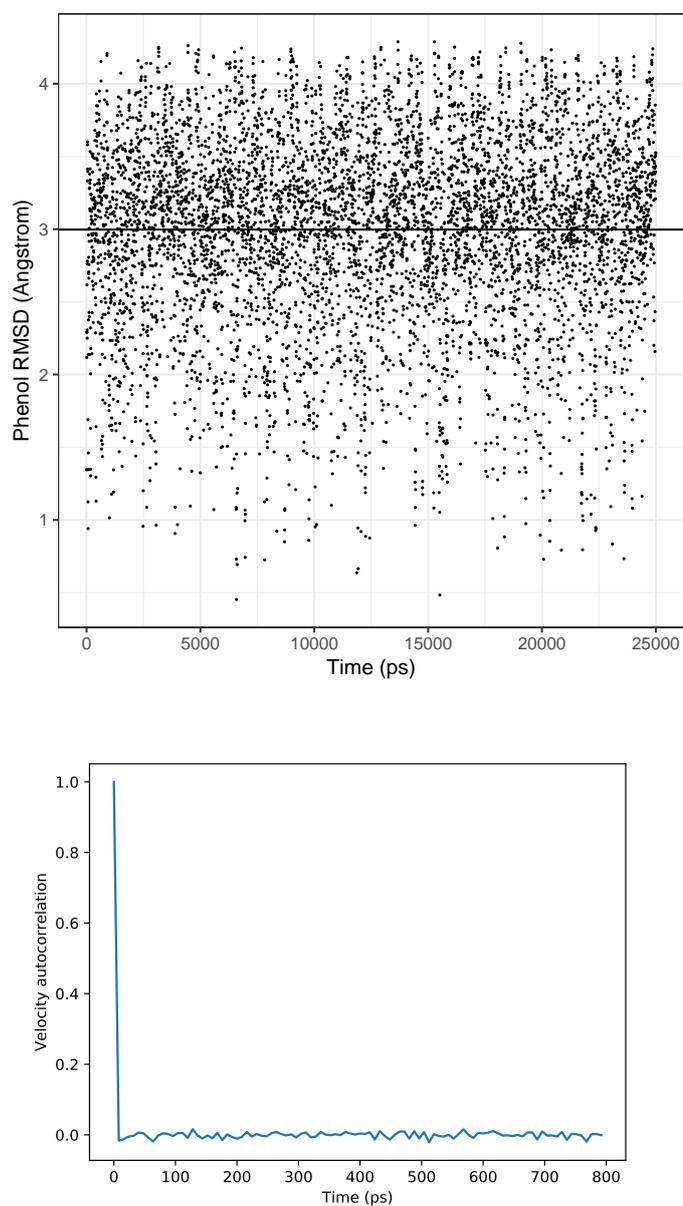
Figure 8.4: Root-mean-square deviation (top) and velocity autocorrelation (bottom) plots for phenol over the course of a single 25 ns simulation. Both show that the simulations are satisfactorily equilibrated, and that independent samples can readily be drawn from the simulations.
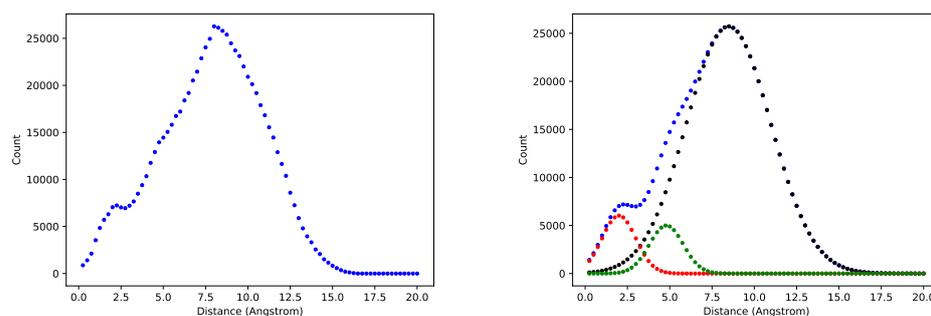
Figure 8.5: The radial distribution function (RDF) between phenol and the water molecules averaged over a single 25 ns simulation. On the left is the raw RDF, and on the right is the RDF fitted as a sum of three Gaussians, showing the peaks in the distribution.

mechanical calculations. Therefore, we choose to take samples at 1 ns intervals, starting at 0.5 ns, from each simulation. This gives us an ensemble of 105 configurations in total.

Figure 8.5 shows the radial distribution function between the phenol and the waters, averaged across an entire simulation. One of the biggest questions when seeing how explicit solvent affects results is how many solvent molecules need to be included to capture all of the effects. This will largely depend on the length scales of both the property to be calculated, and the interactions between the solute and the solvent. As can be seen in the figure, the waters roughly divide into three shells at around 2, 5, and 9 Å. Such peaks imply order, and thus interactions, on fairly long distances - the water molecules beyond the 9 Å mark are essentially randomly distributed in relation to the phenol. In Figure 8.6, we can see explicitly how these shells rapidly expand to include well over a hundred waters.

Investigating this further across all 105 samples, we see in Figure 8.7 that the number of water molecules follows a sigmoid centred around 12 Å. Using this, we can label the shells as the intervals between the sets of points given: the first shell at 2 to 5 Å; second shell between 5 and 7.5 Å; and the third shell from 7.5 to 11 Å. The biggest jump is between the second and third shells. Electronic energy levels are strongly coupled to the conformation of a molecule, and to the interactions between molecules, but the excitations between such levels are not generally thought to be particularly sensitive to long range interactions [412].

In the next section, I will demonstrate how there is a substantial difference between results only using up to the first shell and up to the second shell, but fortunately not such a difference between the second and third. This is perhaps not surprising in consideration of Figure 8.8, which shows the number of hydrogen bonds between phenol and water molecules across the course of a simulation. This averages

	3.3 Å	4.3 Å
Cutoff		
No. of waters	4	13

	6.3 Å	9.3 Å
Cutoff		
No. of waters	48	133
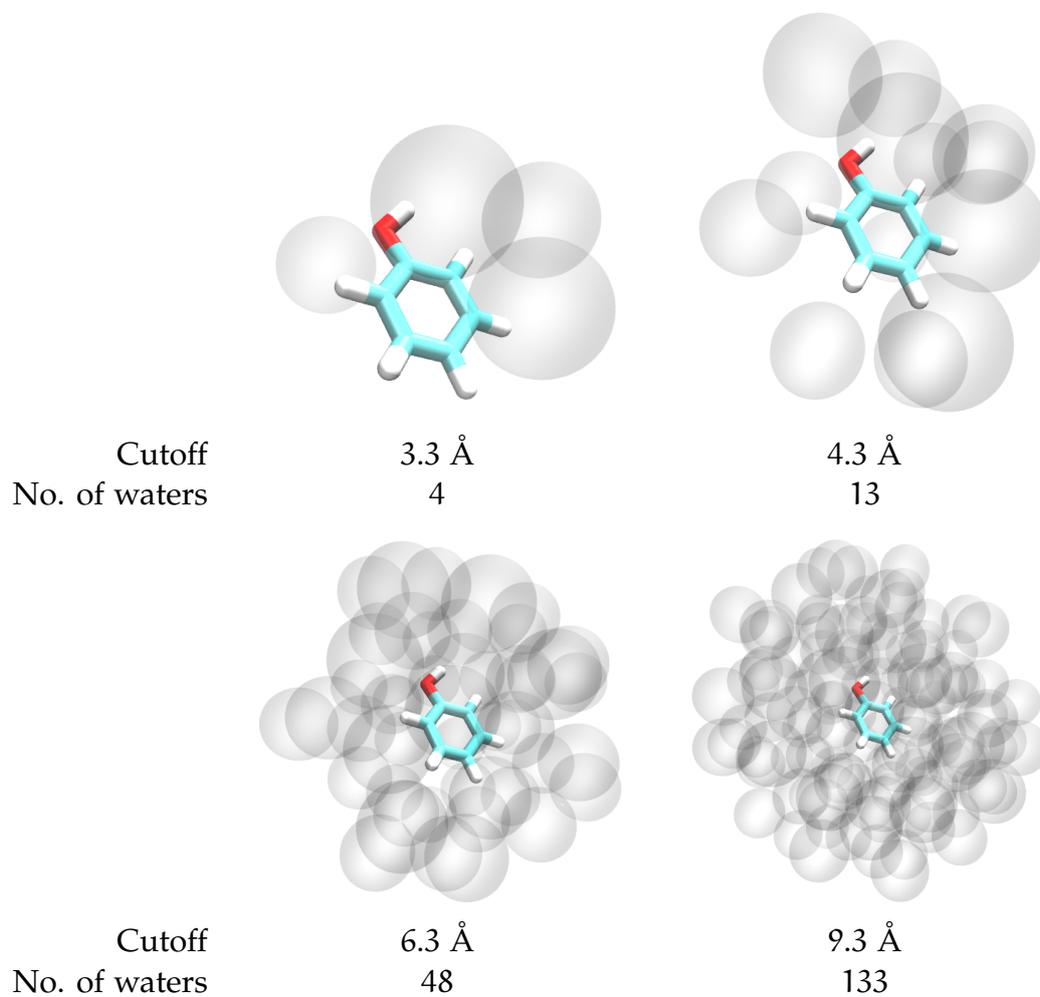
Figure 8.6: Visualisations of the density of water molecules (represented as spheres) surrounding a particular conformation of phenol taken from one MD snapshot, for a variety of cutoff distances from the centre of mass of the phenol. These range from 4 waters in the smallest sphere, to 133 in the largest.
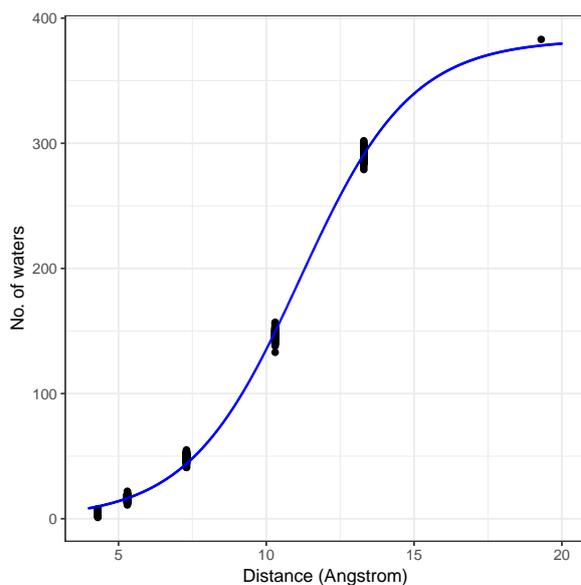
Figure 8.7: The number of water molecules contained in a spherical shell of a given radius with centre at the centre of mass of the phenol molecule, shown for all 105 samples.
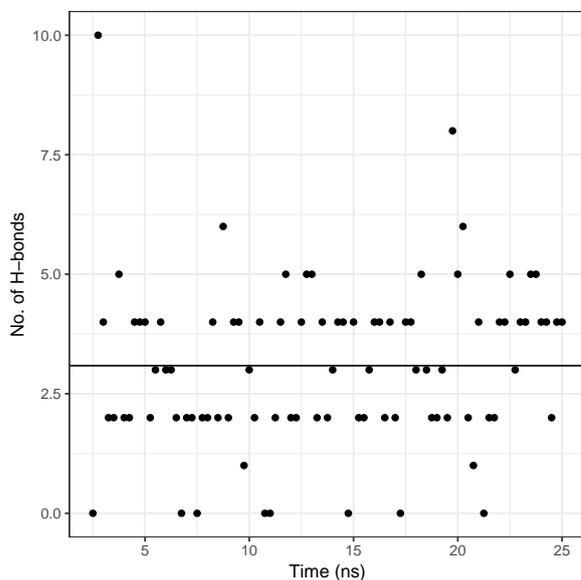
Figure 8.8: The number of hydrogen bonds between the phenol and water molecules over the course of a single 25 ns simulation. Hydrogen bonds were defined to be interactions with a cutoff of 3 Å and bond angles of less than 60 degrees.

to around three such hydrogen bonds, which suggests the strongest effect will always be with the first, and by proxy, second shells. This equates to needing 30 – 50 water molecules in each quantum calculation, as opposed to around 130. Note that even the former would not be remotely possible with traditional excited-state quantum methods. Even with the new linear-scaling method, the difference in computational expense between 50 and 130 solvent molecules is substantial, so it is very useful to know that the spectra converge before this point.

8.3 SIMULATED SPECTRA

All excited state calculations performed in this section were done using GAMMA at either the EOM-CCSD or EOM-rCCD level of theory, with aVDZ basis sets on all atoms [343]. Density fitting was used throughout in both the Fock build and integral transformation steps, with the matching aVDZ auxiliary basis sets [151, 162, 325, 327]. Embedded calculations were calculated using the local ALMO approach with the infinite-order charge-transfer correction; ground-state correlation energies were calculated using the RPAx variant, summed over all excitations. For the ground-state UV absorption spectra, the gas-phase results were validated using the MOLPRO suite of programs [323], and transition dipole moments taken from EOM-CCSD calculations for the lowest five excited states. The peak intensities in the solvated, embedded calculations were estimated from transition dipole moments of the phenol density, averaged over all 105 conformational samples.

For the photoelectron spectra, the vertical ionisation energies were calculated by performing spin-unrestricted ALMO+RPAx calculations on the neutral, singlet ground state, and the cationic, doublet ground state. The summed difference in the HF monomer and RPAx correlation energy was then taken to be the first ionisation energy. The higher-order ionisations were estimated from EOM-rCCD calculations on the cationic state. All spectra were averaged over the entire ensemble of conformations, with the cationic state not allowed to relax.

This is equivalent to assuming we are in the Franck-Condon regime.

To calculate the Dyson orbitals, and thus PES cross-sections, we passed the EOM-rCCD density to the NEWTONX program [413]. The Dyson orbitals were calculated directly from equation 8.7, not approximated using Koopman's theorem. NEWTONX then estimates transition moments by performing CI singles calculations using the reference density provided; the coefficient and overlap cutoffs were set at 0.01 and 0.1, respectively. Finally, the ionisation cross-sections were determined using the Dyson orbital norms, with the electron kinetic energy, eKE, allowed to vary, and the photon energy fixed at two photons of wavelength: 235.5 nm, 253 nm, 265.5 nm, or 275 nm. These
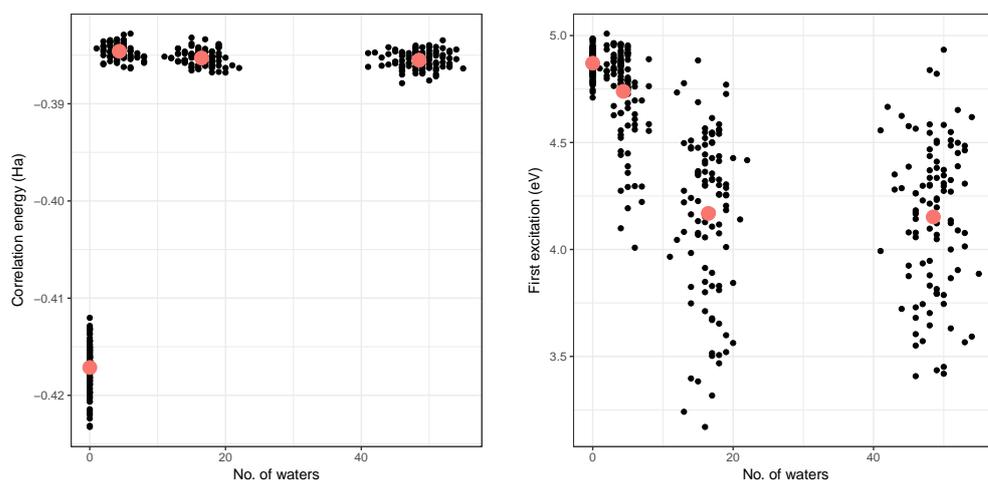
Figure 8.9: The RPAx/aVDZ correlation energy (left), and the first EOM-rCCD/aVDZ vertical excitation energy (right), for phenol as a function of the number of waters, for all 105 samples. The correlation energy in the gas phase calculations is slightly larger due to the full canonical density being used, as opposed to the canonicalised ALMO density in the solvated simulations. Red dots show the mean value at each separation cutoff.

were selected to match with 1+1 UV PES in the reference experiments. The binding energy, eBE, is then $eBE = 2h\nu - eKE$, where ν is the frequency of light. Gas phase calculations were performed on the optimised geometry, while the solvated results were averaged over the ensemble.

8.3.1 Convergence with number of waters

Calculating the PE spectra is expensive, whereas individual correlated calculations are cheap. As such, we wish to determine the number of solvent molecules that need to be included for convergence in the properties of interest to occur. In the current calculations those properties are: the total energy of phenol, in particular the correlation energy contribution; the excitation energies of the neutral and cationic states; and the densities of the neutral and cationic states. The last of these is difficult to quantify, but directly influences the correlation and excitation energies. Therefore, to a good approximation, we can consider the density converged if these energy metrics are converged.

Figure 8.9 shows the full ensemble results for the RPAx correlation energy of the phenol, and the EOM-rCCD first excitation energy. In the former, we see that the correlation energy is largely insensitive to the number of water molecules. This is perhaps unsurprising, as it is dominated by the intramolecular terms involving excitations only within the molecule. The most noticeable difference is between the

Table 8.1: Mean values (across the ensemble) in nm for the first four vertical excitations of phenol in varying amounts of water, as calculated at the EOM-rCCD/aVDZ//ALMO+CT level.

Gas phase	1st shell	2nd shell	3rd shell
257.8	260.5	279.9	280.1
212.9	216.9	225.4	227.0
200.3	204.1	214.8	216.2
196.5	199.3	207.1	209.3

gas-phase fully canonical density calculation and the solvated, canonicalised ALMO density. There is a decrease of about 20 mHa in the correlation energy, which is surprisingly substantial. Redressing this difference is something that needs to be looked at in future work.

More interesting, however, is the marked change in the first excitation energy upon inclusion of the solvent. The difference between the gas-phase and first solvation shell results is not huge, although the excitation energy does decrease by on average 0.2 eV. Upon including the second shell of waters, however, there is a decrease of around 0.6 eV, which is substantial. Moreover, the overall delta of around 0.8 eV between the gas and solvated phases is in almost exact agreement with the experimental results (Figure 1 in ref. [398]). Finally, we note that including the next shell does not significantly affect the excitation energy. Similar results were found for the first four excitation energies, shown in Table 8.1, although there is a slightly greater difference of around 0.05 eV on average upon inclusion of the third shell of waters. We do not expect this to make a quantitative difference to the spectra, however, as the broadening from the ensemble average will be at least an order of magnitude larger than this difference, as can be seen from the spread of points in Figure 8.9. As such, we will simulate all solvated spectra herein with the entire second shell of water molecules included explicitly.

A similar shift is then seen when calculating the first ionisation potential as the difference in the phenol total energy of the neutral singlet ground state and the cationic doublet ground state. Figure 8.10 shows the distribution over the ensemble for both the gas-phase and solvated calculations. Overall, the latter is more broadly distributed and shifted to around 0.8 eV lower energy, once more agreeing with experimental observation [398]. We note, however, that the mean ionisation energies of both are approximately 0.2 eV lower than the experimental values, which suggests that either the basis set or truncated level of theory may not be sufficient to give quantitative agreement. As can be seen in the lower panel of Figure 8.10, increasing the basis set size to a triple-zeta level successfully eliminates this shift. However, the associated increase in compute time is severe, such that we
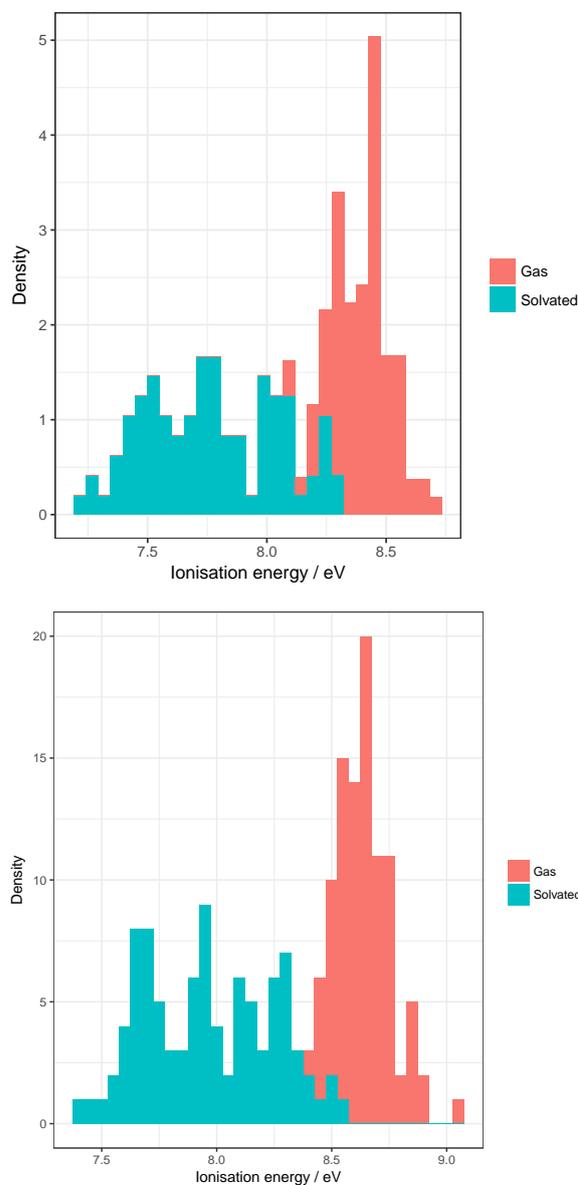
Figure 8.10: The first ionisation potential calculated at the EOM-rCCD/aVDZ level (top) and aVTZ level (bottom) for phenol, as the difference between the neutral and cationic total energies. The distribution over all 105 samples is shown, with the solvated results (blue, embedded in an ALMO+CT density) showing a significant shift to lower energies compared to the gas phase (red) results.
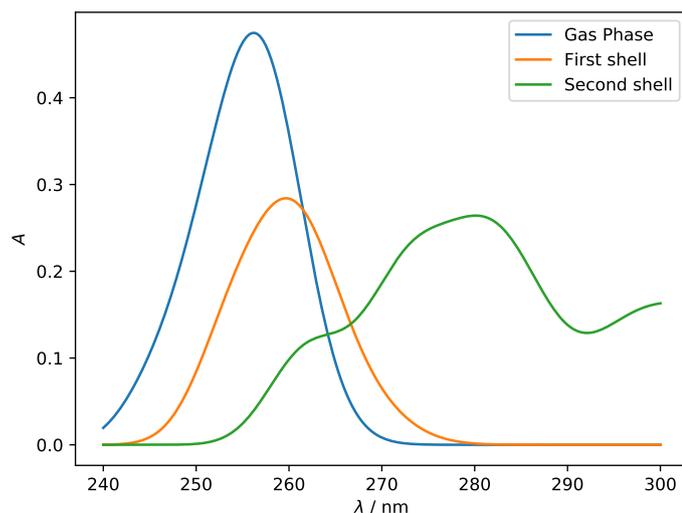
Figure 8.11: The lowest energy band of the UV absorption spectrum of phenol calculated at the EOM-CCSD/aVDZ level in the gas phase, and with the first and second solvation shells. The solvated spectrum is averaged over the entire ensemble of conformational samples. Again, there is a significant shift to lower energies upon inclusion of solvent, in good agreement with the experimental result.

proceed in the following with the smaller basis set. We also note that, while the ionisation potentials shift, the vertical excitation energies *do not* change upon using the larger basis; the average shift was found to be on the order of 0.5 mE_h.

8.3.2 UV spectra

An easier but equally interesting problem to simulating the PE spectrum - which depends on the excited states of the cation - is simulating the UV absorption spectrum, dependent on the excited states of the neutral ground state. As in this case we can use a spin-restricted wavefunction, the calculations are much more stable and less computationally expensive, allowing us to use the full EOM-CCSD calculation on the entire ensemble up to the third solvation shell. We calculated the first excitation energy and transition dipole moment, estimating the latter in the solvated case from the canonicalised ALMO density, then ensemble-averaged these to give the simulated UV spectrum. The resulting spectra are shown in Figure 8.11.

As per Figures 8.9 and 8.10, we see a significant shift to lower energies upon explicitly including water. There is a concomitant broadening and lowering of the peaks, corresponding to the greater conformational variation in solution. The peak centres at around 260 and

Gas phase   With second solvation shell
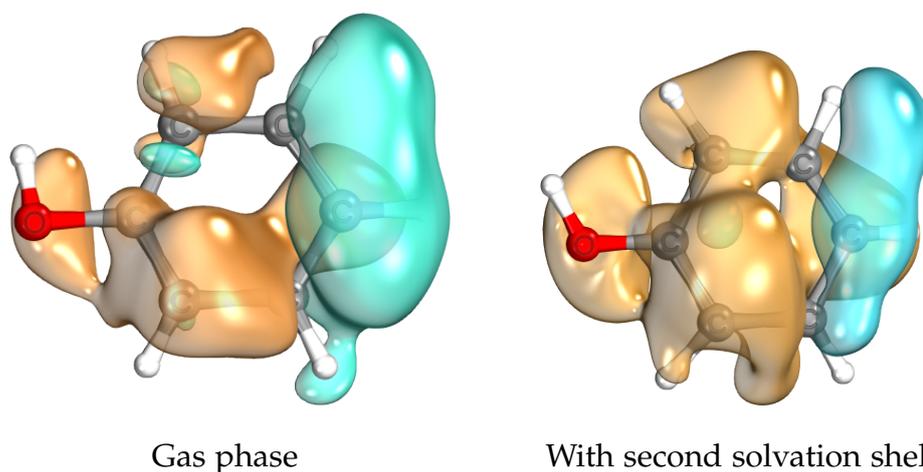
Figure 8.12: The Dyson orbital for the lowest energy transition in the photo-electron spectrum of phenol, calculated at the EOM-rCCD level in both the gas phase (left) and with the second solvation shell (right), for one particular conformational sample. The solvated result shows increased density over the ring and around the hydroxyl group, corresponding to the energy lowering due to interactions with surrounding water molecules. Plotted using IBOView [218].

280 nm for the gas-phase and 'second shell' peaks agrees fairly well with the experimental results in Figure 1 of ref. [398]. This is in contrast to the systematic shift noted for the ionisation potential above, perhaps suggesting that the lack of ladder diagrams in EOM-rCCD is a significant source of error.

8.3.3 Dyson orbitals and photoelectron spectra

Phenol is an organic, aromatic compound, with a slightly polar hydroxyl group. Therefore, it is not expected to be strongly hydrophilic, nor strongly hydrophobic. However, there are clearly significant electronic and structural changes caused by the interaction with the surrounding water molecules. The structural difference is largely due to a lengthening of the O−H bond due to hydrogen bonding with water molecules - averaged over the entire ensemble, this bond is 0.015 Å longer in the solvated structures compared to the gas phase. The electronic changes are then largely due to the subsequent movement in the π-electron density as the oxygen lone pair becomes less available. It is interesting to inspect the energy-equivalent Dyson orbitals with and without solvent, to directly see these electronic effects. These are shown for one sample conformation in Figure 8.12.

