Compact Sparse Coulomb Integrals using a Range-Separated Potential

by
Michael James Lecours

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Trygve Helgaker University of Oslo, Oslo, Norway
Supervisor:	Dr. Marcel Nooijen University of Waterloo, Waterloo, ON, Canada
Supervisor:	Dr. W. Scott Hopkins University of Waterloo, Waterloo, ON, Canada
Internal Member:	Dr. Pierre-Nicholas Roy University of Waterloo, Waterloo, ON, Canada
Internal Member:	Dr. German Sciaini University of Waterloo, Waterloo, ON, Canada
Internal-External Member:	Dr. Roger Melko University of Waterloo, Waterloo, ON, Canada

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The efficient calculation of so-called two-electron integrals is an important component for electronic structure calculations on large molecules and periodic systems at both mean field and post-HF, correlated, levels. In this thesis, a new and fairly complicated representation of the Coulomb interaction is presented. The Coulomb potential is partitioned into short and long-range parts. The short-range interactions are treated analytically using conventional density fitting methods. The long-range interactions are treated numerically through either a Fourier transform in spherical coordinates or through a Cartesian multipole expansion. The Fourier transform is used for intermediate distances, while multipole expansions (up to octupole) are used for longer range, with a switching algorithm to decide between the two. In this range-separated representation, the corresponding two-electron Coulomb integrals can be calculated efficiently and the amount of data scales linearly with respect to system size. Hartree-Fock theory is used as an extensive test of the range-separated method, but the same building blocks can be used in correlated calculations like Second-order Møller-Plesset perturbation theory and (Cluster in Molecule) type Coupled Cluster calculations.

Acknowledgements

Firstly, I would like to acknowledge both of my supervisors as completion of this Ph.D. would not have been possible without the support and guidance from both of them. I would like to thank Dr. W. Scott Hopkins for providing me with the opportunity to engage in interesting research as early as my undergraduate degree. My research experience in the Hopkins lab was predominately experimental research. These experiences include the many CLIO trips and the Waterloo-Bordeaux collaboration, both of which were fruitful experiences. I am grateful for these opportunities as they helped shape the scientist I am today but I am most grateful for the support that Dr. Hopkins has given me as my research interests drifted more towards the theoretical side of physical chemistry. I would like to thank Dr. Marcel Nooijen, who has been my Ph.D. supervisor throughout this thesis work. Dr. Nooijen has been an inspiration and I appreciate his honesty, dedication and support he has shown me throughout my time in the Nooijen group. I would not have finished this thesis without his guidance. I truly grateful to have had two amazing supervisors guide me through this chapter of my life.

I would also like to acknowledge Dr. Kevin Bishop, Dr. Dmitri Iouchtchenko, Neil Raymond and Songhao Bao for their friendship and valuable support. I would like to thank Dr. Patrick Carr for the years of friendship and support he has given as we've worked through our Ph.D.'s. I would also like to thank the many colleagues in the Hopkins, McMahon, Gabelica, Roy and Nooijen groups who have made my graduate degree a wonderful experience. I would like to thank my examining committee for guidance they have shown.

Finally, I'd like to thank my family for their love and support they have given me throughout my degree. Most importantly I would like to thank my almost wife Esmeralda, she has supported me through a pulmonary embolism, a global pandemic and the writing of this thesis. I would not be who I am today without her.

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

AO atomic orbital. 2, 9, 39, 47, 79, 83, 84, 88, 93, 94, 103, 108

CC Coupled Cluster. 1

CFMM continuous fast multipole method. 5, 110

CIM cluster in molecules. 2, 47, 111

COO coordinate. 47

CSC compressed sparse column. 47

CSR compressed sparse row. 47

DF density fitting. 5, 9, 38, 80, 84, 89

DFT Density Functional Theory. 1

DIIS direct inversion of the iterative subspace. 4, 95

DLPNO domain-based local pair natural orbitals. 1, 47

EOMCC Equation of Motion CC. 2

FMM fast multipole method. 5, 6

FTAO Fourier transform atomic orbitals. 58

FTC Fourier-transform Coulomb. 5, 110

GPU graphic processing units. 6

HF Hartree-Fock. 4, 10, 42

LinK linear-in-K. 5

LO local orbitals. 40, 47, 80, 88

MO molecular orbital. 79

PySCF Python-based simulations of chemistry framework. 45

SCF self-consistent field. 4, 43, 79

Chapter 1

Introduction

1.1 Motivation

In this thesis, a report on an alternative implementation of Hartree-Fock calculations is presented. The implementation is based on a new and fairly complicated representation of the Coulomb interaction and the corresponding two-electron integrals. There are a number of excellent implementations of Hartree-Fock and Density Functional Theory (DFT) in the community and the goal of this work is not so much to add another Hartree-Fock code. The Coulomb integrals are needed in all post-Hartree Fock calculations for both ground and excited electronic states. The approach that will be described in this thesis allows for highly efficient calculations of selected Coulomb integrals which is an exceedingly important component of electronic structure calculations for large molecules and periodic solids.

Currently there are basically two main paradigms to perform accurate wave function based electronic structure calculations for large molecules. In all of these approaches Coupled Cluster (CC)[1–4] and low-order perturbation theory (MP2)[5,6] is the main paradigm as it has proven to be highly accurate and efficient for smaller systems, when using canonical, delocalized molecular orbitals. A first, common, step for larger molecules is to localize the occupied orbitals that result from a Hartree-Fock calculation. The various approaches diverge at this point.