From the figure, we can see that the excitation does indeed involve considerably more density over the ring and the hydroxyl group in the solvated case, suggesting more availability of the electron density

265.5 nm   275 nm
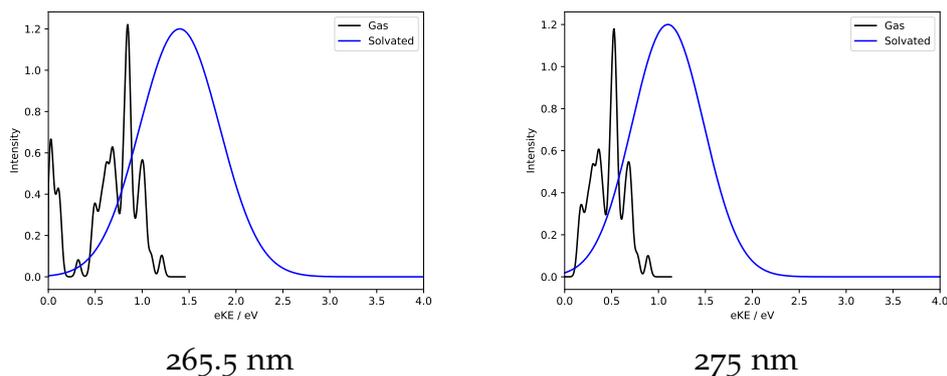
Figure 8.13: The gas-phase (black lines) and solvated (blue lines) photoelectron spectra in terms of electron kinetic energy for the two lowest energy transitions in phenol. The loss of fine structure in the solvated spectrum is due to ensemble averaging, resulting in a Gaussian-like distribution. There is once more a significant energy shift, in good agreement with experiment.

in these regions for excitation. The gas-phase Dyson orbital, in contrast, is much more localised to the carbons adjacent to the oxygen atom. The combination of this and the UV results of Figure 8.11 thus suggest that we will likely see significant changes in the photoelectron spectra when calculated in solvent.

As can be seen in Figures 8.13 and 8.14, there is indeed a substantial difference. There are two modes of plotting the spectra - in terms of electron kinetic or binding energy - which are essentially equivalent but can be read slightly differently. Figure 8.13 shows the excess electron kinetic energy when ionised with two photons of wavelength 265.5 or 275 nm. In the gas-phase case, calculated on the single optimised structure, there is considerable fine structure, not reaching above roughly 1 eV. The solvated results are broadened into a Gaussian, due to ensemble averaging, centred around 0.7 to 0.8 eV higher. This corresponds to the ionisation potential being *lower*, as the electrons have more residual kinetic energy. Comparing with the left-hand panels of Figure 2 of the experimental reference [398], we see that the qualitative agreement is excellent, but with a systematic downshift of around 0.2 eV, probably again due to the use of the ring approximation.

The other view, given in Figure 8.14 for the higher energy photons, is in terms of the amount of energy required to detach the electron from phenol. In these we once again see a loss in fine structure, and an energetic downward shift of around 0.8 eV. Particularly striking is the anisotropy in the 235.5 nm spectrum, which matches that found in experiment [398]. We do once more see a systematic difference of around 0.2 eV, however.

235.5 nm   253 nm
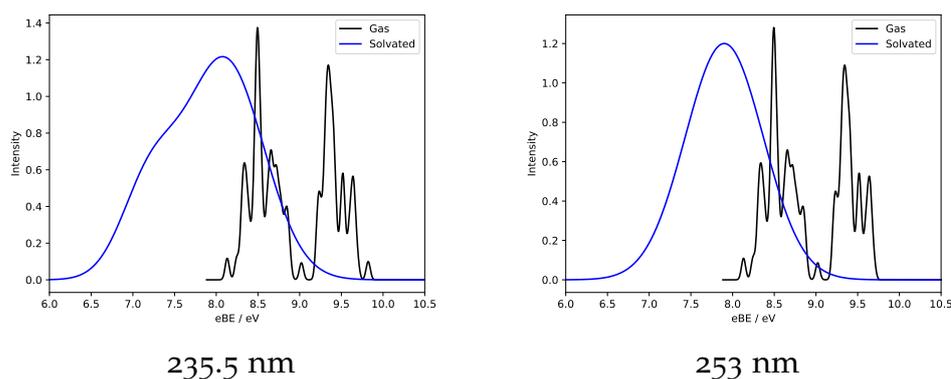
Figure 8.14: The gas-phase (black lines) and solvated (blue lines) photoelectron spectra in terms of electron binding energy for the third- and fourth-lowest energy transitions in phenol. We see a particular anisotropy in the 235 nm solvated spectrum (agreeing with experiment, ref. [398]), although the exact energy transitions are in error compared to experiment by about 0.3 eV.

In summary, we have successfully used our embedded ALMO approach on an ensemble of conformers to calculate electronic excited state spectra. Firstly, we have found that the use of a molecular–dynamics-generated ensemble reproduces the peak broadening seen in experiment. Secondly, the addition of explicit solvent in the equation-of-motion calculations yields significant shifts in the vertical excitation energies in both the neutral and cationic states, again in agreement with experiment; quantitative agreement of the energies, however, seems to be quite sensitive to the level of theory used. Finally, we have seen that the effects of the solvent are fairly long range, only converging between the second and third solvation shells.

Overall, this work highlights the importance of including explicit solvent in the simulation of these spectra, and the determination of electronically excited states. It should be noted that, while several high-level quantum calculations needed to be performed to determine these over an ensemble of 105 conformations, the time spent on these calculations was considerably less than that from calculating the Dyson orbitals and generating the photoelectron spectra. This is largely due to the use of an external package, involving a large amount of writing to and reading from disk. To make these calculations routinely manageable, it would be desirable to reimplement the Dyson orbital calculations directly in GAMMA. The other point of interest is that, a restricted EOM-CCSD calculation on phenol takes roughly 5 times as long as the unrestricted EOM-rCCD equivalent; the unrestricted full EOM-CCSD calculation will thus be even more time consuming. As we have noticed a systematic error that appears to be due to the ring approximation, future work should focus on ap-

proximating this using GW-type approaches, in an effort to improve the accuracy.

9 | CONCLUSIONS AND FUTURE WORK

In this thesis, I have described new methods for the computational study of systems comprising a large number of small fragments - for example, clusters of hundreds of water molecules. The developments can be broadly divided into three areas: theory, implementation, and application. In the following, I briefly summarise the findings presented in this thesis that are pertinent to each of these, and give a perspective on where future work should focus.

Theory

In Chapters 2 and 3, we have presented a novel, correlated method for the treatment of non-covalent interactions, based on absolutely localised molecular orbitals. This method directly gives interaction energies devoid of basis set superposition errors. Furthermore, it automatically yields an energy decomposition analysis in terms of electrostatics, polarisation, charge transfer, and dispersion. The last term is provided by a new, local random phase approximation-based approach, where the double excitations are categorised based on their physical interpretation. Benchmarks, in Chapter 5, have shown the ALMO+RPAxd results to be in good agreement with those obtained using CCSD, across a wide range of systems. In Chapter 6, the theory was extended by the addition of analytical gradients, derived using a Lagrangian approach, allowing for rapid geometry optimisations to be carried out. Then, in Chapter 7, we described how the method can be expanded to include systems with multiconfigurational character. This involved the development of a local, non-orthogonal configuration interaction theory, alongside a second-order orbital optimisation directly including the absolutely localised orbital constraint. Finally, all of the above culminated in an embedding method for the total energy of any given system, based on a definition of the ALMO (or MC-ALMO) SCF fragments in terms of larger groups of molecules.

Given the promising initial results, there are numerous opportunities for future theoretical developments. Most pressing is the contribution of higher-order excitations to energy differences (and thus particularly to interaction energies). It is well-known that the perturbative triples correction used in CCSD(T) is a significant part of the total interaction energy of a system [267, 333], and specifically makes a large difference in dispersion-bound systems. In Chapter 3, we

briefly discussed how such a contribution could be included through exploitation of the connection of CC-type methods with the GW approximation; in particular, by systematic expansion of the vertex function. Attempts could also be based on similar work in local CC theories [414, 415] or higher-order RPA [275]. Further theoretical developments could focus on the extension of the embedding approach to a wider range of theories. As noted at the end of Chapter 7, by canonicalising the embdedded density, any SCF-based theory could readily be inserted into both the energy and gradient calculations. Finally, while we have mostly focused on the liquid condensed phase, the approach could instead be applied to the solid state. This would require careful consideration of periodicity, but would open up the possibility of accurate calculations on surfaces and crystals, using larger unit cells than has previously been possible.

Implementation

The primary intention with the ALMO-based approach was to yield a linear-scaling implementation, which could then be applied to systems containing many hundreds of small molecules. In Chapter 4, we described how this could be achieved for the correlated, BSSE-free interaction energies. In the first instance, density fitting was applied to reduce both the scaling and prefactor of the two-electron integral calculation and transformation steps. In order to reduce the memory imprint, a new fragment-blocked data structure and density fitting algorithm were introduced. This was then coupled with a fragment-based local Fock build procedure. Finally, we devised a pairwise approximation for the RPA dispersion calculations. This latter approximation is fairly severe, in that some fraction of the higher-order many-body terms are necessarily lost, but in Chapter 5 we demonstrated that a substantial amount of the three-body terms at least are retained. In Chapter 6, the same approximations were then able to be applied to the analytical gradients, directly giving a sub-quadratic scaling implementation of these gradients. Similarly, the heterogeneous parallelisation considered in Chapter 4 can be applied to both the energy and force calculations equally. This comprised a coarse, distributed-memory parallelism divided over fragments and fragment pairs, in tandem with shared-memory parallelism of the Fock-build and integration algorithms. GPUs, if available, can then be used to accelerate the most time-consuming linear algebra operations.

The main focus of future work in this area should be on tuning the parallelisation, and in particular expanding the GPU capabilities such that they can be used efficiently, without affecting the non-GPU implementation. Dependencies on external libraries make this a difficult problem to address, and other tensor-contraction libraries should

be considered. Of the algorithms described in Chapter 4, the only significant remaining non-linear term is the Cholesky decomposition of the density fitting metric. It is not clear how this could be linearised without exploiting sparsity, although the work of Schweizer *et al.* [416] could be used as a starting point. A more significant implementation challenge comes from Chapter 7, where only the theory for MC-ALMO SCF was presented. As noted in the chapter, many of the individual parts - especially the effective Fock matrices and integral transformations - can borrow from the single-determinant algorithms. However, the calculation of three- and four-body densities will need to be addressed. One approach would be to expand these approximately in terms of lower-order densities, or transition densities. This is a difficult problem that has received only limited attention in the literature [417]. Finally, the embedding proposed in Chapter 7, implemented in Chapter 8 for equation-of-motion coupled cluster, could be expanded to a wider variety of correlated methods, offering an alternative route to higher-order correlation energies than that described in the previous section.

Application

Any theory is only as good as the results it gives, and how useful those results are in real-world applications. The methods described in this thesis are specialised to large clusters, or extended systems, of small molecules. This was exploited in Chapter 5, where the accuracy of ALMO+RPAxd was demonstrated across a wide range of bimolecular systems, ranging from halogen and hydrogen bonds, to purely dispersive interactions. We also showed the utility on much larger clusters of water molecules (up to 202 monomers) and hydrogen fluoride chains. The ability to perform an energy decomposition on these allows for the investigation of how the physics of systems changes going from the gas to condensed phases. Furthermore, in Chapter 6, we used the analytical gradients to optimise the geometries of the same systems, and calculated the infrared spectrum of the water dimer. The effective densities calculated as part of the gradients could also be used to determine other molecular properties, such as the response to an electromagnetic field. Finally, in Chapter 8, we applied the embedded ALMO method to simulating the UV absorption and photoelectron spectra of phenol in water, reproducing the effects seen in experiment. In future, we hope that these demonstrations of the utility of the theory, and its implementation in GAMMA, will lead to it being applied to a much wider range of real, physical problems.

Appendices

A | PRIMER ON TENSOR NOTATION

This appendix is intended to be a brief introduction to tensors and in particular the notation used to represent and manipulate them, necessary for understanding certain developments in the main text. It is by no means comprehensive, and for a more detailed introduction, the reader is directed to any number of textbooks, for example refs. [418, 419].

A tensor is a mathematical object representing a multilinear transformation of some kind. More precisely, given a finite, n-dimensional linear space $\mathcal{V}$ that can be written as the direct product of n vector spaces, $\mathcal{V}_i$, a tensor, T, is a mapping from $\mathcal{V}$ into some other vector space, $\mathcal{W}$. The multilinearity implies it must satisfy the property that

$$T(v_1, \ldots, \alpha v_k + u, \ldots, v_n) = \alpha T(v_1, \ldots, v_k, \ldots, v_n) + T(v_1, \ldots, u, \ldots, v_n)$$

for all scalars α in the underlying ring of $\mathcal{V}$, for all $u \in \mathcal{V}$, and for each axis, v_i. This condition essentially means that tensors represent well-behaved transformations, such that tensor equations are valid in any coordinate system; this is why they are so useful, and why tensors are the preferred representation of many problems in physics.

It is conventional not to use functional or operator notation for tensors, but rather to choose a particular representation and use index notation. Given a choice of basis $\{e_i\}$ for $\mathcal{V}$, an element of the overall tensor, $\mathbf{T}$, in this representation is denoted $T_{ijk\ldots n}$; note that boldface is used to distinguish the tensor from its elements. The number of indices thus corresponds to the rank of the tensor. The most common examples of tensors are vectors (rank one) and matrices (rank two). In the index, or suffix, notation, familiar operations over these objects can be more succinctly written. For example, a matrix-vector multiplication can be written as

$$(Ab)_i = \sum_j A_{ij} b_j$$

This is an example of **reduction**, so-called as it reduces the rank of the matrix A from two to one. More generally, we can consider **contraction** over an index, where a pair of indices are iterated over (not necessarily in a multiplication) to contract two tensors together:

$$(AB)_{ij\ldots mn} = \sum_k A_{ij\ldots k} B_{k\ldots mn}$$

Clearly, explicitly writing out the summations would become notationally burdensome, especially when multiple indices are contracted over at once. As such, we introduce the *Einstein summation convention*, where it is assumed that - unless explictly stated otherwise - the appearance of two identical indices implies contraction over those indices. That is, for a general scalar operation •, we write

$$\sum_{kl} A_{ij...kl} \bullet B_{kl...mn} \equiv A_{ij...kl} \bullet B_{kl...mn}$$

An important point about this convention is that it only applies when the index in question appears precisely twice in the operation; if summation occurs over the same dummy index in more than two places, e.g. $\sum_j A_{ij} B_{jj}$, then this needs to be written explicitly.

While the index notation is useful, it does not bely anything about the transformation properties of the objects in question, and therefore does not convey the invariance of tensorial equations. To see how this can be achieved, consider a general affine transformation, **A**, of the basis to a new representation:

$$\tilde{\mathbf{e}}_i = A_{ij} \mathbf{e}_j \ \forall i$$

The elements A_{ij} thus represent the projection of $\tilde{\mathbf{e}}_i$ onto $\mathbf{e}_j$. A general point $\mathbf{v} \in \mathcal{V}$ can by definition be written as a linear combination of either set of basis vectors. In the initial representation, it is given by

$$\mathbf{v} = v_i \mathbf{e}_i$$

such that to determine the new representation, we need the projection of $\mathbf{e}_i$ onto $\tilde{\mathbf{e}}_j$, the inverse of the above:

$$\tilde{v}_i = v_j \left(A^{-1} \right)_{ji}$$

A quantity that transforms in this way - in the opposite sense to how the basis transforms - is termed **contravariant**. It is conventional to denote this property by superscripting the indices, e.g the above would become

$$\tilde{v}_i = v^j \left(A^{-1} \right)_{ji}$$

Instead, suppose we consider any mapping $f : \mathcal{V} \to \mathcal{W}$, and in particular its gradient, $\mathbf{w} = \nabla f$, at some point in $\mathcal{V}$. By the chain rule, in the new representation this becomes

$$\tilde{w}_i = \left(\frac{\partial v_j}{\partial \tilde{v}_i} \right) w_j$$

Given that $v_i = A_{ij} \tilde{v}_j$ and **A** does not depend on the v_i, the partial derivative is simply the transformation element, A_{ij}. That is, the gradient transforms in the same sense as the basis, and is termed **covariant**; such quantities are denoted by subscripting their indices. More

generally, different axes of a tensor may transform differently. For example, many rank-2 tensors (matrices) will be mixed, with their rows and columns transforming in opposite senses. Taking the transpose of such an object then leads to ambiguities unless placeholder indices are introduced - this is done by using dots:

$$A_i^j \equiv A_{i\cdot}^{\cdot j} \text{ and } \left(A^\mathsf{T}\right)_i^j \equiv A_{\cdot i}^{j\cdot}$$

The reduction of two tensors to form a scalar is of vital importance in physical applications. In particular, we require the inner product of two tensors to be invariant to changes in coordinate system - physical quantities, i.e. scalar values, should not depend on the particular choice of representation for the system. Clearly, this is not true for the usual inner product between two vectors, $s = u^i v^i$, as this transforms as

$$\tilde{s} = \tilde{a}_i \tilde{b}_i = a^j \left(A^{-1}\right)_{ji} b^k \left(A^{-1}\right)_{ki} = a^j \left[A^{-1}\left(A^{-1}\right)^\mathsf{T}\right]_{jk} b^k$$

which only equates to s in an orthogonal transformation (i.e. where $\mathbf{A}\mathbf{A}^\mathsf{T} = \mathbf{I}$, with $\mathbf{I}$ the identity). Similarly, the inner product between two covariant quantities will not be invariant when defined in this way. However, the product between a co- and contra-variant quantity will be: if $s = a^i b_i$, then

$$\tilde{s} = \tilde{a}^i \tilde{b}_i = a^j \left(A^{-1}A\right)_{jk} b_k = a^j b_j = s$$

We can solve the problem for a pair of contravariant quantities by taking the inner product with respect to a covariant weight. That is, we weight each a^i by w_i such that the new inner product is $s = a^i w_i w_j b^j$, transforming to

$$\tilde{s} = \tilde{a}^i \tilde{w}_i \tilde{w}_j \tilde{b}^j = a^k \left(A^{-1}A\right)_{kl} w_l w_m \left(AA^{-1}\right)_{mn} b^n = a^k w_k w_m b^m = s$$

From this, we see that only the products of the weights are needed, and so we collect these as $g_{ij} = w_i w_j$. In an orthonormal system, this would simply be the identity. Note that both indices are suffixed, implying that both axes transform covariantly.

The quantity g therefore serves as a means of converting between the covariant space and its dual. That is, we clearly have that

$$a_i = g_{ij} a^j \text{ and } a^i = a_j g^{ji}$$

In particular, this means that $a_i = g_{ij} g^{ji} a_i$, such that we must have that the contravariant form of g is the inverse of the covariant form. The co- and contra-variant dual pair thus represent a means of *raising* or *lowering* an index. Essentially, on the curved manifold described by

the nonorthogonal basis, g represents the distance between co- and contra-variant quantities, and thus encodes the information required to transform these quantities correctly. As such, it is called the **metric tensor**, or usually just the **metric**. The contravariant equivalent is then, by the above, the **inverse metric**. For a tensorial equation to be invariant to transformations, we therefore require that all contractions imply the presence of this metric; this is equivalent to requiring that contractions only occur over dual pairs of indices. Thus

$$s = A^{jk}B^j_k$$

would be invalid (that is, would not be invariant to the choice of coordinate system), as j is contravariant in both **A** and **B**. A correct form could instead be either of the following:

$$s = A^{\cdot j}_{k\cdot}B^{j\cdot}_{\cdot k} \text{ or } s = A^{jk}B_{kj}$$

When the manifold is orthogonal, as is often tacitly assumed to be the case (for example in Cartesian or spherical coordinates), the metric is the identity, such that none of these considerations matter. However, for essentially all of modern physics - electromagnetism, relativity, quantum field theory, and so on - it is cumbersome or impossible to impose such orthonormality, and the correct formulation of these theories necessitates the use of tensors. In particular, they provide a natural description for nonorthogonal orbital-based theories in quantum chemistry, where imposition of an orthogonality constraint on the underlying function space introduces several computational problems.

B | THE ALMO ERROR VECTOR FOR DIIS

The ALMO orbitals can be rotated by applying a unitary operator $\hat{U}$, parametrised as

$$\hat{U} = \exp(-\hat{\Delta})$$

where the $\hat{\Delta}$ is some anti-Hermitian (necessary to ensure the unitarity of $\hat{U}$) operator, constrained to only rotate orbitals within each fragment so as to maintain localisation. This means that it splits into disjoint operators $\hat{\Delta}^X$. To first order, the transformed orbitals $\{|\tilde{\phi}\rangle\}$ are thus given by

$$|\tilde{\phi}\rangle = \hat{U}|\phi\rangle \approx (1 - \hat{\Delta})|\phi\rangle$$

From equation 2.49, the density operator therefore transforms as

$$\tilde{P} \approx \sum_{X,Y} (1 - \hat{\Delta}^X)|\phi_{Xi}\rangle\, \tilde{\sigma}^{Xi,Yj}\, \langle\phi_{Yj}|\, (1 + \Delta^{\dagger Y})$$

Taking expectations in the AO basis, the linear response in the density is then given by

$$
\begin{aligned}
\Delta P^{X\mu,Y\nu} &\approx r_1^{X\mu,Y\nu} - r_2^{X\mu,Y\nu} + \left(r_2^{\dagger}\right)^{X\mu,Y\nu} \\
r_1^{X\mu,Y\nu} &= T_{\cdot Xi}^{X\mu\cdot}\, \tilde{\sigma}^{Xi,Yj} T_{Yj\cdot}^{\cdot Y\nu} \\
r_2^{X\mu,Y\nu} &= \left(\Delta^X\right)^{X\mu,X\lambda} S_{X\lambda,X\gamma} T_{\cdot Xi}^{X\gamma\cdot}\, \tilde{\sigma}^{Xi,Yj} T_{Yj\cdot}^{\cdot Y\nu}
\end{aligned}
\tag{B.1}
$$

where we have used the fact that Δ is block-diagonal to eliminate the summation over fragments. This is in terms of the transformed metric, which from equation 2.42 becomes

$$
\begin{aligned}
\tilde{\sigma}_{Xi,Yj} &\approx \sigma_{Xi,Yj} + \langle\phi_{Xi}|\hat{\Delta}^{X\dagger}|\phi_{Yj}\rangle - \langle\phi_{Xi}|\hat{\Delta}^Y|\phi_{Yj}\rangle \\
&= \left[\mathbf{T}_X^X \left(\mathbf{S}_{XY} + \mathbf{S}_{XX}\left(\Delta^X\right)^{\dagger}\mathbf{S}_{XY} - \mathbf{S}_{XY}\Delta^Y\mathbf{S}_{YY}\right)\mathbf{T}_Y^Y\right]_{Xi,Yj}
\end{aligned}
\tag{B.2}
$$

By the chain rule, and the results for the ALMO energy gradient with respect to the coefficients, we have that

$$\frac{\partial E}{\partial \Delta^{X\mu,X\nu}} = \sum_{W,Z} F_{W\gamma,Z\kappa}\, \frac{\partial P^{W\gamma,Z\kappa}}{\partial \Delta^{X\mu,X\nu}}$$

Equation B.1 then leads to

$$\frac{\partial \left(P_\alpha\right)^{A\gamma,B\kappa}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}} = T^{A\gamma\cdot}_{\cdot Ai} \frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}} T^{\cdot B\kappa}_{Bj\cdot} - \delta^{A\gamma}_{X\zeta}\delta^{A\lambda}_{X\tau}S_{A\lambda,A\mu}T^{A\mu\cdot}_{\cdot Ai}\tilde{\sigma}^{Ai,Bj}T^{\cdot B\kappa}_{Bj\cdot}$$

$$- \left(\Delta^A\right)^{A\gamma,A\lambda} S_{A\lambda,A\mu}T^{A\mu\cdot}_{\cdot Ai}\frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}}T^{\cdot B\kappa}_{Bj\cdot}$$

$$-T^{A\gamma\cdot}_{\cdot Ai}\tilde{\sigma}^{Ai,Bj}T^{\cdot B\nu}_{Bj\cdot}S_{B\nu,B\lambda}\delta^{B\lambda}_{X\tau}\delta^{B\kappa}_{X\zeta}$$

$$- T^{A\gamma\cdot}_{\cdot Ai}\frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}}T^{\cdot B\nu}_{Bj\cdot}S_{B\nu,B\lambda}\left(\Delta^B\right)^{B\kappa,B\lambda}$$

Then from equation B.2

$$\frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}} = -\sum_{M,N}\tilde{\sigma}^{Ai,Mm}\frac{\partial \tilde{\sigma}_{Mm,Nn}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}}\tilde{\sigma}^{Nn,Bj}$$

$$= \sum_M \tilde{\sigma}^{Ai,Mm}T^{\cdot M\gamma}_{Mm\cdot}S_{M\gamma,X\zeta}S_{X\tau,X\kappa}T^{X\kappa\cdot}_{\cdot Xn}\tilde{\sigma}^{Xn,Bj}$$

$$-\sum_N \tilde{\sigma}^{Ai,Xm}T^{\cdot X\gamma}_{Xm\cdot}S_{X\gamma,X\zeta}S_{X\tau,N\kappa}T^{N\kappa\cdot}_{\cdot Nn}\tilde{\sigma}^{Nn,Bj}$$

which when put in the above, using the fact that $P^{A\gamma,X\pi} = T^{A\gamma\cdot}_{\cdot Ai}\tilde{\sigma}^{Ai,Mm}T^{\cdot X\pi}_{Mm\cdot}$, implies that

$$T^{A\gamma\cdot}_{\cdot Ai}\frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}}T^{\cdot B\kappa}_{Bj\cdot} = -\sum_N P^{A\gamma,X\pi}S_{X\pi,X\zeta}S_{X\tau,N\eta}P^{N\eta,B\kappa}$$

$$+\sum_M P^{A\gamma,M\pi}S_{M\pi,X\zeta}S_{X\tau,X\eta}P^{X\eta,B\kappa}$$

$$= -P^{A\gamma,X\pi}S_{X\pi,X\zeta}[\mathbf{SP}]^{\cdot B\kappa}_{X\tau\cdot} + [\mathbf{PS}]^{A\gamma\cdot}_{\cdot X\zeta}S_{X\tau,X\eta}P^{X\eta,B\kappa}$$

Therefore, we have that (using the symmetry of the Fock and overlap matrices):

$$\sum_{A,B}\left(F_\alpha\right)_{A\gamma,B\kappa}T^{A\gamma\cdot}_{\cdot Ai}\frac{\partial \tilde{\sigma}^{Ai,Bj}}{\partial \left(\Delta_\alpha\right)^{X\zeta,X\tau}}T^{\cdot B\kappa}_{Bj\cdot} =$$

$$\sum_{A,B}\left(F_\alpha\right)_{A\gamma,B\kappa}\left[-P^{A\gamma,X\pi}S_{X\pi,X\zeta}[\mathbf{SP}]^{\cdot B\kappa}_{X\tau\cdot} + [\mathbf{PS}]^{A\gamma\cdot}_{\cdot X\zeta}S_{X\tau,X\eta}P^{X\eta,B\kappa}\right]$$

$$= -[\mathbf{SPF}_\alpha\mathbf{P}]^{X\pi}_{X\tau\cdot}S_{X\pi,X\zeta} + S_{X\tau,X\eta}[\mathbf{PF}_\alpha\mathbf{PS}]^{X\eta\cdot}_{\cdot X\zeta}$$

Inserting the derivative of the inverse metric into the other terms of the derivative of the density that need it results in those terms cancelling, so that all that is left to add to the energy derivative are terms of the form

$$\sum_{A,B}\delta^{A\gamma}_{X\zeta}\delta^{A\lambda}_{X\tau}\left(F_\alpha\right)_{A\gamma,B\kappa}S_{A\lambda,A\mu}T^{A\mu\cdot}_{\cdot Ai}\tilde{\sigma}^{Ai,Bj}T^{\cdot B\kappa}_{Bj\cdot} =$$

$$S_{X\tau,X\mu}\sum_B\left(F_\alpha\right)_{X\zeta,B\kappa}P^{X\mu,B\kappa} = S_{X\tau,X\mu}[\mathbf{PF}_\alpha]^{X\mu\cdot}_{\cdot X\zeta}$$

Thus, the energy derivative is overall given by

$$\frac{\partial E}{\partial \left(\Delta_\alpha\right)^{X\zeta, X\tau}} = S_{X\tau, X\eta} \left[\mathbf{PF}_\alpha \mathbf{PS} - \mathbf{PF}_\alpha\right]^{X\eta\cdot}_{\cdot X\zeta} - \left[\mathbf{SPF}_\alpha \mathbf{P} - \mathbf{F}_\alpha \mathbf{P}\right]^{\cdot X\pi}_{X\tau\cdot} S_{X\pi, X\zeta}$$

so that the stationarity condition is:

$$\frac{\partial E}{\partial \left(\Delta_\alpha\right)^{X\zeta, X\tau}} = S_{X\tau, X\eta} \left[\mathbf{PF}_\alpha (\mathbf{PS} - \mathbf{1})\right]^{X\eta\cdot}_{\cdot X\zeta} - \left[(\mathbf{SP} - \mathbf{1})\mathbf{F}_\alpha \mathbf{P}\right]^{\cdot X\pi}_{X\tau\cdot} S_{X\pi, X\zeta} = 0$$

$$(\text{B.3})$$

C | CHARGE DECOMPOSITION ANALYSIS

The charge transferred from occupied orbital ϕ_{Xi} into virtual orbital $\overline{\phi}_{Ya}$ is the difference between the electron density before and after relaxation from the ALMO solution. That is, if the transformed total density is $\hat{P} = \hat{U}\hat{P}_0\hat{U}^\dagger$, then the amount of charge transferred is this density projected onto the original occupied space, minus the original density. The total charge transferred would thus be given by

$$\Delta Q = \text{Tr}\left\{\hat{P}_0\hat{P}\hat{P}_0 - \hat{P}_0\right\}$$

This is not very useful, as solving equation 3.15 gives $\hat{X}$, rather than $\hat{U}$, such that $\hat{P}$ cannot be directly formed. Following the same procedure as was used to reach equation 3.15, we rewrite the density transformation as $\hat{P}\hat{U} = \hat{U}\hat{P}_0$, then take just the occupied-occupied block. Note, however, that the idempotency of the density simplifies matters. In the discrete representation, this leads to

$$\mathbf{P}_{OO}\mathbf{U}_{OO} + \mathbf{P}_{OV}\mathbf{U}_{VO} = \mathbf{U}_{OO}$$

which upon multiplying through on the right by $\mathbf{U}_{OO}^{-1}$ gives

$$\mathbf{P}_{OO} + \mathbf{P}_{OV}\mathbf{X}_{VO} = \mathbf{I}_{OO}$$

Given that $\text{Tr}\{\mathbf{P}_0\} = \text{Tr}\{\mathbf{I}_{OO}\}$, this means the total charge transferred is

$$\Delta Q = \text{Tr}\{\mathbf{P}_{OV}\mathbf{X}_{VO}\} \tag{C.1}$$

While this does not remove the need to form the full density, it does simplify the problem somewhat.

It is customary to represent the rotation operator, $\hat{U}$, as an exponential of an antihermitian operator. In the current instance, this would be taken to be $\hat{X} - \hat{X}^\dagger$, which is antihermitian by construction. Alternatively, and entirely equivalently, we can use this as the Cayley generator for the unitary transformation:

$$\hat{U} = \left(\hat{I} + \hat{X} - \hat{X}^\dagger\right)\left(\hat{I} + \hat{X}^\dagger\hat{X} + \hat{X}\hat{X}^\dagger\right)^{-1/2} \tag{C.2}$$

This or the exponential form can then be approximated to first order as

$$\hat{U} = 1 + \hat{X} - \hat{X}^\dagger + \mathcal{O}\left(\hat{X}^2\right)$$

allowing the density to be formed as

$$\hat{P} \approx \left(1 + \hat{X} - \hat{X}^\dagger\right)\hat{P}_0\left(1 + \hat{X}^\dagger - \hat{X}\right) \tag{C.3}$$

Using the fact that $\hat{X} = \hat{Q}_0 \hat{X} \hat{P}_0$, this means that equation C.1 to first order in $\hat{X}$ is given by

$$\Delta Q \approx \text{Tr}\left\{\mathbf{X}^\mathsf{T}\mathbf{X}\right\} \tag{C.4}$$

Finally, we introduce the partition operators of equations 3.17 and 3.18, allowing the separation of charge transfer into terms between fragments. This yields, directly from equation C.1

$$\begin{aligned}
\Delta Q_{\text{CT, M}}^{Y \to Z} &= \sum_{i,a} \text{Tr}\left\{\hat{p}_{Yi} \hat{P} \hat{q}_{Za} \hat{X}\right\} \\
&= \sum_W \langle \phi^{Wk}|\phi_{Yi}\rangle \langle \phi^{Yi}|\hat{P}|\overline{\phi}_{Za}\rangle \langle \overline{\phi}^{Xa}|\hat{X}|\phi_{Wk}\rangle \\
&= \text{Tr}\left\{(\mathbf{P}_{OV})_{YZ}\,(\mathbf{X}_{VO})_{ZY}\right\}
\end{aligned}$$

which is equation 3.20. This requires that the transform matrix $\mathbf{U}$ be explicitly formed from equation C.2, however, which in turn necessitates a singular value decomposition. If such a charge decomposition is required, it is therefore simpler to use equation C.4, resulting in

$$\Delta Q^{Y \to Z} \approx \text{Tr}\left\{\left(\mathbf{X}_{OV}^\mathsf{T}\right)_{YZ}(\mathbf{X}_{VO})_{ZY}\right\} \tag{C.5}$$

D | ELECTRON REPULSION INTEGRALS

Consider a spherical Gaussian charge distribution, ρ_p, centred at $\mathbf{P}$ and with exponent p:

$$\rho_p(\mathbf{r}_P) = \left(\frac{p}{\pi}\right)^{3/2} \exp(-pr_P^2)$$

This is normalised to unity, as it represents a probability density. The repulsion energy between two such distributions in a Coulombic potential is given by

$$V_{pq} = \int \int \frac{\rho_p(\mathbf{r}_1)\rho_q(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 \ d\mathbf{r}_2 \tag{D.1}$$

This can be evaluated explicitly by introducing the integral representation of the potential:

$$\frac{1}{r} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-r^2 t^2) dt$$

The electrostatic potential at a point $\mathbf{s}$ due to the charge distribution, ρ_p, is thus given by

$$V_p(\mathbf{s}) = \int \frac{\rho_p(\mathbf{r})}{|\mathbf{r}-\mathbf{s}|} = \frac{p^{3/2}}{\pi^2} \int \int_{-\infty}^{\infty} \exp\left(-pr^2 - t^2|\mathbf{r}-\mathbf{s}|^2\right) dt \ d\mathbf{r}$$

where we have fixed $\mathbf{P}$ to be the origin for clarity. Using the Gaussian product rule, and defining $\mathbf{S} = t^2\mathbf{s}/(p+t^2)$, this becomes

$$V_p(\mathbf{s}) = \frac{4p^{3/2}}{\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{pt^2}{p+t^2}s^2\right) \int_0^{\infty} r^2 \exp[-(p+t^2)r^2] dr \ dt$$

Evaluating the inner integral simply gives

$$V_p(\mathbf{s}) = \frac{p^{3/2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} (p+t^2)^{-3/2} \exp\left(\frac{-ps^2 t^2}{p+t^2}\right)$$

As the integrand is even, we may restrict the region of integration to the positive half of the real line. Making the substitution $\tau^2 = t^2/(p+t^2)$ then reduces these limits to the interval $[0, 1]$, giving the final form below:

$$V_p(\mathbf{s}) = \sqrt{\frac{4p}{\pi}} \int_0^1 \exp(-ps^2\tau^2) d\tau \tag{D.2}$$

We then define the n-th Boys function, F_n, to be

$$F_n(\tau) = \int_0^1 \tau^{2n} \exp(-x\tau^2) d\tau \qquad \text{(D.3)}$$

such that equation D.2 is clearly just a rescaled Boys function, $F_0(ps^2)$. We can then replace one of the integrals in equation D.1 with this electrostatic potential:

$$V_{pq} = \sqrt{\frac{4p}{\pi}} \left(\frac{q}{\pi}\right)^{3/2} \int F_0(p|\mathbf{r} - \mathbf{P}|^2) \exp(-q|\mathbf{r} - \mathbf{Q}|^2) d\mathbf{r}$$

Applying equation D.3 and using the Gaussian product rule once more, this becomes

$$V_{pq} = \sqrt{\frac{4p}{\pi}} \left(\frac{q}{\pi}\right)^{3/2} \int_0^1 \int \exp\left\{-\frac{pqt^2 R_{PQ}^2}{pt^2 + q} - (pt^2 + q)r^2\right\} d\mathbf{r}\, dt$$

where $R_{PQ} = |\mathbf{P} - \mathbf{Q}|$ is the separation between the two centres. Again, this allows the spatial integration to be performed, leaving

$$V_{pq} = \sqrt{\frac{4pq}{\pi}} \int_0^1 \frac{q}{(pt^2 + q)^{3/2}} \exp\left(-\frac{pqt^2 R_{PQ}^2}{pt^2 + q}\right)$$

Substituting $\tau^2 = (p + q)t^2/(pt^2 + q)$ and defining the reduced exponent $\alpha = pq/(p + q)$ then leads to the final result:

$$V_{pq} = \sqrt{\frac{4\alpha}{\pi}} \int_0^1 \exp(-\alpha R_{PQ}^2 \tau^2) d\tau = \sqrt{\frac{4\alpha}{\pi}} F_0(\alpha R_{PQ}^2) \qquad \text{(D.4)}$$

Note that this is also in terms of the zeroth Boys function; for the general case where the Gaussian distributions have factors of r^n (i.e. higher angular momentum), it can be shown in much the same way as above that the integral follows $F_n(\beta R_{PQ}^2)$, leading to equation 4.5.

E | BENCHMARK DATA

This appendix contains the benchmark data calculated in Chapter 5 for the S66 [267] and X40 [333] datasets. We give results for the following methods: HF, ALMO, ALMO+CT, DF-MP2, DF-SAPT2, M06-2X, CCSD, and the dRPA, SOSEX, and RPAx versions of ALMO+RPA. The SAPT energy decomposition was grouped according to the following scheme, as used by default in Psi4 [328]:

$$E_{\text{elst.}} = E_{\text{elst.}}^{(10)} + E_{\text{elst., r}}^{(12)} + E_{\text{elst., r}}^{(13)}$$

$$E_{\text{ex.}} = E_{\text{ex.}}^{(10)} + E_{\text{ex.}}^{(11)} + E_{\text{ex.}}^{(12)}$$

$$E_{\text{ind.}} = E_{\text{ind., r}}^{(20)} + E_{\text{ex., ind., r}}^{(20)} + E_{\text{ind.}}^{(30)} + E_{\text{ex., ind.}}^{(30)} + E_{\text{ind.}}^{(22)} + E_{\text{ex., ind.}}^{(22)} + \delta_{\text{HF}}^3$$

$$E_{\text{disp.}} = E_{\text{disp.}}^{(20)} + E_{\text{disp.}}^{(21)} + E_{\text{disp.}}^{(22)} + E_{\text{ex., disp.}}^{(20)} + E_{\text{ind., disp.}}^{(20)} + E_{\text{disp.}}^{(30)}$$
$$+ E_{\text{ex., disp.}}^{(30)} + E_{\text{ex., ind., disp.}}^{(30)}$$

S66 BENCHMARKS

The following shorthands are used for the S66 molecule names to save space in tables: benzene, Bz.; uracil, Ur.; peptide, Pep.; pyridine, Pyr.; pentane, Pent.

Table E.1: Uncorrelated interaction energies in kcal mol^{-1} for the molecules in the S66 database using aVDZ basis sets. Counterpoise-corrected Hartree-Fock (HF) is compared with the ALMO + charge transfer (CT) energies, along with the components: CT and local exchange correction (XCorrect) energies, and the ALMO energy (including XCorrect).

ID	Donor	Acceptor	HF	ALMO+CT	ALMO	XCorrect	CT
1	H_2O	H_2O	-3.7319	-3.7622	-2.8451	0.0734	-0.9171
2	H_2O	MeOH	-3.7816	-3.8694	-2.6902	0.0176	-1.1792
3	H_2O	MeNH$_2$	-4.7275	-4.7039	-2.9612	0.0533	-1.7427
4	H_2O	Pep.	-5.9168	-6.0111	-4.3691	0.4926	-1.6419
5	MeOH	MeOH	-3.5277	-3.6347	-2.4209	0.0132	-1.2139
6	MeOH	MeNH$_2$	-4.2225	-4.2921	-2.4212	0.0019	-1.8709
7	MeOH	Pep.	-5.0263	-5.1416	-3.4414	-1.0950	-1.7002
8	MeOH	H_2O	-3.5005	-3.5491	-2.6023	0.0364	-0.9468
9	MeNH$_2$	MeOH	-0.8049	-1.0769	-0.3308	0.0088	-0.7461
10	MeNH$_2$	MeNH$_2$	-1.0839	-1.3463	-0.3690	0.0107	-0.9773
11	MeNH$_2$	Pep.	-1.2614	-1.6540	-0.5340	0.0188	-1.1201
12	MeNH$_2$	H_2O	-4.7543	-4.7765	-2.9145	0.0069	-1.8620
13	Pep.	MeOH	-3.0324	-3.3530	-2.2052	-0.7091	-1.1478
14	Pep.	MeNH$_2$	-3.4517	-3.7538	-2.0242	-0.8402	-1.7296
15	Pep.	Pep.	-4.4891	-4.9058	-3.2124	-0.9758	-1.6934
16	Pep.	H_2O	-3.3832	-3.6300	-2.7369	-0.5422	-0.8931
17	Ur.	Ur. BP	-13.3455	-13.1239	-8.4967	-1.8010	-4.6272
18	H_2O	Pyr.	-4.4513	-4.4266	-2.8885	0.0552	-1.5382
19	MeOH	Pyr.	-4.1911	-4.1797	-2.5478	-0.9639	-1.6320
20	AcOH	AcOH	-15.1633	-14.5807	-9.6235	-2.1712	-4.9572

ID	Donor	Acceptor	HF	ALMO+CT	ALMO	XCorrect	CT
21	$AcNH_2$	$AcNH_2$	-12.6204	-12.4206	-9.1219	-1.6554	-3.2987
22	AcOH	Ur.	-15.6672	-15.3103	-10.8245	-1.8963	-4.4858
23	$AcNH_2$	Ur.	-15.5342	-15.1060	-11.1846	-2.0708	-3.9214
24	Bz.	Bz. pipi	4.3623	3.2432	4.6232	-0.5409	-1.3800
25	Pyr.	Pyr. pipi	3.6080	2.7399	4.0081	0.1186	-1.2682
26	Ur.	Ur. pipi	3.1525	1.6438	4.0957	0.1092	-2.4518
27	Bz.	Pyr. pipi	3.8047	2.8045	4.1153	-0.5102	-1.3107
28	Bz.	Ur. pipi	7.4911	5.9739	8.2716	0.4336	-2.2978
29	Pyr.	Ur. pipi	5.7661	4.5512	6.7267	0.2956	-2.1755
30	Bz.	Ethene	3.3546	3.0932	3.7338	0.0107	-0.6407
31	Ur.	Ethene	1.4819	1.1805	1.7861	0.0420	-0.6056
32	Ur.	Ethyne	0.1296	-0.0970	0.4225	0.0358	-0.5195
33	Pyr.	Ethene	5.6291	5.2505	6.4783	0.0176	-1.2278
34	Pent.	Pent.	4.8194	4.0807	5.2571	0.2041	-1.1764
35	Neopent.	Pent.	3.7167	3.0561	4.0945	-0.6181	-1.0384
36	Neopent.	Neopent.	2.8743	2.3238	3.2047	0.5158	-0.8808
37	Cyclopent.	Neopent.	3.7731	3.1339	4.1566	-0.5974	-1.0227
38	Cyclopent.	Cyclopent.	3.7608	3.2100	4.1813	-0.6752	-0.9713
39	Bz.	Cyclopent.	3.5157	2.4057	3.7659	0.3552	-1.3601
40	Bz.	Neopent.	2.9327	1.8818	3.1615	-0.6338	-1.2797
41	Ur.	Pent.	5.6986	4.5772	6.3512	0.4436	-1.7741
42	Ur.	Cyclopent.	5.3810	4.3184	5.9176	1.0235	-1.5992
43	Ur.	Neopent.	3.7471	2.8434	4.1778	-0.7769	-1.3343
44	Ethene	Pent.	1.9421	1.6493	2.1535	0.0075	-0.5043

ID	Donor	Acceptor	HF	ALMO+CT	ALMO	XCorrect	CT
45	Ethyne	Pent.	1.4908	1.2487	1.6786	0.0126	-0.4299
46	Pep.	Pent.	4.2688	3.3366	4.8002	0.3031	-1.4635
47	Bz.	Bz. TS	1.8333	1.1770	2.0213	-0.2629	-0.8443
48	Pyr.	Pyr. TS	1.0617	0.4361	1.3900	-0.5045	-0.9539
49	Bz.	Pyr. TS	1.3050	0.3301	1.4576	-0.4876	-1.1275
50	Bz.	Ethyne CH-pi	-0.0664	-0.9465	0.0886	-0.4405	-1.0351
51	Ethyne	Ethyne TS	-0.4296	-0.6047	-0.1271	-0.2792	-0.4776
52	Bz.	AcOHOH-pi	-0.5764	-1.4018	-0.2156	-0.4826	-1.1862
53	Bz.	AcNH$_2$NH-pi	0.0772	-0.5431	0.3249	-0.7599	-0.8680
54	Bz.	H$_2$OOH-pi	-0.8199	-1.3626	-0.6412	0.3753	-0.7214
55	Bz.	MeOHOH-pi	0.3270	-0.2328	0.6622	0.3985	-0.8950
56	Bz.	MeNH$_2$NH-pi	1.3754	0.5799	1.5495	0.0458	-0.9695
57	Bz.	Pep. NH-pi	3.9026	2.5914	4.4824	0.0075	-1.8909
58	Pyr.	Pyr. CH-N	-1.2536	-1.5936	-0.7315	-0.5773	-0.8622
59	Ethyne	H$_2$OCH-O	-2.3106	-2.5380	-1.8731	-0.3338	-0.6649
60	Ethyne	AcOHOH-pi	-2.8891	-2.9724	-1.9249	-0.8340	-1.0475
61	Pent.	AcOH	2.3075	1.5962	2.5982	0.0157	-1.0021
62	Pent.	AcNH$_2$	2.2796	1.7092	2.6923	0.1349	-0.9830
63	Bz.	AcOH	1.7734	0.7621	1.9988	-0.4826	-1.2367
64	Pep.	Ethene	0.5280	0.2065	0.8991	0.0107	-0.6927
65	Pyr.	Ethyne	-2.7363	-2.9583	-1.9821	-0.4179	-0.9762
66	MeNH$_2$	Pyr.	0.9678	0.4815	1.4558	0.0151	-0.9743

Table E.2: Correlated aVDZ interaction energies in kcal mol^{-1} for the S66 database, calculated using CCSD, DF-MP2, Mo6-2X, DF-SAPT2, with all but the latter being counterpoise corrected. These are compared with ALMO+RPAd results, with all three RPA variants. The dispersion terms are also given for reference.

Donor	Acceptor	CCSD	MP2	Mo6-2X	SAPT2	ALMO+RPAd			Dispersion		
						dRPA	SOSEX	RPAx	dRPA	SOSEX	RPAx
H$_2$O	H$_2$O	-4.2639	-4.4684	-5.0384	-4.3318	-4.4276	-4.4179	-4.3421	-0.6654	-0.6561	-0.6561
H$_2$O	MeOH	-4.8000	-5.1579	-5.7036	-4.9873	-4.5669	-4.4713	-4.0909	-0.6975	-0.6024	-0.3600
H$_2$O	MeNH$_2$	-5.9442	-6.4418	-6.9445	-6.3798	-5.1785	-5.1859	-4.8128	-0.4746	-0.4832	-0.3280
H$_2$O	Pep.	-6.9628	-7.2155	-8.3496	-7.0732	-6.2232	-6.1345	-5.8179	-0.2121	-0.1242	-0.0454
MeOH	MeOH	-4.8357	-5.2315	-5.5686	-7.1933	-4.9835	-4.8840	-4.2841	-1.3488	-1.2497	-0.7938
MeOH	MeNH$_2$	-6.2431	-6.8694	-7.3870	-5.0742	-6.1652	-5.9248	-5.0570	-1.8731	-1.6340	-0.9879
MeOH	Pep.	-6.8028	-7.1907	-8.1575	-6.8414	-5.7104	-5.8057	-5.3863	-0.5687	-0.6648	-0.4709
MeOH	H$_2$O	-4.2657	-4.4972	-4.9214	-3.2181	-4.3602	-4.3479	-4.0034	-0.8111	-0.7992	-0.5347
MeNH$_2$	MeOH	-2.6283	-2.9940	-3.4123	-4.9256	-2.4682	-2.4683	-1.8247	-1.3913	-1.3915	-0.9295
MeNH$_2$	MeNH$_2$	-3.2117	-3.7372	-4.0738	-2.9988	-2.8217	-2.7941	-2.0888	-1.4754	-1.4480	-0.9564
MeNH$_2$	Pep.	-4.1712	-4.8601	-5.6113	-3.9155	-2.8792	-2.9201	-2.2283	-1.2252	-1.2662	-0.8431
MeNH$_2$	H$_2$O	-6.2235	-6.8198	-7.4666	-6.8289	-6.1989	-6.1526	-5.4152	-1.4224	-1.3773	-0.9058
Pep.	MeOH	-5.1229	-5.6995	-6.0696	-8.0699	-5.1841	-5.2172	-4.3920	-1.8312	-1.8645	-1.2668
Pep.	MeNH$_2$	-6.1507	-6.9596	-7.2590	-5.6841	-5.8864	-5.7823	-4.7666	-2.1325	-2.0294	-1.3176
Pep.	Pep.	-7.2263	-7.9482	-8.6299	-4.6258	-6.9645	-7.0982	-6.0901	-2.0588	-2.1930	-1.5118
Pep.	H$_2$O	-4.3951	-4.6543	-4.9784	-4.2210	-4.6620	-4.6376	-4.1475	-1.0319	-1.0078	-0.6762
Ur.	Ur. BP	-15.1794	-16.1191	-16.9832	-17.0732	-15.8775	-15.9215	-14.5011	-2.7536	-2.8018	-1.8855
H$_2$O	Pyr.	-5.8241	-6.3755	-6.5252	-6.3536	-4.7371	-4.5994	-4.2759	-0.3105	-0.1735	-0.0571
MeOH	Pyr.	-6.1228	-6.8423	-6.7294	-4.3554	-4.4133	-4.3873	-4.1136	-0.2335	-0.2085	-0.1287

Donor	Acceptor	CCSD	MP2	Mo6-2X	SAPT2	ALMO+RPAd			Dispersion		
						dRPA	SOSEX	RPAx	dRPA	SOSEX	RPAx
AcOH	AcOH	-15.9299	-16.4436	-18.7719	-16.9961	-17.2209	-16.9930	-15.6451	-2.6402	-2.4173	-1.5262
AcNH$_2$	AcNH$_2$	-14.2056	-14.6878	-15.9267	-14.9505	-15.3042	-15.0235	-13.7724	-2.8836	-2.6052	-1.6225
AcOH	Ur.	-16.7726	-17.3591	-18.9848	-17.6936	-17.5564	-17.1366	-15.9177	-2.2460	-1.8303	-1.0486
AcNH$_2$	Ur.	-16.8483	-17.4457	-18.6225	-17.6968	-16.7682	-16.7608	-15.8729	-1.6622	-1.6582	-1.0986
Bz.	Bz. pipi	-1.1136	-4.2452	-2.8317	-3.4518	-0.8119	-0.7658	1.0824	-4.0551	-4.0091	-2.6790
Pyr.	Pyr. pipi	-2.0513	-5.4484	-3.8787	-5.4973	-1.7716	-1.6946	0.1509	-4.5115	-4.4346	-2.9426
Ur.	Ur. pipi	-5.7761	-9.6275	-10.4167	-11.3369	-5.3138	-5.4065	-2.3903	-6.9577	-7.0507	-4.7286
Bz.	Pyr. pipi	-1.6957	-4.9115	-3.4640	-3.8878	-1.4632	-1.3902	0.4604	-4.2677	-4.1949	-2.7880
Bz.	Ur. pipi	-1.7075	-6.1978	-5.6770	-7.5117	0.0200	0.1794	2.8520	-5.9538	-5.7948	-3.7953
Pyr.	Ur. pipi	-2.9276	-7.2970	-6.7295	-8.6041	-1.1168	-1.0717	1.3983	-5.6680	-5.6233	-3.7254
Bz.	Ethene	-0.2831	-1.9546	-1.7058	-3.1885	0.3974	0.3801	1.3379	-2.6958	-2.7132	-1.8486
Ur.	Ethene	-2.0769	-3.4187	-3.6297	-3.4943	-1.6562	-1.7264	-0.8078	-2.8367	-2.9069	-1.9788
Ur.	Ethyne	-2.5668	-3.7661	-4.0801	-3.8697	-2.6354	-2.5544	-1.7683	-2.5384	-2.4575	-1.6748
Pyr.	Ethene	0.4145	-2.1186	-1.7702	-2.7545	1.4963	1.4794	2.8311	-3.7542	-3.7715	-2.5598
Pent.	Pent.	-1.8065	-3.1507	-3.9172	-3.9122	-0.0464	-0.2587	1.5079	-4.1272	-4.3395	-2.9790
Neopent.	Pent.	-1.0624	-1.9790	-2.3069	-2.6775	-0.1244	-0.3388	1.1214	-3.1805	-3.3949	-2.3439
Neopent.	Neopent.	-0.6391	-1.2213	-0.9611	-1.7699	-0.2162	-0.3275	0.8699	-2.5401	-2.6513	-1.8085
Cyclopent.	Neopent.	-0.9212	-1.8694	-2.0975	-2.4867	0.1437	-0.1016	1.2694	-2.9903	-3.2356	-2.2494
Cyclopent.	Cyclopent.	-1.4099	-2.4967	-2.5869	-3.0594	-0.0436	-0.1934	1.1954	-3.2536	-3.4034	-2.3294
Bz.	Cyclopent.	-1.9264	-4.0600	-3.5890	-1.9955	-1.3992	-1.3454	0.3896	-3.8049	-3.7513	-2.5028
Bz.	Neopent.	-1.4821	-3.1335	-2.5548	-5.0259	-1.3026	-1.2865	0.2363	-3.1844	-3.1685	-2.1230
Ur.	Pent.	-2.0228	-4.2927	-5.1135	-5.5089	-0.6885	-0.8239	1.4981	-5.2656	-5.4012	-3.6793
Ur.	Cyclopent.	-1.5918	-3.6923	-3.9408	-4.7203	-0.6289	-0.6835	1.4936	-4.9474	-5.0021	-3.3834

Donor	Acceptor	CCSD	MP2	M06-2X	SAPT2	ALMO+RPAd			Dispersion		
						dRPA	SOSEX	RPAx	dRPA	SOSEX	RPAx
Ur.	Neopent.	-1.7374	-3.2615	-3.5897	-4.1698	-0.9114	-1.0219	0.6805	-3.7549	-3.8654	-2.6383
Ethene	Pent.	-1.1005	-1.8211	-2.0734	-4.3269	0.9398	0.9332	1.2794	-0.7095	-0.7161	-0.4847
Ethyne	Pent.	-0.9638	-1.7541	-1.8794	-2.7244	1.2413	1.2471	1.3471	-0.0074	-0.0017	0.0010
Pep.	Pent.	-2.0587	-3.5208	-4.7327	-4.5582	-0.6450	-0.7694	1.0552	-3.9816	-4.1061	-2.7925
Bz.	Bz. TS	-1.6315	-3.3085	-2.5920	-4.1459	-1.5783	-1.5405	-0.3776	-2.7553	-2.7176	-1.8120
Pyr.	Pyr. TS	-2.2652	-3.8855	-3.3253	-3.9973	-2.1879	-2.2386	-0.9934	-2.6240	-2.6748	-1.7957
Bz.	Pyr. TS	-2.0580	-3.6922	-3.1060	-3.0140	-2.3149	-2.2481	-0.9621	-2.6450	-2.5783	-1.7087
Bz.	Ethyne CH-pi	-1.9314	-3.0020	-2.8772	-3.3278	-2.8913	-2.8268	-1.7835	-1.9448	-1.8804	-1.2534
Ethyne	Ethyne TS	-1.0557	-1.3818	-1.3971	-1.8100	-1.0996	-1.0978	-0.8217	-0.4949	-0.4932	-0.3331
Bz.	AcOHOH-pi	-3.2617	-3.9235	-4.7793	-4.5859	-3.8778	-3.7651	-2.6233	-2.4760	-2.3635	-1.5505
Bz.	AcNH$_2$NH-pi	-2.9933	-4.2913	-4.4487	-4.5478	-3.3392	-3.3013	-2.2017	-2.7962	-2.7583	-1.8407
Bz.	H$_2$OOH-pi	-2.3720	-3.0086	-3.5717	-1.8431	-2.9472	-2.9013	-2.1665	-1.5846	-1.5388	-1.0248
Bz.	MeOHOH-pi	-2.8229	-4.0539	-4.4075	-3.2775	-2.9033	-2.8781	-1.8496	-2.6705	-2.6453	-1.7771
Bz.	MeNH$_2$NH-pi	-1.9772	-3.2721	-3.3635	-4.0004	-2.1116	-2.0793	-0.8794	-2.6915	-2.6594	-1.7852
Bz.	Pep. NH-pi	-2.6145	-5.3263	-5.5759	-6.0766	-2.1539	-2.0617	0.0531	-4.7453	-4.6535	-3.0886
Pyr.	Pyr. CH-N	-3.2275	-3.9264	-3.5616	-4.2187	-3.3052	-3.3190	-2.5153	-1.7115	-1.7254	-1.1625
Ethyne	H$_2$OCH-O	-2.4888	-2.6384	-2.8916	-3.7868	-3.0786	-3.0753	-2.7384	-0.5406	-0.5374	-0.3678
Ethyne	AcOHOH-pi	-3.7617	-4.2136	-5.2408	-1.4283	-3.3046	-3.2888	-3.0743	-0.3322	-0.3167	-0.2102
Pent.	AcOH	-1.6367	-2.4715	-2.9103	-2.8130	-1.5010	-1.5432	-0.1498	-3.0972	-3.1394	-2.1076
Pent.	AcNH$_2$	-2.0814	-3.0256	-3.6023	-2.5819	-1.5771	-1.6781	-0.2875	-3.2864	-3.3874	-2.2941
Bz.	AcOH	-2.2788	-4.4359	-4.0921	-4.1297	-2.7496	-2.6992	-1.1501	-3.5117	-3.4615	-2.3061
Pep.	Ethene	-2.0211	-2.7062	-3.0357	-7.0173	-1.9655	-2.0225	-1.1602	-2.1719	-2.2290	-1.5220
Pyr.	Ethyne	-3.4757	-3.9640	-3.7363	-4.1486	-4.2149	-4.2221	-3.5612	-1.2566	-1.2641	-0.8546

Donor	Acceptor	CCSD	MP2	Mo6-2X	SAPT2	ALMO+RPAd dRPA	SOSEX	RPAx	Dispersion dRPA	SOSEX	RPAx
MeNH$_2$	Pyr.	-2.6105	-3.9645	-4.0492	-6.7073	0.0928	0.0679	0.4507	-0.3888	-0.4137	-0.2827

Table E.3: Correlated aVDZ interaction energies in kcal mol^{-1} for the S66 database, calculated using CCSD, DF-MP2, and Mo6-2X, with and without the counterpoise correction.

Donor	Acceptor	CCSD		MP2		Mo6-2X	
		CP	No CP	CP	No CP	CP	No CP
H$_2$O	H$_2$O	-4.2639	-4.2652	-4.4684	-4.2652	-5.0384	-5.0387
H$_2$O	MeOH	-4.8000	-4.8016	-5.1579	-4.8016	-5.7036	-5.7042
H$_2$O	MeNH$_2$	-5.9442	-5.9459	-6.4418	-5.9459	-6.9445	-6.9451
H$_2$O	Pep.	-6.9628	-6.9649	-7.2155	-6.9649	-8.3496	-8.3504
MeOH	MeOH	-4.8357	-4.8376	-5.2315	-4.8376	-5.5686	-5.5692
MeOH	MeNH$_2$	-6.2431	-6.2455	-6.8694	-6.2455	-7.3870	-7.3880
MeOH	Pep.	-6.8028	-6.8055	-7.1907	-6.8055	-8.1575	-8.1583
MeOH	H$_2$O	-4.2657	-4.2670	-4.4972	-4.2670	-4.9214	-4.9217
MeNH$_2$	MeOH	-2.6283	-2.6302	-2.9940	-2.6302	-3.4123	-3.4131
MeNH$_2$	MeNH$_2$	-3.2117	-3.2139	-3.7372	-3.2139	-4.0738	-4.0747
MeNH$_2$	Pep.	-4.1712	-4.1744	-4.8601	-4.1744	-5.6113	-5.6125
MeNH$_2$	H$_2$O	-6.2235	-6.2255	-6.8198	-6.2255	-7.4666	-7.4674
Pep.	MeOH	-5.1229	-5.1256	-5.6995	-5.1256	-6.0696	-6.0707
Pep.	MeNH$_2$	-6.1507	-6.1540	-6.9596	-6.1540	-7.2590	-7.2604
Pep.	Pep.	-7.2263	-7.2304	-7.9482	-7.2304	-8.6299	-8.6313
Pep.	H$_2$O	-4.3951	-4.3971	-4.6543	-4.3971	-4.9784	-4.9792
Ur.	Ur. BP	-15.1794	-15.1845	-16.1191	-15.1845	-16.9832	-16.9894
H$_2$O	Pyr.	-5.8241	-5.8260	-6.3755	-5.8260	-6.5252	-6.5258
MeOH	Pyr.	-6.1228	-6.1250	-6.8423	-6.1250	-6.7294	-6.7301
AcOH	AcOH	-15.9299	-15.9338	-16.4436	-15.9338	-18.7719	-18.7730

Donor	Acceptor	CCSD		MP2		M06-2X	
		CP	No CP	CP	No CP	CP	No CP
AcNH$_2$	AcNH$_2$	-14.2056	-14.2090	-14.6878	-14.2090	-15.9267	-15.9279
AcOH	Ur.	-16.7726	-16.7769	-17.3591	-16.7769	-18.9848	-18.9860
AcNH$_2$	Ur.	-16.8483	-16.8524	-17.4457	-16.8524	-18.6225	-18.6239
Bz.	Bz. pipi	-1.1136	-1.1189	-4.2452	-1.1189	-2.8317	-2.8333
Pyr.	Pyr. pipi	-2.0513	-2.0564	-5.4484	-2.0564	-3.8787	-3.8803
Ur.	Ur. pipi	-5.7761	-5.7817	-9.6275	-5.7817	-10.4167	-10.4256
Bz.	Pyr. pipi	-1.6957	-1.7009	-4.9115	-1.7009	-3.4640	-3.4655
Bz.	Ur. pipi	-1.7075	-1.7171	-6.1978	-1.7171	-5.6770	-5.6841
Pyr.	Ur. pipi	-2.9276	-2.9350	-7.2970	-2.9350	-6.7295	-6.7334
Bz.	Ethene	-0.2831	-0.2857	-1.9546	-0.2857	-1.7058	-1.7067
Ur.	Ethene	-2.0769	-2.0805	-3.4187	-2.0805	-3.6297	-3.6308
Ur.	Ethyne	-2.5668	-2.5697	-3.7661	-2.5697	-4.0801	-4.0808
Pyr.	Ethene	0.4145	0.4079	-2.1186	0.4079	-1.7702	-1.7776
Pent.	Pent.	-1.8065	-1.8088	-3.1507	-1.8088	-3.9172	-3.9197
Neopent.	Pent.	-1.0624	-1.0657	-1.9790	-1.0657	-2.3069	-2.3156
Neopent.	Neopent.	-0.6391	-0.6460	-1.2213	-0.6460	-0.9611	-0.9615
Cyclopent.	Neopent.	-0.9212	-0.9232	-1.8694	-0.9232	-2.0975	-2.1017
Cyclopent.	Cyclopent.	-1.4099	-1.4113	-2.4967	-1.4113	-2.5869	-2.5880
Bz.	Cyclopent.	-1.9264	-1.9324	-4.0600	-1.9324	-3.5890	-3.5909
Bz.	Neopent.	-1.4821	-1.4876	-3.1335	-1.4876	-2.5548	-2.5566
Ur.	Pent.	-2.0228	-2.0263	-4.2927	-2.0263	-5.1135	-5.1213
Ur.	Cyclopent.	-1.5918	-1.5929	-3.6923	-1.5929	-3.9408	-3.9479
Ur.	Neopent.	-1.7374	-1.7377	-3.2615	-1.7377	-3.5897	-3.5912

Donor	Acceptor	CCSD		MP2		Mo6-2X	
		CP	No CP	CP	No CP	CP	No CP
Ethene	Pent.	-1.1005	-1.1030	-1.8211	-1.1030	-2.0734	-2.0744
Ethyne	Pent.	-0.9638	-0.9659	-1.7541	-0.9659	-1.8794	-1.8800
Pep.	Pent.	-2.0587	-2.0666	-3.5208	-2.0666	-4.7327	-4.7367
Bz.	Bz. TS	-1.6315	-1.6366	-3.3085	-1.6366	-2.5920	-2.5934
Pyr.	Pyr. TS	-2.2652	-2.2696	-3.8855	-2.2696	-3.3253	-3.3266
Bz.	Pyr. TS	-2.0580	-2.0632	-3.6922	-2.0632	-3.1060	-3.1075
Bz.	Ethyne CH-pi	-1.9314	-1.9356	-3.0020	-1.9356	-2.8772	-2.8784
Ethyne	Ethyne TS	-1.0557	-1.0572	-1.3818	-1.0572	-1.3971	-1.3976
Bz.	AcOHOH-pi	-3.2617	-3.2660	-3.9235	-3.2660	-4.7793	-4.7801
Bz.	AcNH$_2$NH-pi	-2.9933	-2.9978	-4.2913	-2.9978	-4.4487	-4.4500
Bz.	H$_2$OOH-pi	-2.3720	-2.3747	-3.0086	-2.3747	-3.5717	-3.5723
Bz.	MeOHOH-pi	-2.8229	-2.8270	-4.0539	-2.8270	-4.4075	-4.4086
Bz.	MeNH$_2$NH-pi	-1.9772	-1.9812	-3.2721	-1.9812	-3.3635	-3.3646
Bz.	Pep. NH-pi	-2.6145	-2.6204	-5.3263	-2.6204	-5.5759	-5.5767
Pyr.	Pyr. CH-N	-3.2275	-3.2304	-3.9264	-3.2304	-3.5616	-3.5627
Ethyne	H$_2$OCH-O	-2.4888	-2.4902	-2.6384	-2.4902	-2.8916	-2.8923
Ethyne	AcOHOH-pi	-3.7617	-3.7637	-4.2136	-3.7637	-5.2408	-5.2413
Pent.	AcOH	-1.6367	-1.6410	-2.4715	-1.6410	-2.9103	-2.9120
Pent.	AcNH$_2$	-2.0814	-2.0856	-3.0256	-2.0856	-3.6023	-3.6039
Bz.	AcOH	-2.2788	-2.2842	-4.4359	-2.2842	-4.0921	-4.0938
Pep.	Ethene	-2.0211	-2.0237	-2.7062	-2.0237	-3.0357	-3.0366
Pyr.	Ethyne	-3.4757	-3.4778	-3.9640	-3.4778	-3.7363	-3.7374

Donor	Acceptor	CCSD		MP2		M06-2X	
		CP	No CP	CP	No CP	CP	No CP
$MeNH_2$	Pyr.	-2.6105	-2.6139	-3.9645	-2.6139	-4.0492	-4.0502

Table E.4: The ratio of dispersion to polarisation, $r_{D/P}$ for the complexes in the S66 database, calculated at the aVTZ level using DF-SAPT2 and ALMO+RPAxd. These are used to classify the systems as one of: electrostatically dominated (E, $r_{D/P} < 2$); dispersive (D, $r_{D/P} \geqslant 8$); or mixed (M, $2 \leqslant r_{D/P} < 8$). Note the SAPT polarisation term is calculated as $E_{Ind.} - E_{CT}$.

Donor	Acceptor	$r_{D/P}$		Classification	
		SAPT	ALMO	SAPT	ALMO
H_2O	H_2O	1.0996	1.1363	E	E
H_2O	MeOH	1.1895	0.7290	E	E
H_2O	$MeNH_2$	0.9225	0.5320	E	E
H_2O	Pep.	1.1286	0.0652	E	E
MeOH	MeOH	1.2861	1.2525	E	E
MeOH	$MeNH_2$	1.1180	1.0529	E	E
MeOH	Pep.	1.2441	0.5944	E	E
MeOH	H_2O	1.1889	1.2113	E	E
$MeNH_2$	MeOH	2.9022	2.9504	M	M
$MeNH_2$	$MeNH_2$	2.7331	2.7504	M	M
$MeNH_2$	Pep.	3.0486	1.6564	M	E
$MeNH_2$	H_2O	1.0336	1.0656	E	E
Pep.	MeOH	1.9626	1.9777	E	E
Pep.	$MeNH_2$	1.5598	1.5885	E	E
Pep.	Pep.	1.7348	1.3800	E	E
Pep.	H_2O	1.6348	1.6477	E	E
Ur.	Ur. BP	0.7991	0.6272	E	E
H_2O	Pyr.	1.0165	0.0375	E	E
MeOH	Pyr.	1.1486	0.3009	E	E
AcOH	AcOH	0.5899	0.4789	E	E
$AcNH_2$	$AcNH_2$	0.7992	0.6795	E	E
AcOH	Ur.	0.6553	0.3490	E	E
$AcNH_2$	Ur.	0.7291	0.3304	E	E
Bz.	Bz. pipi	14.2983	14.0566	D	D
Pyr.	Pyr. pipi	12.6527	14.2959	D	D
Ur.	Ur. pipi	7.8886	6.6198	M	M
Bz.	Pyr. pipi	13.2797	14.1122	D	D
Bz.	Ur. pipi	10.9757	9.3313	D	D
Pyr.	Ur. pipi	10.2527	8.7565	D	D
Bz.	Ethene	12.2746	15.7291	D	D
Ur.	Ethene	11.6850	12.4449	D	D
Ur.	Ethyne	10.4475	9.7863	D	D
Pyr.	Ethene	11.9918	17.8708	D	D
Pent.	Pent.	20.9825	40.7263	D	D
Neopent.	Pent.	18.8088	35.1521	D	D
Neopent.	Neopent.	18.1156	35.6381	D	D
Cyclopent.	Neopent.	18.0419	32.4876	D	D
Cyclopent.	Cyclopent.	18.9463	39.0147	D	D
Bz.	Cyclopent.	14.1383	16.6807	D	D

Donor	Acceptor	$r_{D/P}$		Classification	
		SAPT	ALMO	SAPT	ALMO
Bz.	Neopent.	13.2838	15.9362	D	D
Ur.	Pent.	12.3377	12.4501	D	D
Ur.	Cyclopent.	14.9510	16.5343	D	D
Ur.	Neopent.	14.3679	16.0482	D	D
Ethene	Pent.	14.7407	10.4116	D	D
Ethyne	Pent.	12.8984	10.8721	D	D
Pep.	Pent.	9.3676	8.9429	D	D
Bz.	Bz. TS	10.1721	10.8393	D	D
Pyr.	Pyr. TS	8.2129	8.1578	D	D
Bz.	Pyr. TS	8.0469	7.8367	D	M
Bz.	Ethyne CH-pi	4.5897	4.5801	M	M
Ethyne	Ethyne TS	3.0189	3.4605	M	M
Bz.	AcOHOH-pi	2.8023	2.8262	M	M
Bz.	AcNH$_2$NH-pi	3.1488	2.9239	M	M
Bz.	H$_2$OOH-pi	3.2660	3.5820	M	M
Bz.	MeOHOH-pi	4.4019	4.7884	M	M
Bz.	MeNH$_2$NH-pi	7.9365	9.1545	M	D
Bz.	Pep. NH-pi	5.3659	5.0200	M	M
Pyr.	Pyr. CH-N	3.1774	2.7165	M	M
Ethyne	H$_2$OCH-O	1.6967	1.8127	E	E
Ethyne	AcOHOH-pi	1.6756	0.5967	E	E
Pent.	AcOH	11.2297	13.4529	D	D
Pent.	AcNH$_2$	7.0208	6.7911	M	M
Bz.	AcOH	8.2846	8.2215	D	D
Pep.	Ethene	5.9684	5.4634	M	M
Pyr.	Ethyne	1.5647	1.5099	E	E
MeNH$_2$	Pyr.	5.7620	1.8582	M	E

Table E.5: Energy decompositions and interaction energies in kcal mol^{-1} of the complexes in the S66 database, calculated at the aVTZ level using both DF-SAPT2 and the ALMO+RPAxd EDA. The frozen energy in the latter is physically comparable to the combined electrostatic and exchange terms in the former, while removing the charge transfer (CT) from the induction energy gives a term similar to the polarisation energy.

	Total		ALMO EDA				SAPT2				
ID	ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
1	-5.0361	-4.6821	-1.5351	-1.2915	-0.7418	-1.4676	-7.9154	7.5589	-2.3196	-2.0059	-0.4954
2	-4.7713	-5.3586	-1.0517	-1.6102	-0.9354	-1.1739	-9.0308	9.2612	-2.8807	-2.7082	-0.6039
3	-5.5850	-6.8966	-0.4173	-2.3104	-1.6282	-1.2291	-12.4373	12.9972	-4.3206	-3.1358	-0.9213
4	-5.7951	-7.7095	-1.6139	-2.6926	-1.3130	-0.1756	-12.3662	13.1296	-4.4404	-4.0325	-0.8674
5	-5.5956	-5.5255	-0.7128	-1.7324	-0.9806	-2.1698	-9.1655	9.7890	-3.0496	-3.0994	-0.6397
6	-6.7463	-7.5481	0.4401	-2.6249	-1.7977	-2.7638	-13.0888	14.5788	-4.8249	-4.2132	-1.0563
7	-6.7508	-7.9535	-0.7941	-2.8406	-1.4277	-1.6885	-12.7319	14.2142	-4.7336	-4.7022	-0.9539
8	-5.0974	-4.7684	-1.2347	-1.3973	-0.7728	-1.6926	-7.9017	7.8754	-2.4477	-2.2945	-0.5177
9	-2.7937	-2.9718	0.1061	-0.6384	-0.3777	-1.8837	-4.2753	4.9083	-1.1117	-2.4930	-0.2527
10	-3.3702	-4.2242	0.6951	-0.9171	-0.6258	-2.5224	-6.0478	7.1258	-1.6596	-3.6426	-0.3268
11	-3.5241	-5.4436	0.4460	-1.2472	-0.6571	-2.0658	-6.9475	8.4509	-1.9749	-4.9721	-0.3439
12	-6.8405	-7.2475	-0.1442	-2.4201	-1.6972	-2.5789	-12.5411	13.3997	-4.4576	-3.6485	-0.9278
13	-6.1236	-6.1358	-0.8698	-1.5098	-0.7583	-2.9858	-8.0417	8.2480	-2.4116	-3.9304	-0.4090
14	-6.7841	-7.5878	0.2008	-2.1806	-1.3405	-3.4638	-11.1507	12.0913	-3.7463	-4.7821	-0.6806
15	-8.1126	-8.5618	-0.8804	-2.5640	-1.1300	-3.5382	-11.2792	12.2431	-3.9064	-5.6193	-0.6672
16	-5.3804	-5.0005	-1.6517	-1.1954	-0.5635	-1.9698	-6.7229	6.0586	-1.8534	-2.4828	-0.3347
17	-17.2423	-17.0732	-0.9170	-7.9028	-3.4657	-4.9568	-25.8784	29.7095	-12.4885	-8.4158	-1.9566
18	-4.3483	-6.9293	-0.4998	-2.2918	-1.4708	-0.0859	-11.5574	12.1503	-4.1296	-3.3926	-0.7921

ID	Total		ALMO EDA						SAPT2		
	ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
19	-4.8341	-7.5711	0.1316	-2.5651	-1.6288	-0.7718	-12.2387	13.5291	-4.5920	-4.2694	-0.8748
20	-20.0110	-19.2061	1.6778	-11.0212	-5.3898	-5.2778	-32.4439	40.0586	-18.0508	-8.7701	-3.1829
21	-16.6722	-16.0395	-1.7155	-7.1092	-3.0167	-4.8308	-25.5564	27.9215	-11.1023	-7.3023	-1.9649
22	-18.4768	-19.3542	-0.7214	-9.8192	-4.5098	-3.4264	-30.3636	35.2264	-15.6320	-8.5850	-2.5304
23	-17.4375	-18.9737	-2.5722	-8.4284	-3.6517	-2.7852	-28.6159	31.0430	-13.2660	-8.1348	-2.1086
24	-1.2798	-4.7414	4.6703	-0.3638	-0.4719	-5.1143	-1.6614	6.1911	-0.6818	-8.5892	-0.0811
25	-2.3296	-5.9233	4.1739	-0.3893	-0.5486	-5.5655	-3.1237	7.2524	-0.8194	-9.2326	-0.0897
26	-8.0838	-11.3369	2.1843	-1.2337	-0.8675	-8.1669	-9.1685	12.5624	-1.8463	-12.8845	-0.2130
27	-1.9172	-5.4304	4.4105	-0.3843	-0.5205	-5.4229	-2.5280	6.8819	-0.7679	-9.0165	-0.0889
28	-3.1929	-7.5117	4.6834	-0.6905	-0.7421	-6.4436	-4.9114	10.0535	-1.1782	-11.4755	-0.1327
29	-4.9513	-8.6041	3.4015	-0.7791	-0.7513	-6.8223	-5.9300	9.9965	-1.2591	-11.4115	-0.1461
30	-0.2300	-2.2937	3.3576	-0.1938	-0.3458	-3.0480	-0.9397	4.1940	-0.4865	-5.0615	-0.0741
31	-2.7201	-3.9504	2.0251	-0.3224	-0.4106	-4.0122	-3.1936	5.7151	-0.5920	-5.8799	-0.0888
32	-3.3968	-4.3827	1.0462	-0.3761	-0.3861	-3.6808	-3.7710	5.5387	-0.6253	-5.5251	-0.0965
33	-0.8603	-2.7545	3.1772	-0.1938	-0.3801	-3.4636	-1.6343	4.7593	-0.5389	-5.3405	-0.0936
34	-2.1396	-3.9122	3.8823	-0.1344	-0.4124	-5.4751	-1.9136	6.0517	-0.5082	-7.5420	-0.1488
35	-1.5047	-2.6775	2.7169	-0.1079	-0.3198	-3.7939	-1.4032	4.2863	-0.3941	-5.1664	-0.1194
36	-1.1369	-1.7699	1.9649	-0.0782	-0.2371	-2.7865	-0.8100	2.9111	-0.2869	-3.5840	-0.0891
37	-1.3245	-2.4867	2.7154	-0.1108	-0.3301	-3.5991	-1.3892	4.2800	-0.4027	-4.9747	-0.1270
38	-1.6650	-3.0594	3.0233	-0.1080	-0.3673	-4.2131	-1.5665	4.8082	-0.4387	-5.8624	-0.1293
39	-2.2106	-4.5620	3.5159	-0.2961	-0.4905	-4.9398	-2.4056	6.1207	-0.6589	-7.6181	-0.1201
40	-1.8992	-3.6443	2.6547	-0.2446	-0.4107	-3.8985	-1.8871	4.6988	-0.5445	-5.9114	-0.0995
41	-3.2713	-5.5089	4.3510	-0.5222	-0.5981	-6.5019	-2.9693	7.5272	-0.8899	-9.1768	-0.1461

ID	Total		ALMO EDA						SAPT2		
	ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
42	-2.7112	-4.7203	3.8410	-0.3431	-0.5354	-5.6737	-2.4092	6.4037	-0.6596	-8.0551	-0.1208
43	-2.7284	-4.1698	2.6188	-0.2877	-0.4428	-4.6167	-2.6155	5.4693	-0.5500	-6.4736	-0.0994
44	0.5551	-2.1074	2.1622	-0.1185	-0.2548	-1.2339	-1.1222	3.3946	-0.3476	-4.0322	-0.0741
45	1.1484	-2.0908	1.8011	-0.1498	-0.2224	-0.2805	-1.1711	3.0124	-0.3423	-3.5898	-0.0640
46	-2.6231	-4.5582	3.7943	-0.5947	-0.5048	-5.3180	-2.7188	6.8417	-0.9889	-7.6923	-0.1677
47	-1.7882	-3.7484	2.0269	-0.2902	-0.3793	-3.1456	-1.9490	4.2153	-0.6181	-5.3966	-0.0876
48	-2.7622	-4.3285	1.6316	-0.4329	-0.4297	-3.5312	-2.8881	4.8880	-0.7879	-5.5405	-0.1133
49	-2.6555	-4.1932	1.6937	-0.4421	-0.4429	-3.4642	-2.4112	4.4349	-0.7664	-5.4505	-0.0891
50	-2.8949	-3.4921	0.6679	-0.5571	-0.4542	-2.5515	-2.1929	3.2779	-0.8939	-3.6833	-0.0914
51	-1.4429	-1.6306	0.0381	-0.2609	-0.3174	-0.9027	-1.7503	2.0883	-0.5435	-1.4251	-0.0714
52	-4.4890	-5.1777	1.3998	-1.3036	-0.9010	-3.6843	-4.2155	6.2521	-2.0661	-5.1481	-0.2290
53	-3.8868	-4.6924	0.8055	-1.0331	-0.6384	-3.0208	-4.9277	6.4869	-1.6779	-4.5737	-0.2253
54	-3.2172	-3.4249	0.3261	-0.6642	-0.4998	-2.3793	-2.9519	3.7988	-1.1065	-3.1654	-0.1373
55	-3.2092	-4.5455	1.4316	-0.6945	-0.6210	-3.3253	-3.4178	5.3113	-1.3183	-5.1208	-0.1550
56	-2.5877	-3.7727	1.8274	-0.3901	-0.4535	-3.5716	-2.5308	4.6341	-0.7508	-5.1252	-0.1050
57	-3.8987	-6.0766	1.9511	-0.8655	-0.6395	-4.3448	-4.0224	6.4621	-1.4659	-7.0504	-0.1520
58	-3.7782	-4.5213	0.2259	-0.9469	-0.4852	-2.5721	-4.9264	5.7914	-1.4679	-3.9184	-0.2347
59	-3.4262	-2.8738	-1.2324	-0.6553	-0.3506	-1.1879	-3.8692	3.4039	-1.0115	-1.3970	-0.1882
60	-3.3439	-4.9258	0.1730	-1.4503	-1.2013	-0.8653	-6.9831	8.4229	-2.6771	-3.6885	-0.4758
61	-2.2683	-3.0092	2.6884	-0.3140	-0.4191	-4.2236	-1.8115	4.7153	-0.5856	-5.3275	-0.1112
62	-2.6572	-3.7014	3.2257	-0.6954	-0.4648	-4.7228	-2.4540	5.9867	-1.0554	-6.1787	-0.1753
63	-3.3507	-4.5609	2.1609	-0.5404	-0.5279	-4.4433	-3.1699	5.8425	-0.8703	-6.3632	-0.1022
64	-2.4967	-3.1353	1.4061	-0.5479	-0.3618	-2.9931	-2.7361	4.6239	-0.8312	-4.1918	-0.1289

	Total		ALMO EDA				SAPT2				
ID	ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
65	-4.3664	-4.3230	-0.6989	-1.1895	-0.6818	-1.7961	-6.0210	6.1151	-1.9065	-2.5107	-0.3019
66	-0.6426	-4.3648	1.5830	-0.5912	-0.5358	-1.0986	-4.4462	6.4650	-1.1249	-5.2586	-0.2123

X40 BENCHMARKS

The following IDs are used to shorthand the X40 donors and acceptors:

ID	Donor	Acceptor
1	C_6H_5Br	Acetone
2	C_6H_5Cl	CH_3SH
3	$C_6H_3F_3$	Trimethylamine
4	C_6F_6	Benzene
5	C_6H_5I	Formaldehyde
6	CH_3Br	CH_3Cl
7	CF_3Br	CH_3F
8	CF_3I	Methane
9	CH_3I	Methanol
10	CH_3Cl	Methylamine
11	CF_3OH	Br_2
12	CF_3Cl	Cl_2
13	CH_3F	F_2
14	HBr	I_2
15	HCl	Water
16	HF	—
17	HI	—
18	Methane	—
19	Methanol	—
20	$CHCl_3$	—
21	CHF_3	—
22	CCl_3OH	—

Table E.6: Uncorrelated interaction energies in kcal mol^{-1} for the molecules in the X40 database using aVDZ basis sets. Counterpoise-corrected Hartree-Fock (HF) is compared with the ALMO + infinite-order charge transfer (CT) energies, along with the three different orders of charge transfer: CT1 (2nd order), CT2 (2nd + 4th order), and CT3 (infinite order).

Donor	Acceptor	HF	ALMO+CT3	CT1	CT2	CT3
1	1	0.2437	0.0320	-0.6210	-0.6200	-0.4620
1	2	0.5782	0.4170	-0.5750	-0.5750	-0.4900
1	3	0.6809	0.4170	-1.3460	-1.3450	-1.0810
2	1	0.5993	0.4040	-0.3640	-0.3640	-0.2540
2	3	0.8524	0.5640	-0.7390	-0.7390	-0.5270
3	4	4.4722	2.8150	-1.9150	-1.9150	-1.3010
4	4	-0.6358	-0.9070	-0.9100	-0.9090	-0.6770
5	1	0.4848	0.2870	-1.0240	-1.0230	-0.8710
5	2	-0.0227	-0.3180	-2.1280	-2.1260	-1.6900

Donor	Acceptor	HF	ALMO+CT3	CT1	CT2	CT3
5	3	0.2869	0.1910	-0.4010	-0.4010	-0.3300
6	5	0.6242	0.1840	-0.8440	-0.8430	-0.6510
7	4	0.7078	0.1580	-1.3130	-1.3120	-1.0010
8	4	1.9577	1.6820	-0.5600	-0.5600	-0.4440
6	4	2.1920	1.7910	-0.8670	-0.8660	-0.6960
9	4	-0.0912	-0.2660	-0.2190	-0.2190	-0.1380
10	6	0.3546	0.2200	-0.2640	-0.2640	-0.1970
10	5	0.4297	0.2160	-0.2230	-0.2230	-0.1560
10	8	-1.4697	-1.5820	-0.6910	-0.6910	-0.5690
7	5	-0.8067	-0.9660	-0.4800	-0.4800	-0.3670
12	5	-2.4054	-2.5740	-1.0470	-1.0470	-0.8250
8	5	-0.9443	-1.2150	-0.3320	-0.3320	-0.1700
13	7	0.2828	0.0600	-0.2180	-0.2180	-0.1420
13	8	-2.8177	-2.5060	-1.7710	-1.7700	-1.6420
14	9	-4.0203	-3.7720	-1.8610	-1.8590	-1.6940
15	10	-11.8829	-11.3750	-4.0140	-4.0080	-3.4560
15	9	-8.1716	-8.0080	-2.0110	-2.0090	-1.7800
16	10	-1.4034	-1.2110	-1.3500	-1.3490	-1.2460
16	9	-0.1555	-0.3070	-0.6580	-0.6580	-0.5240
17	9	0.4862	0.3100	-0.2780	-0.2780	-0.2260
9	5	0.3910	0.1770	-0.2660	-0.2660	-0.1960
18	11	0.2081	0.0860	-0.1860	-0.1860	-0.0990
18	12	0.4853	0.2780	-0.2910	-0.2910	-0.2350
18	13	-1.3830	-1.6920	-0.7410	-0.7410	-0.5680
18	14	-2.3436	-2.5680	-0.7010	-0.7010	-0.5570
19	6	0.6513	0.3890	-0.2760	-0.2760	-0.1880
19	7	-7.1755	-7.1670	-2.5940	-2.5920	-2.2960
20	8	0.2157	-0.0340	-0.2360	-0.2360	-0.1570
22	15	-7.6370	-7.6610	-2.1030	-2.1020	-1.8740

Table E.7: Correlated aVDZ interaction energies in kcal mol^{-1} for the X40 database, calculated using CCSD, DF-MP2, Mo6-2X, DF-SAPT2, with all but the latter counterpoise corrected. These are compared with ALMO+RPAd results, with all three RPA variants. The dispersion terms are also given for reference.

Donor	Acceptor	CCSD	MP2	Mo6-2X	SAPT2	ALMO+RPAd dRPA	SOSEX	RPAx	Dispersion dRPA	SOSEX	RPAx
1	1	-1.6393	-2.1614	-2.2392	-2.3416	-1.5190	-1.5590	-1.0550	-1.5510	-1.5910	-1.0870
1	2	-1.4304	-2.1142	-2.2166	-2.1315	-1.1910	-1.2120	-0.6910	-1.6080	-1.6290	-1.1080
1	3	-2.5006	-3.6274	-3.5803	-3.8121	-1.4030	-1.7630	-1.1970	-1.8200	-2.1800	-1.6140
2	1	-0.8833	-1.2605	-1.1011	-1.4622	-0.7810	-0.7980	-0.4090	-1.1850	-1.2020	-0.8130
2	3	-1.2807	-1.9508	-1.7173	-2.2062	-0.9590	-1.1140	-0.6190	-1.5230	-1.6780	-1.1830
3	4	-2.2277	-5.9017	-4.4702	-6.1122	-2.5360	-2.6700	-0.9090	-5.3510	-5.4850	-3.7240
4	4	-2.4889	-3.0816	-3.4035	-3.3386	-2.5290	-2.6040	-2.0830	-1.6220	-1.6970	-1.1760
5	1	-1.8749	-2.7934	-3.1791	-2.8610	-1.4880	-1.5540	-0.9860	-1.7750	-1.8410	-1.2730
5	2	-3.9743	-5.5953	-5.6136	-5.9546	-2.0360	-2.7290	-2.2450	-1.7180	-2.4110	-1.9270
5	3	-1.0611	-1.4235	-1.6592	-1.4631	-0.8450	-0.8610	-0.5290	-1.0360	-1.0520	-0.7200
6	5	-1.9792	-3.3193	-3.2293	-3.5095	-1.6670	-1.7230	-1.1260	-1.8510	-1.9070	-1.3100
7	4	-2.3331	-4.0453	-4.2403	-4.4443	-2.3260	-2.3450	-1.5240	-2.4840	-2.5030	-1.6820
8	4	-0.8141	-2.1337	-1.8221	-2.1937	-0.2560	-0.3020	0.3180	-1.9380	-1.9840	-1.3640
6	4	-1.0938	-2.7908	-2.6845	-2.9078	-0.4390	-0.5220	0.1870	-2.2300	-2.3130	-1.6040
9	4	-0.9090	-1.1327	-1.0681	-1.1637	-0.6250	-0.6020	-0.4770	-0.3590	-0.3360	-0.2110
10	6	-0.6471	-0.8716	-0.9195	-0.9371	-0.4990	-0.5080	-0.2760	-0.7190	-0.7280	-0.4960
10	5	-0.6016	-0.7681	-0.9444	-0.8155	-0.5440	-0.5480	-0.3010	-0.7600	-0.7640	-0.5170
10	8	-2.3914	-2.6745	-3.1257	-2.8224	-2.7600	-2.7850	-2.4070	-1.1780	-1.2030	-0.8250
7	5	-1.6416	-1.8559	-2.0980	-2.0096	-1.9800	-1.9890	-1.6600	-1.0140	-1.0230	-0.6940

Donor	Acceptor	CCSD	MP2	Mo6-2X	SAPT2	ALMO+RPAd			Dispersion		
						dRPA	SOSEX	RPAx	dRPA	SOSEX	RPAx
12	5	-3.1906	-3.5359	-4.2499	-3.8058	-3.8390	-3.8920	-3.4900	-1.2650	-1.3180	-0.9160
8	5	-1.3411	-1.3468	-1.4405	-1.3620	-1.7230	-1.7300	-1.5580	-0.5080	-0.5150	-0.3430
13	7	-0.4655	-0.5282	-0.6637	-0.5638	-0.4980	-0.5000	-0.3160	-0.5580	-0.5600	-0.3760
13	8	-4.0336	-5.0508	-5.0754	-4.7702	-3.4860	-3.7590	-3.4650	-0.9800	-1.2530	-0.9590
14	9	-5.0620	-5.9740	-6.0207	-5.7324	-4.1990	-4.2280	-4.0880	-0.4270	-0.4560	-0.3160
15	10	-12.6144	-13.2433	-14.6183	-13.1249	-11.8710	-11.8990	-11.7270	-0.4960	-0.5240	-0.3520
15	9	-8.3012	-8.5849	-9.5880	-8.2645	-8.3310	-8.3520	-8.2420	-0.3230	-0.3440	-0.2340
16	10	-2.7654	-3.7440	-3.8985	-3.5364	-2.3060	-2.5160	-2.1830	-1.0950	-1.3050	-0.9720
16	9	-1.5725	-2.0480	-2.4705	-2.1039	-1.4670	-1.4970	-1.1270	-1.1600	-1.1900	-0.8200
17	9	-0.6924	-1.0716	-1.3044	-1.0929	-0.0050	0.0490	0.1660	-0.3150	-0.2610	-0.1440
9	5	-0.5869	-0.8837	-0.9729	-0.9050	0.0690	0.0880	0.1280	-0.1080	-0.0890	-0.0490
18	11	-0.2625	-0.3810	-0.4099	-0.3594	-0.2230	-0.2100	-0.1060	-0.3090	-0.2960	-0.1920
18	12	-0.7177	-1.1049	-1.4706	-1.1562	0.0280	0.0580	0.1510	-0.2500	-0.2200	-0.1270
18	13	-2.7889	-3.1833	-3.6242	-3.1068	-2.0550	-2.0230	-1.8940	-0.3630	-0.3310	-0.2020
18	14	-3.2099	-3.3568	-3.6657	-3.2397	-3.3430	-3.4080	-3.1510	-0.7750	-0.8400	-0.5830
19	6	-0.6485	-0.9471	-1.1214	-1.0114	-0.6560	-0.6610	-0.3240	-1.0450	-1.0500	-0.7130
19	7	-8.6387	-9.4581	-10.7642	-9.4063	-8.4020	-8.7100	-8.3270	-1.2350	-1.5430	-1.1600
20	8	-0.3997	-0.4447	-0.5369	-0.4706	-0.5350	-0.5350	-0.3660	-0.5010	-0.5010	-0.3320
22	15	-8.3052	-8.6838	-9.6393	-8.6425	-8.5980	-8.8100	-8.5130	-0.9370	-1.1490	-0.8520

Table E.8: Energy decompositions and interaction energies in kcal mol^{-1} of the complexes in the X40 database, calculated at the aVDZ level using both DF-SAPT2 and the ALMO+RPAxd EDA. The frozen energy in the latter is physically comparable to the combined electrostatic and exchange terms in the former, while removing the charge transfer (CT) from the induction energy gives a term similar to the polarisation energy.

		Total		ALMO EDA				SAPT2				
Don.	Acc.	ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
1	1	-1.5590	-2.3416	1.2470	-0.5940	-0.6210	-1.5910	-2.6251	4.3569	-1.0452	-3.0366	-0.1712
1	2	-1.2120	-2.1315	1.1800	-0.1880	-0.5750	-1.6290	-2.3748	3.9284	-0.6949	-2.9172	-0.1701
1	3	-1.7630	-3.8121	2.7500	-0.9870	-1.3460	-2.1800	-6.4042	10.0177	-2.2055	-4.9702	0.1912
2	1	-0.7980	-1.4622	1.1000	-0.3320	-0.3640	-1.2020	-1.0727	2.4158	-0.5783	-2.1888	-0.2097
2	3	-1.1140	-2.2062	1.7710	-0.4680	-0.7390	-1.6780	-3.0337	5.3056	-1.0272	-3.2168	-0.7903
3	4	-2.6700	-6.1122	5.1760	-0.4460	-1.9150	-5.4850	-4.5000	10.4494	-0.9764	-9.5974	-0.2615
4	4	-5.1740	-7.9261	4.2360	-0.6980	-2.4210	-6.2910	-6.1337	11.3736	-1.1900	-10.7527	-0.2743
5	1	-2.6040	-3.3386	0.9760	-0.9730	-0.9100	-1.6970	-4.2659	6.0526	-1.6491	-3.5088	-0.1442
5	2	-1.5540	-2.8610	1.7380	-0.4270	-1.0240	-1.8410	-4.1623	6.5297	-1.3874	-3.7593	-0.2075
5	3	-2.7290	-5.9546	3.9670	-2.1570	-2.1280	-2.4110	-11.2620	16.5346	-4.0763	-6.8148	3.2444
6	5	-0.8610	-1.4631	0.8540	-0.2620	-0.4010	-1.0520	-1.8368	2.9481	-0.5619	-2.0317	-0.1486
7	4	-1.7230	-3.5095	1.4180	-0.3900	-0.8440	-1.9070	-2.4687	4.4885	-0.9490	-4.1324	-0.1293
8	4	-2.3450	-4.4443	2.2500	-0.7790	-1.3130	-2.5030	-3.9548	7.1371	-1.7359	-5.3718	-0.0688
6	4	-0.3020	-2.1937	2.4250	-0.1830	-0.5600	-1.9840	-1.1773	3.7290	-0.5376	-3.6356	-0.1363
9	4	-0.5220	-2.9078	2.9750	-0.3170	-0.8670	-2.3130	-2.4171	5.7846	-0.8862	-4.6952	-0.1339
10	6	-0.6020	-1.1637	0.0200	-0.0670	-0.2190	-0.3360	-0.7304	0.9961	-0.1588	-1.2674	-0.0417
10	5	-0.5080	-0.9371	0.6280	-0.1440	-0.2640	-0.7280	-0.8214	1.5986	-0.3015	-1.4484	-0.0884
10	8	-0.5480	-0.8155	0.4910	-0.0520	-0.2230	-0.7640	-0.3660	0.9756	-0.0986	-1.3517	-0.0180

Don.	Acc.	Total		ALMO EDA						SAPT2		
		ALMO	SAPT2	Frz.	Pol.	CT	Disp.	Elst.	Exch.	Ind.	Disp.	CT
7	5	-2.7850	-2.8224	-0.3630	-0.5280	-0.6910	-1.2030	-3.7512	4.2773	-1.1292	-2.3240	-0.2330
12	5	-1.9890	-2.0096	-0.1900	-0.2960	-0.4800	-1.0230	-2.3984	2.8217	-0.6629	-1.8209	-0.2257
8	5	-3.8920	-3.8058	-0.6160	-0.9110	-1.0470	-1.3180	-5.4258	6.0702	-1.8016	-2.7816	-0.2625
13	7	-1.7300	-1.3620	-0.7580	-0.1250	-0.3320	-0.5150	-1.2351	0.8353	-0.1789	-0.9098	-0.0349
13	8	-0.5000	-0.5638	0.3190	-0.0410	-0.2180	-0.5600	-0.2306	0.6400	-0.0713	-0.9714	-0.0122
14	9	-3.7590	-4.7702	1.3970	-2.1320	-1.7710	-1.2530	-9.0699	12.2454	-4.3416	-3.7060	-1.5643
15	10	-6.1780	-11.5263	8.7690	-7.6580	-6.4300	-0.8590	-24.1283	35.7660	-16.2068	-7.2446	-7.2459
15	9	-4.2280	-5.7324	0.3570	-2.2680	-1.8610	-0.4560	-10.0810	12.3714	-4.6491	-3.5364	-1.6510
16	10	-11.8990	-13.1249	-2.2710	-5.0900	-4.0140	-0.5240	-21.8978	23.3982	-10.5745	-4.7152	-4.5882
16	9	-8.3520	-8.2645	-3.1560	-2.8410	-2.0110	-0.3440	-12.8730	12.8587	-5.6633	-3.1488	-2.1859
17	9	-2.5160	-3.5364	1.8090	-1.6700	-1.3500	-1.3050	-6.9094	10.0691	-3.2272	-3.4528	-1.1076
9	5	-1.4970	-2.1039	0.8110	-0.4600	-0.6580	-1.1900	-3.1317	4.3946	-0.9278	-2.4404	-0.1969
18	11	0.0490	-1.0929	0.7330	-0.1450	-0.2780	-0.2610	-0.7943	1.7569	-0.3582	-1.6410	-0.5128
18	12	0.0880	-0.9050	0.5260	-0.0830	-0.2660	-0.0890	-0.5890	1.2948	-0.2320	-1.3303	-0.0528
18	13	-0.2100	-0.3594	0.2790	-0.0070	-0.1860	-0.2960	-0.2239	0.5399	-0.0754	-0.5964	-0.0036
18	14	0.0580	-1.1562	0.7610	-0.1920	-0.2910	-0.2200	-0.8873	1.9103	-0.4022	-1.7289	-0.0568
19	6	-2.0230	-3.1068	-0.4220	-0.5290	-0.7410	-0.3310	-3.9880	4.6076	-1.2127	-2.6791	-0.4114
19	7	-3.4080	-3.2397	-1.3110	-0.5560	-0.7010	-0.8400	-4.6431	4.4347	-1.1337	-2.1277	-0.4197
20	8	-0.6610	-1.0114	0.6850	-0.0200	-0.2760	-1.0500	-0.4022	1.2230	-0.0658	-1.7338	-0.0195
22	15	-8.7100	-9.4063	-1.1130	-3.4600	-2.5940	-1.5430	-16.0525	18.2789	-7.0470	-5.0992	-2.8625
21	8	-0.5350	-0.4706	0.2120	-0.0100	-0.2360	-0.5010	-0.1244	0.4257	-0.0234	-0.8211	-0.0030
11	15	-8.8100	-8.6425	-2.7420	-2.8160	-2.1030	-1.1490	-13.9141	14.0020	-5.6294	-3.6805	-2.2281

F | ADDITIONAL DERIVATIONS FOR GRADIENTS

To determine the analytical derivatives in Chapter 6, a number of intermediate derivatives of the transformed quantities in the orthogonalised basis are needed in terms of the same quantities in the ALMO basis. We start by considering those intermediates in equations 6.23, which are for the transformed orbital coefficients and amplitudes in terms of the ALMO orbital coefficients. Common to all of these is the need for derivatives of the square-root inverse metrics, $\boldsymbol{\sigma}^{-1/2}$ and $\boldsymbol{\pi}^{-1/2}$. We will use a common notation as follows:

$$
\begin{aligned}
\gamma^o_{ikj\mu} &= \frac{\partial}{\partial T^\mu_j} \left[\boldsymbol{\sigma}^{-1/2} \right]_{ik} \\
\gamma^v_{acb\mu} &= \frac{\partial}{\partial V^\mu_b} \left[\boldsymbol{\pi}^{-1/2} \right]_{ac}
\end{aligned}
\tag{F.1}
$$

These derivatives are complicated due to the square root, and require the solution of sets of linear equations as follows. We note the result from functional analysis that

$$
\delta s^{-1} = -s^{-1} (\delta s) \, s^{-1}
$$

and thus write that, for example for the occupied inverse metric (the result for the virtual metric is equivalent):

$$
\frac{\partial \sigma^{il}}{\partial T_{jv}} = -\sigma^{ij} \left[\mathbf{ST}\boldsymbol{\sigma}^{-1} \right]_{vl} - \left[\boldsymbol{\sigma}^{-1}\mathbf{T}^\dagger\mathbf{S} \right]_{iv} \sigma^{jl}
$$

The inverse can of course be expanded in terms of its square root as

$$
\sigma^{il} = \left[\boldsymbol{\sigma}^{-1/2} \right]_{ik} \left[\boldsymbol{\sigma}^{-1/2} \right]_{kl}
$$

The linear equations for the occupied γ are thus

$$
\gamma^o_{ikjv} \left[\boldsymbol{\sigma}^{-1/2} \right]_{kl} + \left[\boldsymbol{\sigma}^{-1/2} \right]_{ik} \gamma^o_{kljv} = -\sigma^{ij} \left[\mathbf{ST}\boldsymbol{\sigma}^{-1} \right]_{vl} - \left[\boldsymbol{\sigma}^{-1}\mathbf{T}^\dagger\mathbf{S} \right]_{iv} \sigma^{jl}
\tag{F.2}
$$

and equivalently for γ^v, but with the virtual indices and metric.

The transformed coefficients can be written as follows:

$$
\tilde{\mathbf{T}} = \boldsymbol{\sigma}^{-1/2}\mathbf{T} \text{ and } \tilde{\mathbf{V}} = \boldsymbol{\pi}^{-1/2}\mathbf{VQ}
$$

For the occupied transformation, we thus have the simple result that

$$
\mathcal{T}_{\mu i v j} = \delta_{\mu v} \left[\boldsymbol{\sigma}^{-1/2} \right]_{ij} + T^\mu_k \gamma^o_{ikjv}
\tag{F.3}
$$

For the virtual transformation, however, we have the added complication of the projector, $\mathbf{Q}$, which itself depends on the occupied coefficents. We thus split it into two parts. The first, in terms of the virtual ALMO coefficients, is equivalent to equation F.3 above:

$$\mathcal{V}_{\mu a v b} = \left[\boldsymbol{\pi}^{-1/2}\right]_{ab} Q_{\mu v} + [\mathbf{VQ}]_{c\mu}\, \gamma^{v}_{acbv} \tag{F.4}$$

In terms of the occupied coefficients, however, we have the problem that $\boldsymbol{\pi}$ also depends on $\mathbf{Q}$, as

$$\boldsymbol{\pi} = \mathbf{V}^{\dagger}\mathbf{Q}^{\dagger}\mathbf{SQV} = \mathbf{V}^{\dagger}\mathbf{SV} - \mathbf{V}^{\dagger}\mathbf{PSSV} - \mathbf{V}^{\dagger}\mathbf{SSPV} + \mathbf{V}^{\dagger}\mathbf{PSSSPV} \tag{F.5}$$

Therefore we need to form a further γ tensor, γ^{vo}, where the right hand side of equation F.2 changes to now give

$$\frac{\partial \pi^{ac}}{\partial T_{jv}} = \pi^{ad} \left\{ V_{d\tau}\left[\mathbf{SSQV}\right]_{\rho e} + \left[\mathbf{V}^{\dagger}\mathbf{Q}^{\dagger}\mathbf{SS}\right]_{d\tau} V_{e\rho} \right\} \frac{\partial P_{\tau\rho}}{\partial T_{jv}} \pi^{ec} \tag{F.6}$$

where the density derivative is known from equation 2.47. This means we have a third set of linear equations:

$$\gamma^{vo}_{acjv}\left[\boldsymbol{\pi}^{-1/2}\right]_{cb} + \left[\boldsymbol{\pi}^{-1/2}\right]_{ac}\gamma^{o}_{cbjv} = \frac{\partial \pi^{ac}}{\partial T_{jv}} \tag{F.7}$$

The occupied-virtual transformation tensor is then

$$\mathcal{V}_{\mu a v j} = -\left[\boldsymbol{\pi}^{-1/2}\right]_{ab}\left[\mathbf{V}^{\dagger}\mathbf{S}\right]_{b\tau}\frac{\partial P_{\tau\mu}}{\partial T_{vj}} + [\mathbf{QV}]_{\mu b}\,\gamma^{vo}_{abjv} \tag{F.8}$$

The final intermediate from equations 6.23 is that of the transformed residual amplitudes. These are expanded as

$$\tilde{t}^{ab}_{ij} = \left[\boldsymbol{\sigma}^{-1/2}\right]_{ik}\left[\boldsymbol{\pi}^{-1/2}\right]_{ac}t^{cd}_{kl}\left[\boldsymbol{\pi}^{-1/2}\right]_{db}\left[\boldsymbol{\sigma}^{-1/2}\right]_{lj}$$

Thus the derivatives are simply expanded in terms of the same γ tensors above, as an exercise in the product rule of differentiation:

$$\begin{aligned}
\mathcal{Z}_{ijabvk} =&\gamma^{o}_{imkv}\left[\boldsymbol{\sigma}^{1/2}\right]_{mn}\tilde{t}^{ab}_{nj} + \tilde{t}^{ab}_{im}\left[\boldsymbol{\sigma}^{1/2}\right]_{mn}\gamma^{o}_{njkv} \\
&+ \gamma^{vo}_{adkv}\left[\boldsymbol{\pi}^{1/2}\right]_{de}\tilde{t}^{eb}_{ij} + \tilde{t}^{ad}_{ij}\left[\boldsymbol{\pi}^{1/2}\right]_{de}\gamma^{vo}_{ebkv} \\
\mathcal{Z}_{ijabvc} =&\gamma^{v}_{adcv}\left[\boldsymbol{\pi}^{1/2}\right]_{de}\tilde{t}^{ec}_{ij} + \tilde{t}^{ad}_{ij}\left[\boldsymbol{\pi}^{1/2}\right]_{de}\gamma^{v}_{ebcv}
\end{aligned} \tag{F.9}$$

The second set of intermediates we need are those of equations 6.32, concerning the derivatives of the transformed orbital coefficients and residual amplitudes with respect to the AO overlap matrix. The former are reasonably simple, although again the projector in the virtual coefficients adds complexity. We once more need derivatives of the

square-root inverse metrics, which again leads to sets of linear equations. We have very simply that $\boldsymbol{\sigma} = \mathbf{T}^\dagger \mathbf{S} \mathbf{T}$, such that

$$\frac{\partial \sigma_{ij}}{\partial S_{\mu\nu}} = T_i^\mu T_j^\nu$$

while for the virtuals, we need to work from equation F.5, also containing the density, which itself depends on the AO overlap. The result is messy:

$$\frac{\partial \pi_{ab}}{\partial S_{\mu\nu}} = V_a^\mu V_b^\nu - \left[\mathbf{V}^\dagger\mathbf{P}\right]_{a\mu} [\mathbf{S}\mathbf{V}]_{\nu b} - \left[\mathbf{V}^\dagger\mathbf{P}\mathbf{S}\right]_{a\mu} V_b^\nu - V_a^\tau \frac{\partial P_{\tau\rho}}{\partial S_{\mu\nu}} [\mathbf{S}\mathbf{S}\mathbf{V}]_{\rho b}$$
$$- V_a^\mu [\mathbf{S}\mathbf{P}\mathbf{V}]_{\nu b} - \left[\mathbf{V}^\dagger\mathbf{S}\right]_{a\mu} [\mathbf{P}\mathbf{V}]_{\nu b} - \left[\mathbf{V}^\dagger\mathbf{S}\mathbf{S}\right]_{a\tau} \frac{\partial P_{\tau\rho}}{\partial S_{\mu\nu}} V_b^\rho$$
$$+ \left[\mathbf{V}^\dagger\mathbf{P}\right]_{a\mu} [\mathbf{S}\mathbf{S}\mathbf{P}\mathbf{V}]_{\nu b} + \left[\mathbf{V}^\dagger\mathbf{P}\mathbf{S}\right]_{a\mu} [\mathbf{S}\mathbf{P}\mathbf{V}]_{\nu b} + \left[\mathbf{V}^\dagger\mathbf{P}\mathbf{S}\mathbf{S}\right]_{a\mu} [\mathbf{P}\mathbf{V}]_{\nu b}$$
$$+ V_a^\tau \frac{\partial P_{\tau\rho}}{\partial S_{\mu\nu}} [\mathbf{S}\mathbf{S}\mathbf{S}\mathbf{P}\mathbf{V}]_{\rho b} + \left[\mathbf{V}^\dagger\mathbf{P}\mathbf{S}\mathbf{S}\mathbf{S}\right]_{a\tau} \frac{\partial P_{\tau\rho}}{\partial S_{\mu\nu}} V_b^\rho$$

From these, however, we can deduce the derivatives of the square-root inverse metrics in exactly the same manner as equation F.2. With these determined, we directly get the derivatives of the transformed coefficients by multiplying the aforementioned derivatives by the untransformed coefficients. Similarly, the derivatives of the transformed residual amplitude are then just a series of linear combinations of the square-root inverse metric derivatives with the untransformed amplitudes. The desired results then follow immediately.

BIBLIOGRAPHY

(1) H. Stoll, G. Wagenblast and H. Preuss, "On the Use of Local Basis-Sets for Localized Molecular-Orbitals", *Theoretica Chimica Acta*, 1980, **57**, 169–178.

(2) P. R. Horn, E. J. Sundstrom, T. A. Baker and M. Head-Gordon, "Unrestricted absolutely localized molecular orbitals for energy decomposition analysis: Theory and applications to intermolecular interactions involving radicals", *The Journal of Chemical Physics*, 2013, **138**, 134119.

(3) R. A. DiStasio, O. A. von Lilienfeld and A. Tkatchenko, "Collective many-body van der Waals interactions in molecular systems", *Proceedings of the National Academy of Sciences*, 2012, **109**, 14791–14795.

(4) R. J. Full, K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch, W. P. Chan, T. W. Kenny and R. Fearing, "Adhesive force of a single gecko foot-hair", *Nature*, 2000, **405**, 681–685.

(5) K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili and R. J. Full, "Evidence for van der Waals adhesion in gecko setae", *Proceedings of the National Academy of Sciences*, 2002, **99**, 12252–12256.

(6) B. Rozitis, E. MacLennan and J. P. Emery, "Cohesive forces prevent the rotational breakup of rubble-pile asteroid (29075) 1950 DA", *Nature*, 2014, **512**, 174–176.

(7) G. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, New York, 1999.

(8) M. K. Gilson and H.-X. Zhou, "Calculation of Protein-Ligand Binding Affinities", *Annual Review of Biophysics and Biomolecular Structure*, 2007, **36**, 21–42.

(9) M. Muthukumar, C. Ober and E. Thomas, "Competing Interactions and Levels of Ordering in Self-Organizing Polymeric Materials", *Science*, 1997, **277**, 1225–1232.

(10) A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, 2nd ed., 2013.

(11) Y. Mao, Q. Ge, P. R. Horn and M. Head-Gordon, "On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes", *Journal of Chemical Theory and Computation*, 2018, **14**, 2401–2417.

(12) R. A. Shaw, J. G. Hill and A. C. Legon, "Halogen Bonding with Phosphine: Evidence for Mulliken Inner Complexes and the Importance of Relaxation Energy", *The Journal of Physical Chemistry A*, 2016, **120**, 8461–8468.

(13) G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, "The Halogen Bond", *Chemical Reviews*, 2016, **116**, 2478–2601.

(14) J. G. Hill and A. C. Legon, "On the directionality and non-linearity of halogen and hydrogen bonds", *Physical Chemistry Chemical Physics*, 2014, **17**, 858–867.

(15) A. J. Stone, "Are halogen bonded structures electrostatically driven?", *Journal of the American Chemical Society*, 2013, **135**, 7005–9.

(16) A. J. Stone and A. J. Misquitta, "Charge-transfer in Symmetry-Adapted Perturbation Theory", *Chemical Physics Letters*, 2009, **473**, 201–205.

(17) L. Yang, C. Adam, G. S. Nichol and S. L. Cockroft, "How much do van der Waals dispersion forces contribute to molecular recognition in solution?", *Nature Chemistry*, 2013, **5**, 1006–1010.

(18) E. G. Hohenstein and C. D. Sherrill, "Wavefunction methods for non-covalent interactions", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 304–326.

(19) S. Grimme, "Density functional theory with London dispersion corrections", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 211–228.

(20) L. A. Burns, A. V. Mayagoitia, B. G. Sumpter and C. D. Sherrill, "Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals", *The Journal of Chemical Physics*, 2011, **134**, 084107.

(21) K. E. Riley, M. Pitoňák, P. Jurečka and P. Hobza, "Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories", *Chemical Reviews*, 2010, **110**, 5023–5063.

(22) K. E. Riley and P. Hobza, "Noncovalent interactions in biochemistry", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 3–17.

(23) S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, "Dispersion-Corrected Mean-Field Electronic Structure Methods", *Chemical Reviews*, 2016, **116**, 5105–5154.

(24) P. Jurečka, J. Šponer, J. Černý and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs", *Physical Chemistry Chemical Physics*, 2006, **8**, 1985–1993.

(25) B. W. Hopkins and G. S. Tschumper, "Ab Initio Studies of $\pi \cdots \pi$ Interactions: The Effects of Quadruple Excitations", *The Journal of Physical Chemistry A*, 2004, **108**, 2941–2948.

(26) M. Pitoňák, P. Neogrády, J. Řezáč, P. Jurečka, M. Urban and P. Hobza, "Benzene Dimer: High-Level Wave Function and Density Functional Theory Calculations", *Journal of Chemical Theory and Computation*, 2008, **4**, 1829–1834.

(27)   L. A. Burns, M. S. Marshall and C. D. Sherrill, "Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches", *The Journal of Chemical Physics*, 2014, **141**, 234111.

(28)   W. A. de Jong, E. Bylaska, N. Govind, C. L. Janssen, K. Kowalski, T. Müller, I. M. B. Nielsen, H. J. J. van Dam, V. Veryazov and R. Lindh, "Utilizing high performance computing for chemistry: parallel computational chemistry", *Physical Chemistry Chemical Physics*, 2010, **12**, 6896.

(29)   M. O. Sinnokrot and C. D. Sherrill, "Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations", *The Journal of Physical Chemistry A*, 2004, **108**, 10200–10207.

(30)   C. D. Sherrill, T. Takatani and E. G. Hohenstein, "An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H$_2$S", *The Journal of Physical Chemistry A*, 2009, **113**, 10146–10159.

(31)   T. Takatani and C. David Sherrill, "Performance of spin-component-scaled Møller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions", *Physical Chemistry Chemical Physics*, 2007, **9**, 6106.

(32)   V. Tajti and G. Czakó, "Benchmark ab Initio Characterization of the Complex Potential Energy Surface of the F$^-$ + CH$_3$CH$_2$Cl Reaction", *The Journal of Physical Chemistry A*, 2017, **121**, 2847–2854.

(33)   J. Koput, "Ab initio potential energy surface and vibration-rotation energy levels of beryllium monohydroxide", *Journal of Computational Chemistry*, 2017, **38**, 37–43.

(34)   T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic Structure Theory*, John Wiley & Sons, Inc., 2000, pp. 315–327.

(35)   J. F. Ouyang and R. P. A. Bettens, "Many-body basis set superposition effect", *Journal of Chemical Theory and Computation*, 2015, **11**, 5132–5143.

(36)   J. F. Ouyang and R. P. A. Bettens, "When are Many-Body Effects Significant?", *Journal of Chemical Theory and Computation*, 2016, **12**, 5860–5867.

(37)   M. Schütz, G. Rauhut and H.-J. Werner, "Local Treatment of Electron Correlation in Molecular Clusters: Structures and Stabilities of (H$_2$O)$_n$, n = 2–4", *The Journal of Physical Chemistry A*, 1998, **102**, 5997–6003.

(38)   S. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors", *Molecular Physics*, 1970, **19**, 553–566.

(39) J. M. Cullen, "An examination of the effects of basis set and charge transfer in hydrogen-bonded dimers with a constrained Hartree-Fock method", *International Journal of Quantum Chemistry*, 1991, **40**, 193–207.

(40) E. Gianinetti, M. Raimondi and E. Tornaghi, "Modification of the Roothaan equations to exclude BSSE from molecular interaction calculations", *Internation Journal of Quantum Chemistry*, 1996, **60**, 157–166.

(41) T. Nagata, O. Takahashi, K. Saito and S. Iwata, "Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism", *The Journal of Chemical Physics*, 2001, **115**, 3553–3560.

(42) R. Z. Khaliullin, M. Head-Gordon and A. T. Bell, "An efficient self-consistent field method for large systems of weakly interacting components", *The Journal of Chemical Physics*, 2006, **124**, 204105.

(43) F. B. van Duijneveldt, J. van Duijneveldt-van de Rijdt and J. H. van Lenthe, "State of the Art in Counterpoise Theory", *Chemical Reviews*, 1994, **94**, 1873–1885.

(44) I. Mayer, "The chemical Hamiltonian approach for treating the BSSE problem of intermolecular interactions", *International Journal of Quantum Chemistry*, 1998, **70**, 41–63.

(45) B. Brauer, M. K. Kesharwani and J. M. L. Martin, "Some observations on counterpoise corrections for explicitly correlated calculations on noncovalent interactions", *Journal of Chemical Theory and Computation*, 2014, **10**, 3791–3799.

(46) R. M. Richard, K. U. Lao and J. M. Herbert, "Approaching the complete-basis limit with a truncated many-body expansion", *The Journal of Chemical Physics*, 2013, **139**, 224102.

(47) J. F. Ouyang, M. W. Cvitkovic and R. P. A. Bettens, "Trouble with the many-body expansion", *Journal of Chemical Theory and Computation*, 2014, **10**, 3699–3707.

(48) M. Urban and P. Hobza, "Weak intermolecular interaction", *Theoretica Chimica Acta*, 1975, **36**, 207–214.

(49) R. J. Vos, J. H. van Lenthe and F. B. van Duijneveldt, "Convergence to the configuration-set limit in multireference configuration-interaction calculations on the He dimer", *The Journal of Chemical Physics*, 1990, **93**, 643–651.

(50) T. van Mourik, "Basis Set Superposition Error Effects Cause the Apparent Nonexistence of the Ethene/Benzenium Ion Complex on the MP2 Potential Energy Surface", *The Journal of Physical Chemistry A*, 2008, **112**, 11017–11020.

(51) D. Toroz and T. van Mourik, "Structure of the gas-phase glycine tripeptide", *Physical Chemistry Chemical Physics*, 2010, **12**, 3463.

(52)   M. O. Sinnokrot and C. D. Sherrill, "High-Accuracy Quantum Mechanical Studies of $\pi \cdots \pi$ Interactions in Benzene Dimers", *The Journal of Physical Chemistry A*, 2006, **110**, 10656–10668.

(53)   K. R. Liedl, "Dangers of counterpoise corrected hypersurfaces. Advantages of basis set superposition improvement", *The Journal of Chemical Physics*, 1998, **108**, 3199–3204.

(54)   T. H. Dunning, "A Road Map for the Calculation of Molecular Binding Energies", *The Journal of Physical Chemistry A*, 2000, **104**, 9062–9080.

(55)   A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen and P. R. Taylor, "Basis set convergence of the interaction energy of hydrogen-bonded complexes", *The Journal of Chemical Physics*, 1999, **111**, 9157–9167.

(56)   T. Schwabe and S. Grimme, "Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability", *Physical Chemistry Chemical Physics*, 2007, **9**, 3397.

(57)   L. A. Burns, M. S. Marshall and C. D. Sherrill, "Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions", *Journal of Chemical Theory and Computation*, 2014, **10**, 49–57.

(58)   P. Valiron, Á. Vibók and I. Mayer, "Comparison of a posteriori and a priori BSSE correction schemes for SCF intermolecular energies", *Journal of Computational Chemistry*, 1993, **14**, 401–409.

(59)   P. W. Fowler and A. D. Buckingham, "The long range model of intermolecular forces", *Molecular Physics*, 1983, **50**, 1349–1361.

(60)   L. Turi and J. J. Dannenberg, "Correcting for basis set superposition error in aggregates containing more than two molecules: ambiguities in the calculation of the counterpoise correction", *The Journal of Physical Chemistry*, 1993, **97**, 2488–2490.

(61)   B. H. Wells and S. Wilson, "van der Waals interaction potentials: Many-body basis set superposition effects", *Chemical Physics Letters*, 1983, **101**, 429–434.

(62)   R. M. Richard, K. U. Lao and J. M. Herbert, "Achieving the CCSD(T) Basis-Set Limit in Sizable Molecular Clusters: Counterpoise Corrections for the Many-Body Expansion", *The Journal of Physical Chemistry Letters*, 2013, **4**, 2674–2680.

(63)   P. Valiron and I. Mayer, "Hierarchy of counterpoise corrections for N-body clusters: generalization of the Boys-Bernardi scheme", *Chemical Physics Letters*, 1997, **275**, 46–55.

(64)   B. Jeziorski, R. Moszynski and K. Szalewicz, "Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes", *Chemical Reviews*, 1994, **94**, 1187–1930.

(65)   A. Hesselmann, G. Jansen and M. Schütz, "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies", *The Journal of Chemical Physics*, 2005, **122**, 014103.

(66) Y. Mo, L. Song and Y. Lin, "Block-localized wavefunction (BLW) method at the density functional theory (DFT) level", *Journal of Physical Chemistry A*, 2007, **111**, 8291–8301.

(67) M. Sironi, A. Genoni, M. Civera, S. Pieraccini and M. Ghitti, "Extremely localized molecular orbitals: Theory and applications", *Theoretical Chemistry Accounts*, 2007, **117**, 685–698.

(68) M. Couty, C. A. Bayse and M. B. Hall, "Extremely localized molecular orbitals (ELMO): a non-orthogonal Hartree-Fock method", *Theor. Chem. Acc.*, 1997, **97**, 96–109.

(69) B. Meyer, B. Guillot, M. F. Ruiz-Lopez and A. Genoni, "Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability", *Journal of Chemical Theory and Computation*, 2016, **12**, 1052–1067.

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

(71) S Saebo and P Pulay, "Local Treatment of Electron Correlation", *Annual Reviews in Physical Chemistry*, 1993, **44**, 213–236.

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

(73) C. Hättig, W. Klopper, A. Köhn and D. P. Tew, "Explicitly Correlated Electrons in Molecules", *Chem. Rev.*, 2012, **112**, 4–74.

(74) P. Ettenhuber, P. Baudin, T. Kjærgaard, P. Jørgensen and K. Kristensen, "Orbital spaces in the divide-expand-consolidate coupled cluster method", *The Journal of Chemical Physics*, 2016, **144**, 164116.

(75) M. Saitow, U. Becker, C. Riplinger, E. F. Valeev and F. Neese, "A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory", *The Journal of Chemical Physics*, 2017, **146**, 164105.

(76) K. Wolinski and P. Pulay, "Second-order Møller-Plesset calculations with dual basis sets", *The Journal of Chemical Physics*, 2003, **118**, 9497–9503.

(77) R. P. Steele, R. A. DiStasio, Y. Shao, J. Kong and M. Head-Gordon, "Dual-basis second-order Møller-Plesset perturbation theory: A reduced-cost reference for correlation calculations", *The Journal of Chemical Physics*, 2006, **125**, 074108.

(78) J. Deng and P. M. W. Gill, "Communication: A new approach to dual-basis second-order Møller-Plesset calculations", *The Journal of Chemical Physics*, 2011, **134**, 1–4.

(79) A. Hamza, Á Vibók, G. J. Halász and I. Mayer, "BSSE-free SCF theories: A comment", *Journal of Molecular Structure: THEOCHEM*, 2000, **501-502**, 427–434.

(80) I. Mayer, "Hermitian "Chemical" Hamiltonian: an alternative ab initio method", *Theoretical Chemistry Accounts*, 2015, **134**, 86.

(81) H. L. Williams and C. F. Chabalowski, "Using Kohn-Sham Orbitals in Symmetry-Adapted Perturbation Theory to Investigate Intermolecular Interactions", *The Journal of Physical Chemistry A*, 2001, **105**, 646–659.

(82) R. Podeszwa and K. Szalewicz, "Accurate interaction energies for argon, krypton, and benzene dimers from perturbation theory based on the Kohn-Sham model", *Chemical Physics Letters*, 2005, **412**, 488–493.

(83) R. Z. Khaliullin, A. T. Bell and M. Head-Gordon, "Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals", *The Journal of Chemical Physics*, 2008, **128**, 184112.

(84) S. Karthikeyan, R. Sedlak and P. Hobza, "on the Nature of Stabilization in Weak, Medium, and Strong Charge-Transfer Complexes: CCSD(T)/CBS and SAPT Calculations", *The Journal of Physical Chemistry A*, 2011, **115**, 9422–9428.

(85) O. A. Stasyuk, R. Sedlak, C. F. Guerra and P. Hobza, "Comparison of the DFT-SAPT and Canonical EDA Schemes for the Energy Decomposition of Various Types of Noncovalent Interactions", *Journal of Chemical Theory and Computation*, 2018, **14**, 3440–3450.

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

(87) L. Shirkov and J. Makarewicz, "Does DFT-SAPT method provide spectroscopic accuracy?", *The Journal of Chemical Physics*, 2015, **142**, 064102.

(88) I. M. Høyvik and P. Jørgensen, "Characterization and Generation of Local Occupied and Virtual Hartree-Fock Orbitals", *Chemical Reviews*, 2016, **116**, 3306–3327.

(89) W. Ryback, R. Poirier and R. Kari, "An application of the method of conjugate gradients to the calculation of minimal basis set localized orbitals", *International Journal of Quantum Chemistry*, 1978, **13**, 1–16.

(90) T. Morikawa and Y. J. Ihaya, "Maximum and minimum overlap, localized, and hybrid orbitals for atoms and molecules by means of orthonormality-constrained variation", *International Journal of Quantum Chemistry*, 1978, **13**, 199–206.

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

(92) M. D. Newton, E. Switkes and W. N. Lipscomb, "Localized Bonds in SCF Wavefunctions for Polyatomic Molecules. III C–H and C–C Bonds", *The Journal of Chemical Physics*, 1970, **53**, 2645–2657.

(93) M. Levy, "Transferability of intrageminal correlation energies and oc-cupation coefficients", *The Journal of Chemical Physics*, 1974, **61**, 1857–1864.

(94) A. Famulari, E. Gianinetti, M. Raimondi and M. Sironi, "Implementa-tion of gradient-optimization algorithms and force constant compu-tations in BSSE-free direct and conventional SCF approaches", *Inter-national Journal of Quantum Chemistry*, 1998, **69**, 151–158.

(95) S. Iwata, "Absolutely local occupied and excited molecular orbitals in the third-order single excitation perturbation theory for molecular interaction", *Journal of Physical Chemistry A*, 2010, **114**, 8697–8704.

(96) S. Iwata, "Absolutely local excited orbitals in the higher order per-turbation expansion for the molecular interaction", *Journal of Physical Chemistry B*, 2008, **112**, 16104–16109.

(97) T. Nagata and S. Iwata, "Perturbation expansion theory corrected from basis set superposition error. I. Locally projected excited or-bitals and single excitations", *The Journal of Chemical Physics*, 2004, **120**, 3555–3562.

(98) S. Iwata, "The single excitation perturbation expansion theory based on the locally projected molecular orbitals for molecular interaction: Comparison with the counterpoise corrected energy", *Chemical Physics Letters*, 2006, **431**, 204–209.

(99) R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell and M. Head-Gordon, "Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals", *Journal of Physical Chemistry A*, 2007, **111**, 8753–8765.

(100) R. J. Azar and M. Head-Gordon, "An energy decomposition analysis for intermolecular interactions from an absolutely localized molecu-lar orbital reference at the coupled-cluster singles and doubles level", *The Journal of Chemical Physics*, 2012, **136**, 024103.

(101) J. Thirman and M. Head-Gordon, "Efficient Implementation of En-ergy Decomposition Analysis for Second-Order Møller-ĂIJPlesset Per-turbation Theory and Application to Anion-π Interactions", *The Jour-nal of Physical Chemistry A*, 2017, **121**, 717–728.

(102) Q. Ge, Y. Mao and M. Head-Gordon, "Energy decomposition analy-sis for exciplexes using absolutely localized molecular orbitals", *The Journal of Chemical Physics*, 2018, **148**, 064105.

(103) Q. Ge and M. Head-Gordon, "Energy Decomposition Analysis for Ex-cimers Using Absolutely Localized Molecular Orbitals within Time-Dependent Density Functional Theory and Configuration Interaction with Single Excitations", *Journal of Chemical Theory and Computation*, 2018, **14**, 5156–5168.

(104) R. Staub, M. Iannuzzi, R. Z. Khaliullin and S. N. Steinmann, "Energy Decomposition Analysis for Metal Surface-Adsorbate Interactions by Block Localized Wave Functions", *Journal of Chemical Theory and Com-putation*, 2019, **15**, 265–275.

(105) I. Mayer and Á Vibök, "SCF theory of intermolecular interactions without basis set superposition error", *Chemical Physics Letters*, 1987, **136**, 115–121.

(106) I. Mayer and L. Turi, "An analytical investigation into the bsse problem", *Journal of Molecular Structure: THEOCHEM*, 1991, **227**, 43–65.

(107) A. Conway and J. Murrell, "The dependence of exchange and coulomb energies on wave functions of the interacting systems", *Molecular Physics*, 1972, **23**, 1143–1146.

(108) Á Vibók and I. Mayer, "Intermolecular SCF theory without BSSE: The equations and some applications for small systems", *Journal of Molecular Structure: THEOCHEM*, 1988, **170**, 9–17.

(109) I. Mayer, "BSSE-free second-order intermolecular perturbation theory", *Molecular Physics*, 1997, **92**, 503–510.

(110) I. Mayer, "Bond order and valence indices: A personal account", *Journal of Computational Chemistry*, 2007, **28**, 204–221.

(111) I. Mayer, "Improved definition of bond orders for correlated wave functions", *Chemical Physics Letters*, 2012, **544**, 83–86.

(112) I. Mayer, "Improved chemical energy component analysis", *Physical Chemistry Chemical Physics*, 2012, **14**, 337–344.

(113) C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, "The continuous fast multipole method", *Chemical Physics Letters*, 1994, **230**, 8–16.

(114) C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, "Linear scaling density functional calculations via the continuous fast multipole method", *Chemical Physics Letters*, 1996, **253**, 268–278.

(115) C. A. White and M. Head-Gordon, "A J matrix engine for density functional theory calculations", *The Journal of Chemical Physics*, 1996, **104**, 2620–2629.

(116) M. C. Strain, G. E. Scuseria and M. J. Frisch, "Achieving Linear Scaling for the Electronic Quantum Coulomb Problem", *Science*, 1996, **271**, 51–53.

(117) Y. Shao and M. Head-Gordon, "An improved J matrix engine for density functional theory calculations", *Chemical Physics Letters*, 2000, **323**, 425–433.

(118) L. Füsti-Molnár and J. Kong, "Fast and accurate Coulomb calculation with Gaussian functions", *The Journal of Chemical Physics*, 2005, **122**, 074108.

(119) F. R. Manby, P. J. Knowles and A. W. Lloyd, "The Poisson equation in density fitting for the Kohn-Sham Coulomb problem", *The Journal of Chemical Physics*, 2001, **115**, 9144–9148.

(120) E. Schwegler, M. Challacombe and M. Head-Gordon, "Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build", *The Journal of Chemical Physics*, 1997, **106**, 9708–9717.

(121) C. Ochsenfeld, C. A. White and M. Head-Gordon, "Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices", *The Journal of Chemical Physics*, 1998, **109**, 1663–1669.

(122) R. Polly, H.-J. Werner, F. R. Manby and P. J. Knowles, "Fast Hartree-Fock theory using local density fitting approximations", *Molecular Physics*, 2004, **102**, 2311–2321.

(123) C. Köppl and H.-J. Werner, "Parallel and low-order scaling implementation of Hartree-Fock exchange using local density fitting", *Journal of Chemical Theory and Computation*, 2016, **12**, 3122–3134.

(124) G. Golub and C. van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 4th ed., 2012.

(125) W. Klopper and W. Kutzelnigg, "Møller-plesset calculations taking care of the correlation cusp", *Chemical Physics Letters*, 1987, **134**, 17–22.

(126) M. R. Nyden and G. A. Petersson, "Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions", *The Journal of Chemical Physics*, 1981, **75**, 1843–1862.

(127) C. Schwartz, "Importance of Angular Correlations between Atomic Electrons", *Physical Review*, 1962, **126**, 1015–1019.

(128) T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen", *The Journal of Chemical Physics*, 1989, **90**, 1007–1023.

(129) T. H. Dunning, K. A. Peterson and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited", *The Journal of Chemical Physics*, 2001, **114**, 9244–9253.

(130) L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, "Gaussian-2 theory for molecular energies of first- and second-row compounds", *The Journal of Chemical Physics*, 1991, **94**, 7221–7230.

(131) S. Havriliak and H. F. King, "Rydberg radicals. 1. Frozen-core model for Rydberg levels of the ammonium radical", *Journal of the American Chemical Society*, 1983, **105**, 4–12.

(132) S. Havriliak, T. R. Furlani and H. F. King, "Rydberg levels of the sodium atom and the ammonium radical calculated by perturbation theory", *Canadian Journal of Physics*, 1984, **62**, 1336–1346.

(133) R. Jurgens-Lutovsky and J. Almlöf, "Dual basis sets in calculations of electron correlation", *Chemical Physics Letters*, 1991, **178**, 451–454.

(134) R. P. Steele, Y. Shao, R. A. DiStasio and M. Head-Gordon, "Dual-basis analytic gradients. 1. Self-consistent field theory", *Journal of Physical Chemistry A*, 2006, **110**, 13915–13922.

(135) R. P. Steele, M. Head-Gordon and J. C. Tully, "Ab Initio Molecular Dynamics with Dual Basis Set Methods", *The Journal of Physical Chemistry A*, 2010, **114**, 11853–11860.

(136) J. Zhang and M. Dolg, "Third-Order Incremental Dual-Basis Set Zero-Buffer Approach: An Accurate and Efficient Way To Obtain CCSD and CCSD(T) Energies", *Journal of Chemical Theory and Computation*, 2013, **9**, 2992–3003.

(137) J. Deng, A. T. B. Gilbert and P. M. W. Gill, "MP2[V]: A Simple Approximation to Second-Order Møller-Plesset Perturbation Theory", *Journal of Chemical Theory and Computation*, 2015, **11**, 1639–1644.

(138) L. Kong and E. F. Valeev, "Perturbative correction for the basis set incompleteness error of complete-active-space self-consistent field", *The Journal of Chemical Physics*, 2010, **133**, 174126.

(139) L. Kong, F. A. Bischoff and E. F. Valeev, "Explicitly correlated R12/F12 methods for electronic structure", *Chemical Reviews*, 2011, **112**, 75–107.

(140) T. B. Adler, G. Knizia and H.-J. Werner, "A simple and efficient CCSD(T)-F12 approximation", *The Journal of Chemical Physics*, 2007, **127**, 221106.

(141) T. H. Rasmussen, Y. M. Wang, T. Kjærgaard and K. Kristensen, "The explicitly correlated same number of optimized parameters (SNOOP-F12) scheme for calculating intermolecular interaction energies", *The Journal of Chemical Physics*, 2016, **144**, 204102.

(142) D. P. Tew, "Explicitly correlated coupled-cluster theory with Brueckner orbitals", *The Journal of Chemical Physics*, 2016, **145**, 074103.

(143) R. A. Shaw and J. G. Hill, "Approaching the Hartree-Fock Limit through the Complementary Auxiliary Basis Set Singles Correction and Auxiliary Basis Sets", *Journal of Chemical Theory and Computation*, 2017, **13**, 1691–1698.

(144) W. Klopper, F. R. Manby, S. Ten-No and E. F. Valeev, "R12 methods in explicitly correlated molecular electronic structure theory", *Internation Reviews in Physical Chemistry*, 2006, **25**, 427–468.

(145) J. Garza, J.-Z. Ramírez and R. Vargas, "Role of Hartree-Fock and Kohn-Sham Orbitals in the Basis Set Superposition Error for Systems Linked by Hydrogen Bonds", *The Journal of Physical Chemistry A*, 2005, **109**, 643–651.

(146) J. L. Whitten, "Coulombic potential energy integrals and approximations", *The Journal of Chemical Physics*, 1973, **58**, 4496–4501.

(147) B. I. Dunlap, J. W. D. Connolly and J. R. Sabin, "On the applicability of LCAO-X methods to molecules containing transition metal atoms: The nickel atom and nickel hydride", *International Journal of Quantum Chemistry*, 1977, **12**, 81–87.

(148) B. I. Dunlap, J. W. D. Connolly and J. R. Sabin, "On some approximations in applications of X -theory", *The Journal of Chemical Physics*, 1979, **71**, 3396–3402.

(149) O. Vahtras, J. Almlöf and M. Feyereisen, "Integral approximations for LCAO-SCF calculations", *Chemical Physics Letters*, 1993, **213**, 514–518.

(150) A. P. Rendell and T. J. Lee, "Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks", *The Journal of Chemical Physics*, 1994, **101**, 400–408.

(151) F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency", *Physical Chemistry Chemical Physics*, 2002, **4**, 4285–4291.

(152) H.-J. Werner, F. R. Manby and P. J. Knowles, "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations", *J. Chem. Phys.*, 2003, **118**, 8149–8160.

(153) S. Höfener, "Coupled-cluster frozen-density embedding using resolution of the identity methods", *Journal of Computational Chemistry*, 2014, **35**, 1716–1724.

(154) H. F. Schurkus, A. Luenser and C. Ochsenfeld, "Communication: Almost error-free resolution-of-the-identity correlation methods by null space removal of the particle-hole interactions", *The Journal of Chemical Physics*, 2017, **146**, 211106.

(155) J. Boström, F. Aquilante, T. B. Pedersen and R. Lindh, "Ab Initio Density Fitting: Accuracy Assessment of Auxiliary Basis Sets from Cholesky Decompositions", *Journal of Chemical Theory and Computation*, 2009, **5**, 1545–1553.

(156) F. Aquilante, T. B. Pedersen and R. Lindh, "Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals", *The Journal of Chemical Physics*, 2007, **126**, 194106.

(157) M. S. Marshall, J. S. Sears, L. A. Burns, J.-L. Brédas and C. D. Sherrill, "An Error and Efficiency Analysis of Approximations to Møller-Plesset Perturbation Theory", *Journal of Chemical Theory and Computation*, 2010, **6**, 3681–3687.

(158) H. Koch, A. Sánchez de Merás and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions", *The Journal of Chemical Physics*, 2003, **118**, 9481–9484.

(159) R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz and Y. Cao, "Correlated ab Initio Electronic Structure Calculations for Large Molecules", *The Journal of Physical Chemistry A*, 1999, **103**, 1913–1928.

(160) R. Łazarski, A. M. Burow and M. Sierka, "Density Functional Theory for Molecular and Periodic Systems Using Density Fitting and Continuous Fast Multipole Methods", *Journal of Chemical Theory and Computation*, 2015, **11**, 3029–3041.

(161) R. Łazarski, A. M. Burow, L. Grajciar and M. Sierka, "Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole method: Analytical gradients", *Journal of Computational Chemistry*, 2016, **37**, 2518–2526.

(162)   F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn", *Physical Chemistry Chemical Physics*, 2006, **8**, 1057.

(163)   J. C. Womack and F. R. Manby, "Density fitting for three-electron integrals in explicitly correlated electronic structure theory", *The Journal of Chemical Physics*, 2014, **140**, 044118.

(164)   F. R. Manby, "Density fitting in second-order linear-r12 Møller-Plesset perturbation theory", *The Journal of Chemical Physics*, 2003, **119**, 4607–4613.

(165)   M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, "Fragmentation Methods: A Route to Accurate Calculations on Large Systems", *Chemical Reviews*, 2012, **112**, 632–672.

(166)   K. Raghavachari and A. Saha, "Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules", *Chemical Reviews*, 2015, **115**, 5643–5677.

(167)   Y. Okiyama, T. Nakano, C. Watanabe, K. Fukuzawa, Y. Mochizuki and S. Tanaka, "Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson-Boltzmann Equation: Implementation and DNA Study", *The Journal of Physical Chemistry B*, 2018, **122**, 4457–4471.

(168)   V. Q. Vuong, Y. Nishimoto, D. G. Fedorov, B. G. Sumpter, T. A. Niehaus and S. Irle, "The Fragment Molecular Orbital Method Based on Long-Range Corrected Density-Functional Tight-Binding", *Journal of Chemical Theory and Computation*, 2019, **15**, 3008–3020.

(169)   J. Lennard-Jones, "The Molecular Orbital Theory of Chemical Valency. I. The Determination of Molecular Orbitals", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1949, **198**, 1–13.

(170)   S. F. Boys, "Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another", *Reviews of Modern Physics*, 1960, **32**, 296–299.

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

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

(173)   G. J. O. Beran and S. Hirata, "Fragment and localized orbital methods in electronic structure theory", *Physical Chemistry Chemical Physics*, 2012, **14**, 7559.

(174)   J. J. Eriksen, P. Baudin, P. Ettenhuber, K. Kristensen, T. Kjærgaard and P. Jørgensen, "Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide-Expand-Consolidate CCSD(T) Model", *Journal of Chemical Theory and Computation*, 2015, **11**, 2984–2993.

(175) T. Kjærgaard, P. Baudin, D. Bykov, K. Kristensen and P. Jørgensen, "The divide-expand-consolidate coupled cluster scheme", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2017, e1319.

(176) F. Neese, F. Wennmohs and A. Hansen, "Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method", *The Journal of Chemical Physics*, 2009, **130**, 114108.

(177) D. G. Liakos and F. Neese, "Domain Based Pair Natural Orbital Coupled Cluster Studies on Linear and Folded Alkane Chains", *Journal of Chemical Theory and Computation*, 2015, **11**, 2137–2143.

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

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

(180) G. Schmitz, C. Hättig and D. P. Tew, "Explicitly correlated PNO-MP2 and PNO-CCSD and their application to the S66 set and large molecular systems", *Physical Chemistry Chemical Physics*, 2014, **16**, 22167–22178.

(181) Q. Ma and H.-J. Werner, "Accurate Intermolecular Interaction Energies Using Explicitly Correlated Local Coupled Cluster Methods [PNO-LCCSD(T)-F12]", *Journal of Chemical Theory and Computation*, 2019, **15**, 1044–1052.

(182) P. Y. Ayala and G. E. Scuseria, "Linear scaling second-order Møller-Plesset theory in the atomic orbital basis for large molecular systems", *The Journal of Chemical Physics*, 1999, **110**, 3660–3671.

(183) P. Y. Ayala, K. N. Kudin and G. E. Scuseria, "Atomic orbital Laplace-transformed second-order Møller-Plesset theory for periodic systems", *The Journal of Chemical Physics*, 2001, **115**, 9698–9707.

(184) A. El Azhary, G. Rauhut, P. Pulay and H.-J. Werner, "Analytical energy gradients for local second-order Møller-Plesset perturbation theory", *The Journal of Chemical Physics*, 1998, **108**, 5185–5193.

(185) G. Rauhut, P. Pulay and H.-J. Werner, "Integral transformation with low-order scaling for large local second-order Møller-Plesset calculations", *Journal of Computational Chemistry*, 1998, **19**, 1241–1254.

(186) K. Morokuma, "Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity", *Accounts of Chemical Research*, 1977, **10**, 294–300.

(187) K. Kitaura and K. Morokuma, "A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation", *International Journal of Quantum Chemistry*, 1976, **10**, 325–340.

(188) R. F. Frey and E. R. Davidson, "Energy partitioning of the self-consistent field interaction energy of ScCO", *The Journal of Chemical Physics*, 1989, **90**, 5555–5562.

(189) M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys and W. J. Stevens, "The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry", *The Journal of Physical Chemistry A*, 2001, **105**, 293–307.

(190) M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko and J. A. Boatz, "Accurate Methods for Large Molecular Systems", *The Journal of Physical Chemistry B*, 2009, **113**, 9646–9663.

(191) P. K. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. I. Krylov and L. V. Slipchenko, "Extension of the Effective Fragment Potential Method to Macromolecules", *The Journal of Physical Chemistry B*, 2016, **120**, 6562–6574.

(192) E. D. Glendening and A. Streitwieser, "Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding, strong ionic, and moderate donor-acceptor interactions", *The Journal of Chemical Physics*, 1994, **100**, 2900–2909.

(193) E. D. Glendening, "Natural Energy Decomposition Analysis: Extension to Density Functional Methods and Analysis of Cooperative Effects in Water Clusters", *The Journal of Physical Chemistry A*, 2005, **109**, 11936–11940.

(194) F. Weinhold, C. Landis and E. Glendening, "What is NBO analysis and how is it useful?", *International Reviews in Physical Chemistry*, 2016, **35**, 399–440.

(195) A. J. Stone, "Natural Bond Orbitals and the Nature of the Hydrogen Bond", *The Journal of Physical Chemistry A*, 2017, **121**, 1531–1534.

(196) M. P. Mitoraj, A. Michalak and T. Ziegler, "A Combined Charge and Energy Decomposition Scheme for Bond Analysis", *Journal of Chemical Theory and Computation*, 2009, **5**, 962–975.

(197) A. J. Sadlej, "Long range induction and dispersion interactions between Hartree-Fock subsystems", *Molecular Physics*, 1980, **39**, 1249–1264.

(198) J. Thirman and M. Head-Gordon, "An energy decomposition analysis for second-order Møller-Plesset perturbation theory based on absolutely localized molecular orbitals", *The Journal of Chemical Physics*, 2015, **143**, 084124.

(199) Q. Wu, P. W. Ayers and Y. Zhang, "Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies", *The Journal of Chemical Physics*, 2009, **131**, 164112.

(200) M. H. Cohen, A. Wasserman, R. Car and K. Burke, "Charge Transfer in Partition Theory", *The Journal of Physical Chemistry A*, 2009, **113**, 2183–2192.

(201) P. Elliott, K. Burke, M. H. Cohen and A. Wasserman, "Partition density-functional theory", *Physical Review A*, 2010, **82**, 024501.

(202) M. Domagała, A. Lutyńska and M. Palusiak, "Halogen bond versus hydrogen bond: The many-body interactions approach", *International Journal of Quantum Chemistry*, 2017, **117**, e25348.

(203) R. Luyckx, P. Coulon and H. N. W. Lekkerkerker, "Dispersion forces between noble gas atoms", *The Journal of Chemical Physics*, 1978, **69**, 2424.

(204) A. Altun, M. Saitow, F. Neese and G. Bistoni, "Local Energy Decomposition of Open-Shell Molecular Systems in the Domain-Based Local Pair Natural Orbital Coupled Cluster Framework", *Journal of Chemical Theory and Computation*, 2019, **15**, 1616–1632.

(205) H. Scheiber, Y. Shi and R. Z. Khaliullin, "Communication: Compact orbitals enable low-cost linear-scaling *ab initio* molecular dynamics for weakly-interacting systems", *The Journal of Chemical Physics*, 2018, **148**, 231103.

(206) B. F. E. Curchod and T. J. Martínez, "Ab Initio Nonadiabatic Quantum Molecular Dynamics", *Chemical Reviews*, 2018, **118**, 3305–3336.

(207) M. Biczysko, J. Bloino and C. Puzzarini, "Computational challenges in Astrochemistry: Computational challenges in Astrochemistry", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2018, **8**, e1349.

(208) A. Kleshchonok and A. Tkatchenko, "Tailoring van der Waals dispersion interactions with external electric charges", *Nature Communications*, 2018, **9**, 3017.

(209) S. Saha, M. K. Mishra, C. M. Reddy and G. R. Desiraju, "From Molecules to Interactions to Crystal Engineering: Mechanical Properties of Organic Solids", *Accounts of Chemical Research*, 2018, **51**, 2957–2967.

(210) A Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill, New York, 1st ed., reprint, 1989.

(211) D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion", *Mathematical Proceedings of the Cambridge Philosophical Society*, 1928, **24**, 111–132.

(212) J. C. Slater, "The Self Consistent Field and the Structure of Atoms", *Physical Review*, 1928, **32**, 339–348.

(213) V. Fock, "Selfconsistent field mit Austausch für Natrium", *Zeitschrift für Physik*, 1930, **62**, 795–805.

(214) B. Levy and G. Berthier, "Generalized Brillouin theorem for multiconfigurational SCF theories", *International Journal of Quantum Chemistry*, 1968, **2**, 307–319.

(215) E. Kreyszig, *Introductory functional analysis with applications*, OCLC: 256624333, Wiley, New York, 1989.

(216) R. S. Mulliken, "Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I", *The Journal of Chemical Physics*, 1955, **23**, 1833–1840.

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

(218) G. Knizia, "Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts", *Journal of Chemical Theory and Computation*, 2013, **9**, 4834–4843.

(219) A. J. Garza and G. E. Scuseria, "Comparison of self-consistent field convergence acceleration techniques", *The Journal of Chemical Physics*, 2012, **137**, 054110.

(220) T. Rohwedder and R. Schneider, "An analysis for the DIIS acceleration method used in quantum chemistry calculations", *Journal of Mathematical Chemistry*, 2011, **49**, 1889–1914.

(221) P. Pulay, "Improved SCF convergence acceleration", *Journal of Computational Chemistry*, 1982, **3**, 556–560.

(222) I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics*, Cambridge University Press, Cambridge, UK, 1st ed., 2009.

(223) W. Liang and M. Head-Gordon, "An exact reformulation of the diagonalization step in electronic structure calculations as a set of second order nonlinear equations", *The Journal of Chemical Physics*, 2004, **120**, 10379–10384.

(224) J. J. Sakurai, *Modern Quantum Mechanics: Revised Edition*, ed. S. F. Tuan, Addison-Wesley Publishing Company, Inc., Boston, MA, 1994.

(225) A. D. McLachlan, "Retarded Dispersion Forces Between Molecules", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1963, **271**, 387–401.

(226) J. Hermann, R. A. DiStasio and A. Tkatchenko, "First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications", *Chemical Reviews*, 2017, **117**, 4714–4758.

(227) A. Zee, *Quantum Field Theory in a Nutshell*, Princeton University Press, Princeton, NJ, 2nd ed., 2010.

(228) M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,* Dover Publications, Incorporated, New York, 1974.

(229) A. D. McLachlan and M. A. Ball, "Time-Dependent Hartree-Fock Theory for Molecules", *Reviews of Modern Physics*, 1964, **36**, 844–855.

(230) J. Toulouse, I. C. Gerber, G. Jansen, A. Savin and J. G. Ángyán, "Adiabatic-Connection Fluctuation-Dissipation Density-Functional Theory Based on Range Separation", *Physical Review Letters*, 2009, **102**, 096404.

(231) D. Bohm and D. Pines, "A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas", *Physical Review*, 1953, **92**, 609–625.

(232) X. Ren, P. Rinke, C. Joas and M. Scheffler, "Random-phase approximation and its applications in computational chemistry and materials science", *Journal of Materials Science*, 2012, **47**, 7447–7471.

(233) E. Wigner and F. Seitz, "On the Constitution of Metallic Sodium", *Physical Review*, 1933, **43**, 804–810.

(234) E. Wigner and F. Seitz, "On the Constitution of Metallic Sodium. II", *Physical Review*, 1934, **46**, 509–524.

(235) F. Aryasetiawan and O. Gunnarsson, "The GW method", *Reports on Progress in Physics*, 1998, **61**, 237–312.

(236) A. Tkatchenko, A. Ambrosetti and R. A. DiStasio, "Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem", *The Journal of Chemical Physics*, 2013, **138**, 074106.

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

(238) D. Bohm and D. Pines, "A Collective Description of Electron Interactions. I. Magnetic Interactions", *Physical Review*, 1951, **82**, 625–634.

(239) D. Pines and D. Bohm, "A Collective Description of Electron Interactions: II. Collective vs. Individual Particle Aspects of the Interactions", *Physical Review*, 1952, **85**, 338–353.

(240) D. Pines, "A Collective Description of Electron Interactions: IV. Electron Interaction in Metals", *Physical Review*, 1953, **92**, 626–636.

(241) P. T. Landsberg, "A Contribution to the Theory of Soft X-ray Emission Bands of Sodium", *Proceedings of the Physical Society. Section A*, 1949, **62**, 806–816.

(242) E. Wohlfarth, "XLIX. The influence of exchange and correlation forces on the specific heat of free electrons in metals", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1950, **41**, 534–542.

(243) M. Gell-Mann and K. A. Brueckner, "Correlation Energy of an Electron Gas at High Density", *Physical Review*, 1957, **106**, 364–368.

(244) J. Goldstone, "Derivation of the Brueckner Many-Body Theory", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1957, **239**, 267–279.

(245) A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, "Long-range correlation energy calculated from coupled atomic response functions", *The Journal of Chemical Physics*, 2014, **140**, 18A508.

(246) A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, "Accurate and Efficient Method for Many-Body van der Waals Interactions", *Physical Review Letters*, 2012, **108**, 236402.

(247) J. Toulouse, W. Zhu, J. G. Ángyán and A. Savin, "Range-separated density-functional theory with the random-phase approximation: Detailed formalism and illustrative applications", *Physical Review A*, 2010, **82**, 032502.

(248) V. G. Ruiz, W. Liu and A. Tkatchenko, "Density-functional theory with screened van der Waals interactions applied to atomic and molecular adsorbates on close-packed and non-close-packed surfaces", *Physical Review B*, 2016, **93**, 035118.

(249) G. P. Chen, V. K. Voora, M. M. Agee, S. G. Balasubramani and F. Furche, "Random-Phase Approximation Methods", *Annual Review of Physical Chemistry*, 2017, **68**, 421–445.

(250) H. Eshuis, J. E. Bates and F. Furche, "Electron correlation methods based on the random phase approximation", *Theoretical Chemistry Accounts*, 2012, **131**, 1084.

(251) A.-S. Hehn, D. P. Tew and W. Klopper, "Explicitly correlated ring-coupled-cluster-doubles theory", *The Journal of Chemical Physics*, 2015, **142**, 194106.

(252) G. Jansen, R.-F. Liu and J. G. Ángyán, "On the equivalence of ring-coupled cluster and adiabatic connection fluctuation-dissipation theorem random phase approximation correlation energy expressions", *The Journal of Chemical Physics*, 2010, **133**, 154106.

(253) G. E. Scuseria, T. M. Henderson and D. C. Sorensen, "The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach", *The Journal of Chemical Physics*, 2008, **129**, 231101.

(254) G. E. Scuseria, T. M. Henderson and I. W. Bulik, "Particle-particle and quasiparticle random phase approximations: Connections to coupled cluster theory", *The Journal of Chemical Physics*, 2013, **139**, 104113.

(255) J. Toulouse, W. Zhu, A. Savin, G. Jansen and J. G. Ángyán, "Closed-shell ring coupled cluster doubles theory with range separation applied on weak intermolecular interactions", *The Journal of Chemical Physics*, 2011, **135**, 084119.

(256) R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, Dover Publications, Inc., New York, 2nd ed., 1992.

(257) L. Jódar, "A formula for the general solution of Riccati type matrix differential equations", *Systems & Control Letters*, 1989, **12**, 39–43.

(258) A. J. W. Thom, "Stochastic Coupled Cluster Theory", *Physical Review Letters*, 2010, **105**, 263004.

(259) M. Schütz and H.-J. Werner, "Local perturbative triples correction (T) with linear cost scaling", *Chemical Physics Letters*, 2000, **318**, 370–378.

(260) Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo and F. Neese, "Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]", *The Journal of Chemical Physics*, 2018, **148**, 011101.

(261) D. Kats and A. Köhn, "On the distinguishable cluster approximation for triple excitations", *The Journal of Chemical Physics*, 2019, **150**, 151101.

(262) D. I. Lyakh, V. V. Ivanov and L. Adamowicz, "Automated generation of coupled-cluster diagrams: Implementation in the multireference state-specific coupled-cluster approach with the complete-active-space reference", *The Journal of Chemical Physics*, 2005, **122**, 024108.

(263) A. A. Auer, G. Baumgartner, D. E. Bernholdt, A. Bibireata, V. Choppella, D. Cociorva, X. Gao, R. Harrison, S. Krishnamoorthy, S. Krishnan, C.-C. Lam, Q. Lu, M. Nooijen, R. Pitzer, J. Ramanujam, P. Sadayappan and A. Sibiryakov, "Automatic code generation for many-body electronic structure methods: the tensor contraction engine", *Molecular Physics*, 2006, **104**, 211–228.

(264) M. Krupička, K. Sivalingam, L. Huntington, A. A. Auer and F. Neese, "A toolchain for the automatic generation of computer codes for correlated wavefunction calculations", *Journal of Computational Chemistry*, 2017, **38**, 1853–1868.

(265) K. Z. Ibrahim, E. Epifanovsky, S. Williams and A. I. Krylov, "Cross-scale efficient tensor contractions for coupled cluster computations through multiple programming model backends", *Journal of Parallel and Distributed Computing*, 2017, **106**, 92–105.

(266) T. D. Crawford, T. J. Lee, N. C. Handy and H. F. Schaefer, "Spin-restricted Brueckner orbitals for coupled-cluster wavefunctions", *The Journal of Chemical Physics*, 1997, **107**, 9980–9984.

(267) J. Řezáč, K. E. Riley and P. Hobza, "S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures", *Journal of Chemical Theory and Computation*, 2011, **7**, 2427–2438.

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

(269) J. F. Gonthier and M. Head-Gordon, "Assessing Electronic Structure Methods for Long-Range Three-Body Dispersion Interactions: Analysis and Calculations on Well-Separated Metal Atom Trimers", *Journal of Chemical Theory and Computation*, 2019, **15**, 4351–4361.

(270) A. Altun, F. Neese and G. Bistoni, "Effect of Electron Correlation on Intermolecular Interactions: A Pair Natural Orbitals Coupled Cluster Based Local Energy Decomposition Study", *Journal of Chemical Theory and Computation*, 2019, **15**, 215–228.

(271) M. J. Deegan and P. J. Knowles, "Perturbative corrections to account for triple excitations in closed and open shell coupled cluster theories", *Chemical Physics Letters*, 1994, **227**, 321–326.

(272) K. S. Singwi, M. P. Tosi, R. H. Land and A. Sjolander, "Electron Correlations at Metallic Densities", *Physical Review*, 1968, **176**, 589–599.

(273) J. G. Ángyán, R.-F. Liu, J. Toulouse and G. Jansen, "Correlation Energy Expressions from the Adiabatic-Connection Fluctuation-Dissipation Theorem Approach", *Journal of Chemical Theory and Computation*, 2011, **7**, 3116–3130.

(274) J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Gruneis and G. Kresse, "Hybrid functionals including random phase approximation correlation and second-order screened exchange", *The Journal of Chemical Physics*, 2010, **132**, 094103.

(275) A. Heßelmann, "Third-order corrections to random-phase approximation correlation energies", *The Journal of Chemical Physics*, 2011, **134**, 204107.

(276) M. Beuerle and C. Ochsenfeld, "Short-range second order screened exchange correction to RPA correlation energies", *The Journal of Chemical Physics*, 2017, **147**, 204107.

(277) G. P. Chen, M. M. Agee and F. Furche, "Performance and Scope of Perturbative Corrections to Random-Phase Approximation Energies", *Journal of Chemical Theory and Computation*, 2018, **14**, 5701–5714.

(278) X. Ren, N. Marom, F. Caruso, M. Scheffler and P. Rinke, "Beyond the GW approximation: A second-order screened exchange correction", *Physical Review B*, 2015, **92**, 081104.

(279) L. Hedin, "On correlation effects in electron spectroscopies and the GW approximation", *Journal of Physics: Condensed Matter*, 1999, **11**, R489–R528.

(280) C. Friedrich and A. Schindlmayr, in *Computational Nanoscience: Do It Yourself!*, ed. J. Grotendorst, S. Blügel and J. von Neumann-Institut für Computing, NIC-Secretariat, Research Centre Jülich, Jülich, 2006, pp. 335–355.

(281) A. Schindlmayr and R. W. Godby, "Systematic Vertex Corrections through Iterative Solution of Hedin's Equations Beyond the GW Approximation", *Physical Review Letters*, 1998, **80**, 1702–1705.

(282) F. Bruneval, F. Sottile, V. Olevano, R. Del Sole and L. Reining, "Many-Body Perturbation Theory Using the Density-Functional Concept: Beyond the GW Approximation", *Physical Review Letters*, 2005, **94**, 186402.

(283) P. Romaniello, S. Guyot and L. Reining, "The self-energy beyond GW: Local and nonlocal vertex corrections", *The Journal of Chemical Physics*, 2009, **131**, 154111.

(284) M. Grumet, P. Liu, M. Kaltak, J. Klime and G. Kresse, "Beyond the quasiparticle approximation: Fully self-consistent G W calculations", *Physical Review B*, 2018, **98**, 155143.

(285) F. Bechstedt, "Correlation beyond the random phase approximation: A consistent many-body perturbation theory approach", *Physical Review B*, 2018, **97**, 241109.

(286) H. Ma, M. Govoni, F. Gygi and G. Galli, "A Finite-Field Approach for GW Calculations beyond the Random Phase Approximation", *Journal of Chemical Theory and Computation*, 2019, **15**, 154–164.

(287) M. Kállay, "Linear-scaling implementation of the direct random-phase approximation", *The Journal of Chemical Physics*, 2015, **142**, 204105.

(288) C. Kalai and J. Toulouse, "A general range-separated double-hybrid density-functional theory", *The Journal of Chemical Physics*, 2018, **148**, 164105.

(289) F. Weigend, "Hartree-Fock exchange fitting basis sets for H to Rn", *Journal of Computational Chemistry*, 2008, **29**, 167–175.

(290) M. Schutz, R. Lindh and H.-J. Werner, "Integral-direct electron correlation methods", *Molecular Physics*, 1999, **96**, 719–733.

(291) J. Almlöf, K. Faegri and K. Korsell, "Principles for a direct SCF approach to calculations", *Journal of Computational Chemistry*, 1982, **3**, 385–399.

(292) S. A. Maurer, D. S. Lambrecht, D. Flaig and C. Ochsenfeld, "Distance-dependent Schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. Rigorous upper bounds", *The Journal of Chemical Physics*, 2012, **136**, 144107.

(293) Q. Wang, X. Zhang, Y. Zhang and Q. Yi, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '13, ACM Press, Denver, Colorado, 2013, pp. 1–12.

(294) V. Volkov and J. Demmel, 2008 SC - International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE, Austin, TX, 2008, pp. 1–11.

(295) E. Agullo, J. Demmel, J. Dongarra, B. Hadri, J. Kurzak, J. Langou, H. Ltaief, P. Luszczek and S. Tomov, "Numerical linear algebra on emerging architectures: The PLASMA and MAGMA projects", *Journal of Physics: Conference Series*, 2009, **180**, 012037.

(296) J. D. C. Maia, G. A. Urquiza Carvalho, C. P. Mangueira, S. R. Santana, L. A. F. Cabral and G. B. Rocha, "GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations", *Journal of Chemical Theory and Computation*, 2012, **8**, 3072–3081.

(297) B. Korte and J. Vygen, *Combinatorial optimization: theory and algorithms*, OCLC: 254652176, Springer, Berlin, 3. ed, 2006.

(298) G. Finn and E. Horowitz, "A linear time approximation algorithm for multiprocessor scheduling", *BIT*, 1979, **19**, 312–320.

(299) R. E. Korf, Eighteenth National Conference on Artificial Intelligence, event-place: Edmonton, Alberta, Canada, American Association for Artificial Intelligence, Menlo Park, CA, USA, 2002, pp. 731–736.

(300) J. Nieplocha, B. Palmer, V. Tipparaju, M. Krishnan, H. Trease and E. Aprà, "Advances, Applications and Performance of the Global Arrays Shared Memory Programming Toolkit", *The International Journal of High Performance Computing Applications*, 2006, **20**, 203–231.

(301)   J. Nickolls, I. Buck, M. Garland and K. Skadron, "Scalable parallel programming with CUDA", *Queue*, 2008, **6**, 40.

(302)   J. Nickolls and W. J. Dally, "The GPU Computing Era", *IEEE Micro*, 2010, **30**, 56–69.

(303)   D. Kirk and W.-m. Hwu, *Programming massively parallel processors hands-on with CUDA*, Morgan Kaufmann Publishers, Burlington, MA, 2010.

(304)   D. Luebke, 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, Paris, France, 2008, pp. 836–838.

(305)   V. V. Kindratenko, J. J. Enos, G. Shi, M. T. Showerman, G. W. Arnold, J. E. Stone, J. C. Phillips and W.-m. Hwu, 2009 IEEE International Conference on Cluster Computing and Workshops, IEEE, New Orleans, LA, USA, 2009, pp. 1–8.

(306)   J. A. Anderson, C. D. Lorenz and A. Travesset, "General purpose molecular dynamics simulations fully implemented on graphics processing units", *Journal of Computational Physics*, 2008, **227**, 5342–5359.

(307)   A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born", *Journal of Chemical Theory and Computation*, 2012, **8**, 1542–1555.

(308)   S. Le Grand, A. W. Götz and R. C. Walker, "SPFP: Speed without compromise: A mixed precision model for GPU accelerated molecular dynamics simulations", *Computer Physics Communications*, 2013, **184**, 374–380.

(309)   S. I. Allec, Y. Sun, J. Sun, C.-e. A. Chang and B. M. Wong, "Heterogeneous CPU+GPU-Enabled Simulations for DFTB Molecular Dynamics of Large Chemical and Biological Systems", *Journal of Chemical Theory and Computation*, 2019, **15**, 2807–2815.

(310)   Y. Zhao, "Lattice Boltzmann based PDE solver on the GPU", *The Visual Computer*, 2008, **24**, 323–333.

(311)   E. F. Bollig, N. Flyer and G. Erlebacher, "Solution to PDEs using radial basis function finite-differences (RBF-FD) on multiple GPUs", *Journal of Computational Physics*, 2012, **231**, 7133–7151.

(312)   A. Gloster, L. Ó Náraigh and K. E. Pang, "cuPentBatchA batched pentadiagonal solver for NVIDIA GPUs", *Computer Physics Communications*, 2019, **241**, 113–121.

(313)   A. Abdelfattah, M. Baboulin, V. Dobrev, J. Dongarra, C. Earl, J. Falcou, A. Haidar, I. Karlin, T. Kolev, I. Masliah and S. Tomov, "High-performance Tensor Contractions for GPUs", *Procedia Computer Science*, 2016, **80**, 108–118.

(314)   B. Liu, C. Wen, A. D. Sarwate and M. M. Dehnavi, 2017 IEEE International Conference on Cluster Computing (CLUSTER), IEEE, Honolulu, HI, USA, 2017, pp. 47–57.

(315) J. E. Stone, D. Gohara and G. Shi, "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems", *Computing in Science & Engineering*, 2010, **12**, 66–73.

(316) G. Guennebaud, B. Jacob and Others, *Eigen v3*, http://eigen.tuxfamily.org, date accessed: 27/08/19.

(317) M Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015.

(318) R. Olivares-Amaya, M. A. Watson, R. G. Edgar, L. Vogt, Y. Shao and A. Aspuru-Guzik, "Accelerating Correlated Quantum Chemistry Calculations Using Graphical Processing Units and a Mixed Precision Matrix Multiplication Library", *Journal of Chemical Theory and Computation*, 2010, **6**, 135–144.

(319) V. P. Vysotskiy and L. S. Cederbaum, "Accurate Quantum Chemistry in Single Precision Arithmetic: Correlation Energy", *Journal of Chemical Theory and Computation*, 2011, **7**, 320–326.

(320) P. Pokhilko, E. Epifanovsky and A. I. Krylov, "Double Precision Is Not Needed for Many-Body Calculations: Emergent Conventional Wisdom", *Journal of Chemical Theory and Computation*, 2018, **14**, 4088–4096.

(321) E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton and J. Demmel, "A massively parallel tensor contraction framework for coupled-cluster computations", *Journal of Parallel and Distributed Computing*, 2014, **74**, 3176–3190.

(322) E. F. Valeev, *Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions*, http://libint.valeyev.net/, version 2.3.1, date accessed: 27/08/19.

(323) H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and M. Wang, *MOLPRO, version 2015.1, a package of ab initio programs*, Cardiff, UK, 2015.

(324) D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon", *J. Chem. Phys.*, 1993, **98**, 1358–1371.

(325) F. Weigend, A. Köhn and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations", *The Journal of Chemical Physics*, 2002, **116**, 3175–3183.

(326) *Correlation consistent basis set repository*, http://bit.ly/ccBasis, date accessed: 27/08/19.

(327) C. Hättig, "Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Corevalence and quintuple- basis sets for H to Ar and QZVPP basis sets for Li to Kr", *Physical Chemistry Chemical Physics*, 2005, **7**, 59–66.

(328) R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, "Ps14 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability", *Journal of Chemical Theory and Computation*, 2017, **13**, 3185–3197.

(329) D. Manna, M. K. Kesharwani, N. Sylvetsky and J. M. L. Martin, "Conventional and Explicitly Correlated ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets", *Journal of Chemical Theory and Computation*, 2017, **13**, 3136–3152.

(330) M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers", *SoftwareX*, 2015, **1-2**, 19–25.

(331) K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell Jr., "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields", *Journal of Computational Chemistry*, 2010, **31**, 671–690.

(332) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics*, 1983, **79**, 926–935.

(333) J. Řezáč, K. E. Riley and P. Hobza, "Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules", *Journal of Chemical Theory and Computation*, 2012, **8**, 4285–4292.

(334) P. Metrangolo, G. Resnati, T. Pilati, R. Liantonio and F. Meyer, "Engineering functional materials by halogen bonding", *Journal of Polymer Science Part A: Polymer Chemistry*, 2007, **45**, 1–15.

(335) A. Priimagi, G. Cavallo, P. Metrangolo and G. Resnati, "The Halogen Bond in the Design of Functional Supramolecular Materials: Recent Advances", *Accounts of Chemical Research*, 2013, **46**, 2686–2695.

(336) A. E. DePrince, M. R. Kennedy, B. G. Sumpter and C. D. Sherrill, "Density-fitted singles and doubles coupled cluster on graphics processing units", *Molecular Physics*, 2014, **112**, 844–852.

(337) P. Politzer, J. S. Murray and T. Clark, "Halogen bonding and other σ-hole interactions: a perspective.", *Physical Chemistry Chemical Physics*, 2013, **15**, 11178–89.

(338) A. Karpfen, J. Ladik, P. Russegger, P. Schuster and S. Suhai, "Hydrogen bonding in long chains of hydrogen fluoride and long chains and large clusters of water molecules", *Theoretica Chimica Acta*, 1974, **34**, 115–127.

(339) S. Hirata and S. Iwata, "Ab Initio Hartree-Fock and Density Functional Studies on the Structures and Vibrations of an Infinite Hydrogen Fluoride Polymer", *The Journal of Physical Chemistry A*, 1998, **102**, 8426–8436.

(340) S. Maheshwary, N. Patel, N. Sathyamurthy, A. D. Kulkarni and S. R. Gadre, "Structure and stability of water clusters $(H_2O)_n$, $n = 8$–20: An ab initio investigation", *Journal of Physical Chemistry A*, 2001, **105**, 10525–10537.

(341) E. Rudberg, E. H. Rubensson and P. Sałek, "Kohn-Sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage", *Journal of Chemical Theory and Computation*, 2011, **7**, 340–350.

(342) K. U. Lao and J. M. Herbert, "Energy Decomposition Analysis with a Stable Charge-Transfer Term for Interpreting Intermolecular Interactions", *Journal of Chemical Theory and Computation*, 2016, **12**, 2569–2582.

(343) R. A. Kendall, T. H. Dunning and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions", *The Journal of Chemical Physics*, 1992, **96**, 6796–6806.

(344) K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, "Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements", *The Journal of Chemical Physics*, 2003, **119**, 11113.

(345) K. A. Peterson, B. C. Shepler, D. Figgen and H. Stoll, "On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions", *The Journal of Physical Chemistry A*, 2006, **110**, 13877–13883.

(346) C. Hättig, D. P. Tew and B. Helmich, "Local explicitly correlated second- and third-order Møller-Plesset perturbation theory with pair natural orbitals", *The Journal of Chemical Physics*, 2012, **136**, 204105.

(347) T. Helgaker, in *Encyclopedia of Computational Chemistry*, ed. P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer and P. R. Schreiner, John Wiley & Sons, Ltd, Chichester, UK, 2002.

(348) L. M. Milne-Thomson, *The calculus of finite differences*, AMS Chelsea Pub, Providence, R.I, 2000.

(349) P. Salvador and M. M. Szczniak, "Counterpoise-corrected geometries and harmonic frequencies of N-body clusters: Application to $(HF)_n$ ($n = 3,4$)", *The Journal of Chemical Physics*, 2003, **118**, 537–549.

(350) J. A. Pople, R. Krishnan, H. B. Schlegel and J. S. Binkley, "Derivative studies in hartree-fock and møller-plesset theories", *International Journal of Quantum Chemistry*, 1979, **16**, 225–241.

(351) M. Head-Gordon and J. A. Pople, "A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations", *The Journal of Chemical Physics*, 1988, **89**, 5777–5786.

(352) T. Helgaker and P. Jørgensen, in *Advances in Quantum Chemistry*, Elsevier, 1988, vol. 19, pp. 183–245.

(353) P. Pulay, in *Advances in Chemical Physics*, ed. K. P. Lawley, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1987, pp. 241–286.

(354) R. P. Feynman, "Forces in Molecules", *Physical Review*, 1939, **56**, 340–343.

(355) *Hans Hellmann: Einführung in die Quantenchemie*, ed. D. Andrae, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

(356) J. Boström, F. Aquilante, T. B. Pedersen and R. Lindh, "Analytical Gradients of Hartree-Fock Exchange with Density Fitting Approximations", *Journal of Chemical Theory and Computation*, 2013, **9**, 204–212.

(357) U. Bozkaya, "Derivation of general analytic gradient expressions for density-fitted post-Hartree-Fock methods: An efficient implementation for the density-fitted second-order Møller–Plesset perturbation theory", *The Journal of Chemical Physics*, 2014, **141**, 124108.

(358) U. Bozkaya and C. D. Sherrill, "Analytic energy gradients for the coupled-cluster singles and doubles method with the density-fitting approximation", *The Journal of Chemical Physics*, 2016, **144**, 174103.

(359) J. Rekkedal, S. Coriani, M. F. Iozzi, A. M. Teale, T. Helgaker and T. B. Pedersen, "Communication: Analytic gradients in the random-phase approximation", *The Journal of Chemical Physics*, 2013, **139**, 081101.

(360) A. M. Burow, J. E. Bates, F. Furche and H. Eshuis, "Analytical First-Order Molecular Properties and Forces within the Adiabatic Connection Random Phase Approximation", *Journal of Chemical Theory and Computation*, 2014, **10**, 180–194.

(361) B. Mussard, P. G. Szalay and J. G. Ángyán, "Analytical Energy Gradients in Range-Separated Hybrid Density Functional Theory with Random Phase Approximation", *Journal of Chemical Theory and Computation*, 2014, **10**, 1968–1979.

(362) J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, New York, 1999.

(363) V. Bakken and T. Helgaker, "The efficient optimization of molecular geometries using redundant internal coordinates", *The Journal of Chemical Physics*, 2002, **117**, 9160–9174.

(364) A. K. Wilson, D. E. Woon, K. A. Peterson and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton", *The Journal of Chemical Physics*, 1999, **110**, 7667–7676.

(365) B. Brauer, M. K. Kesharwani, S. Kozuch and J. M. L. Martin, "The S66x8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory", *Physical Chemistry Chemical Physics*, 2016, **18**, 20905–20925.

(366) P.-O. Löwdin, "Quantum Theory of Many-Particle Systems. III. Extension of the Hartree-Fock Scheme to Include Degenerate Systems and Correlation Effects", *Physical Review*, 1955, **97**, 1509–1520.

(367) J. Olsen, "The CASSCF method: A perspective and commentary", *International Journal of Quantum Chemistry*, 2011, **111**, 3267–3272.

(368) W. Jiang, N. J. DeYonker and A. K. Wilson, "Multireference Character for 3d Transition-Metal-Containing Molecules", *Journal of Chemical Theory and Computation*, 2012, **8**, 460–468.

(369) J. Wang, S. Manivasagam and A. K. Wilson, "Multireference Character for 4d Transition Metal-Containing Molecules", *Journal of Chemical Theory and Computation*, 2015, **11**, 5865–5872.

(370) C. D. Sherrill and H. F. Schaefer, in *Advances in Quantum Chemistry*, Elsevier, 1999, vol. 34, pp. 143–269.

(371) J. Olsen, "Novel methods for configuration interaction and orbital optimization for wave functions containing non-orthogonal orbitals with applications to the chromium dimer and trimer", *The Journal of Chemical Physics*, 2015, **143**, 114102.

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

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

(374) J. Olsen, B. O. Roos, P. Jørgensen and H. J. A. Jensen, "Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces", *The Journal of Chemical Physics*, 1988, **89**, 2185–2192.

(375) P. A. Malmqvist, A. Rendell and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach", *The Journal of Physical Chemistry*, 1990, **94**, 5477–5482.

(376) A. Hibbert, C. Froese Fischer and M. Godefroid, "Non-orthogonal orbitals in MCHF or configuration interaction wave functions", *Computer Physics Communications*, 1988, **51**, 285–293.

(377) K. Hirao, H. Nakano, K. Nakayama and M. Dupuis, "A complete active space valence bond (CASVB) method", *The Journal of Chemical Physics*, 1996, **105**, 9227–9239.

(378) D. L. Cooper, T. Thorsteinsson and J. Gerratt, in *Advances in Quantum Chemistry*, Elsevier, 1998, vol. 32, pp. 51–67.

(379) N. Handy, "Multi-root configuration interaction calculations", *Chemical Physics Letters*, 1980, **74**, 280–283.

(380) P. Knowles and N. Handy, "A new determinant-based full configuration interaction method", *Chemical Physics Letters*, 1984, **111**, 315–321.

(381) E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices", *Journal of Computational Physics*, 1975, **17**, 87–94.

(382) B Liu, *The simultaneous expansion method for the iterative solution of several of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices*, tech. rep. LBL-8158, Lawrence Berkeley Laboratory, University of California, Berkeley, 1978.

(383) J. C. Slater, "The Theory of Complex Spectra", *Physical Review*, 1929, **34**, 1293–1322.

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

(385) H. B. Schlegel and J. J. W. McDouall, in *Computational Advances in Organic Chemistry: Molecular Structure and Reactivity*, ed. C. Ögretir and I. G. Csizmadia, Springer Netherlands, Dordrecht, 1991, pp. 167–185.

(386) D. A. Kreplin, P. J. Knowles and H.-J. Werner, "Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence", *The Journal of Chemical Physics*, 2019, **150**, 194106.

(387) H. Werner and P. J. Knowles, "A second order multiconfiguration SCF procedure with optimum convergence", *The Journal of Chemical Physics*, 1985, **82**, 5053–5063.

(388) N. Mataga and H. Miyasaka, in *Advances in Chemical Physics*, ed. I. Prigogine and S. A. Rice, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1999, pp. 431–496.

(389) R. A. Krueger and G. Blanquart, "Multireference exciplex binding energies: Basis set convergence and error", *International Journal of Quantum Chemistry*, 2019, **119**, e25819.

(390) K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski, "Second-order perturbation theory with a CASSCF reference function", *The Journal of Physical Chemistry*, 1990, **94**, 5483–5488.

(391) K. Andersson, P. Malmqvist and B. O. Roos, "Second-order perturbation theory with a complete active space self-consistent field reference function", *The Journal of Chemical Physics*, 1992, **96**, 1218–1226.

(392) R McWeeny, "The density matrix in many-electron quantum mechanics I. Generalized product functions. Factorization and physical interpretation of the density matrices", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 1959, **253**, 242–259.

(393) E. Davidson, *Reduced Density Matrices in Quantum Chemistry.* OCLC: 1044726220, Elsevier Science, Saint Louis, 1976.

(394) A. V. Luzanov, A. A. Sukhorukov and V. Umanskii, "Application of transition density matrix for analysis of excited states", *Theoretical and Experimental Chemistry*, 1976, **10**, 354–361.

(395) K. Truong and M. Ikura, "The use of FRET imaging microscopy to detect proteinprotein interactions and protein conformational changes in vivo", *Current Opinion in Structural Biology*, 2001, **11**, 573–578.

(396) S. Jana, M. Diez-Castellnou, E. R. Kay and C. Penedo, "Twin-FRET: A New Molecular Ruler for Biomolecules", *Biophysical Journal*, 2019, **116**, 565a.

(397) S. E. Bradforth, R. Jimenez, F. van Mourik, R. van Grondelle and G. R. Fleming, "Excitation Transfer in the Core Light-Harvesting Complex (LH-1) of Rhodobacter sphaeroides: An Ultrafast Fluorescence Depolarization and Annihilation Study", *The Journal of Physical Chemistry*, 1995, **99**, 16179–16191.

(398) J. W. Riley, B. Wang, J. L. Woodhouse, M. Assmann, G. A. Worth and H. H. Fielding, "Unravelling the Role of an Aqueous Environment on the Electronic Structure and Ionization of Phenol Using Photoelectron Spectroscopy", *The Journal of Physical Chemistry Letters*, 2018, **9**, 678–682.

(399) P. R. Tentscher, R. Seidel, B. Winter, J. J. Guerard and J. S. Arey, "Exploring the Aqueous Vertical Ionization of Organic Molecules by Molecular Simulation and Liquid Microjet Photoelectron Spectroscopy", *The Journal of Physical Chemistry B*, 2015, **119**, 238–256.

(400) P. G. Szalay and R. J. Bartlett, "Multi-reference averaged quadratic coupled-cluster method: a size-extensive modification of multi-reference CI", *Chemical Physics Letters*, 1993, **214**, 481–488.

(401) J. F. Stanton and R. J. Bartlett, "The equation of motion coupledcluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties", *The Journal of Chemical Physics*, 1993, **98**, 7029–7039.

(402) M. Nooijen and R. J. Bartlett, "A new method for excited states: Similarity transformed equation-of-motion coupled-cluster theory", *The Journal of Chemical Physics*, 1997, **106**, 6441–6448.

(403) T. C. Berkelbach, "Communication: Random-phase approximation excitation energies from approximate equation-of-motion coupled-cluster doubles", *The Journal of Chemical Physics*, 2018, **149**, 041103.

(404) M. F. Lange and T. C. Berkelbach, "On the Relation between Equation-of-Motion Coupled-Cluster Theory and the *GW* Approximation", *Journal of Chemical Theory and Computation*, 2018, **14**, 4224–4236.

(405) W. Arbelo-González, R. Crespo-Otero and M. Barbatti, "Steady and Time-Resolved Photoelectron Spectra Based on Nuclear Ensembles", *Journal of Chemical Theory and Computation*, 2016, **12**, 5037–5049.

(406) M. C. Oana and A. I. Krylov, "Dyson orbitals for ionization from the ground and electronically excited states within equation-of-motion coupled-cluster formalism: Theory, implementation, and examples", *The Journal of Chemical Physics*, 2007, **127**, 234106.

(407) G. Torres-Vega and J. H. Frederick, "Quantum mechanics in phase space: New approaches to the correspondence principle", *The Journal of Chemical Physics*, 1990, **93**, 8862–8874.

(408) K. Vanommeslaeghe and A. D. MacKerell, "Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing", *Journal of Chemical Information and Modeling*, 2012, **52**, 3144–3154.

(409) K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges", *Journal of Chemical Information and Modeling*, 2012, **52**, 3155–3168.

(410) M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, non-polarizable potential functions", *The Journal of Chemical Physics*, 2000, **112**, 8910–8922.

(411) K. Sargsyan, C. Grauffel and C. Lim, "How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations", *Journal of Chemical Theory and Computation*, 2017, **13**, 1518–1524.

(412) A. Puglisi, T. Giovannini, L. Antonov and C. Cappelli, "Interplay between conformational and solvent effects in UV-visible absorption spectra: curcumin tautomers as a case study", *Physical Chemistry Chemical Physics*, 2019, **21**, 15504–15514.

(413) M. Barbatti, M. Ruckenbauer, F. Plasser, J. Pittner, G. Granucci, M. Persico and H. Lischka, "Newton-X: a surface-hopping program for nonadiabatic molecular dynamics: Newton-X", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 26–33.

(414) G. Schmitz and C. Hättig, "Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12*) using Laplace transformation techniques", *The Journal of Chemical Physics*, 2016, **145**, 234107.

(415)   J. Lang, J. Brabec, M. Saitow, J. Pittner, F. Neese and O. Demel, "Perturbative triples correction to domain-based local pair natural orbital variants of Mukherjee's state specific coupled cluster method", *Physical Chemistry Chemical Physics*, 2019, **21**, 5022–5038.

(416)   S. Schweizer, J. Kussmann, B. Doser and C. Ochsenfeld, "Linear-scaling Cholesky decomposition", *Journal of Computational Chemistry*, 2008, **29**, 1004–1010.

(417)   F. Colmenero, C. Pérez del Valle and C. Valdemoro, "Approximating q -order reduced density matrices in terms of the lower-order ones. I. General relations", *Physical Review A*, 1993, **47**, 971–978.

(418)   P. Grinfeld, *Introduction to Tensor Analysis and the Calculus of Moving Surfaces*, Springer-Verlag New York, 1st ed., 2013.

(419)   Z. Ahsan, *Tensors: Mathematics of Differential Geometry and Relativity*, PHI Learning, 1st ed., 2015.