In the so-called domain-based local pair natural orbitals (DLPNO) approaches[7–18] the primary idea is to define a small set of virtual orbitals associated with each pair of localized occupied orbitals (LMO's). In practice, one might use 5 'pair natural orbitals'[19] (or PNO's) when the orbitals are distant, and up to about 40 orbitals when the LMO's are close. This is an enormous reduction of the complete virtual space (1000's of virtuals). To achieve sufficient accuracy second order perturbation theory (MP2) is required for all distant pairs, while CC is used for close pairs. As a result, the number of double excitation amplitudes is small enough and can

be solved for. Current implementations scale (nearly) linear with the size of molecules. There are a number of disadvantages, however, that are related to the fact that all virtual orbitals are very different depending on the pair of occupied orbitals. This leads to a very large number of atomic orbital (AO) and LMO/PNO integrals and integral transformations. The CC equations are significantly more complicated due to the presence of numerous overlap integrals between PNO's corresponding to different pairs. The DLPNO-CC equations refer to the complete molecule and the solution of these equations is quite hard to parallelize efficiently. The main drawback from our perspective is that one has to develop equations and computer code for each approach, e.g. CC for ground states, Equation of Motion CC (EOMCC)[20] and similarity transformed EOMCC[21] for excited states, multireference approaches and so on. Each of these implementations is very time-consuming and not so interesting in itself (technical).

The second paradigm, denoted as a cluster in molecules (CIM) calculations[22–39] divides the full calculation into a (very) large number of smaller calculations that each require a subset of orbitals. The scheme is most easily explained for the ground state, based on the expression for the (closed-shell) correlation energy:

$$E_{corr} = \sum_{i}^{Nocc} \Delta E_i \tag{1.1}$$

$$\Delta E_i = \frac{1}{2} \sum_{j,a,b \in I} V_{ij}^{ab} \tau_{ab}^{ij} \tag{1.2}$$

There is a sum over localized occupied orbitals i, and associated with this central orbital there is a limited set of localized occupied orbitals j, and localized virtual orbitals a,b, that are in the vicinity of i, optimized for maximum effect, which define an orbital domain I. The term V_{ij}^{ab} is the anti-symmetrized two-electron integrals and $\tau_{ab}^{ij} = t_{ij}^{ab}$ for MP2 or $\tau_{ab}^{ij} = t_{ij}^{ab} + t_a^i t_b^j - t_a^j t_b^i$ for CC. The big difference between CIM and PNO approaches is that each calculation corresponding to an orbital domain I is completely independent and is just a canonical calculation using a set of orthonormal orbitals. The equations are the same and can be run using standard implementations. In a CIM calculation one would typically perform a Hartree-Fock and MP2 calculation for the complete molecule, but the highest layer of calculation (coupled cluster) is divided up in small independent calculations that only require the Hamiltonian integrals over a suitable small subset of orbitals,

irrespective of the total size of the molecule. The orbital domains in CIM calculations are on the order of 200-300 orbitals and canonical calculations with this number of orbitals are routine. The advantages of the CIM paradigm are clear. The hardest part of the calculation uses existing canonical codes and each of them can be run independently without any communication, leading to trivial parallelization. The demanding part of an implementation is to carry out the numerous integral transformations (one for each occupied orbital), and to make judicious use of potential sparsity during the integral recalculation and transformation. The long-range nature of the Coulomb potential provides a challenge for efficient screening techniques. Currently the method is only used for ground state energy calculations. The approaches can be generalized for excited states, especially for methods that have been developed in the Nooijen lab (STEOM-CCSD and MREOM-CCSD), which require a dressed Hamiltonian in a limited (active) orbital space.

Both LPNO and CIM approaches can benefit from the representation of integrals which is the main subject of the thesis, and this is the main reason for our interest. The implementation of Hartree-Fock is of some interest but it is more important as an extensive test of the approach. The implementation of Laplace MP2 is in progress and our approach is potentially quite fruitful here[40].

There is one more challenge in electronic structure calculations that can benefit from the integral representation in this thesis and this concerns Gaussian basis set calculations for periodic solids[41–46] (in 1, 2, or 3 dimensions). These calculations are not nearly as well developed as calculations for molecules, and even Hartree-Fock calculations are not very efficient. Most of the current routine calculations for solids use DFT and plane wave basis sets. Such calculations have limited accuracy and the availability of efficient Quantum Chemistry machinery would constitute a real advance in the field. This research project started out as a new approach for Hartree-Fock calculations and representations of Coulomb integrals for solids. The partitioning of the Coulomb potential into a short-range and long-range part is very beneficial for solids. The short-range integrals can easily be deployed using density fitting 3-center integrals. Most of the troubles for periodic systems arise from long-range interactions. They can be dealt with using a Fourier technique in spherical coordinates to cancel the g^{-2} singularity in the Fourier transform of r_{ij}^{-1}. Beyond Hartree-Fock, a CIM approach can be adopted, which would yield a localized approach to electron correlation. Unfortunately, this interesting approach to solids

will not be discussed further as (too) many challenges needed to be overcome to get things to work for molecules. The extension to solids will be left for future investigations.

In the remaining part of this introduction a thorough review of the current status of Hartree-Fock calculations is presented, and then a concise overview of the scope of this thesis is provided.

1.2 Overview of Linear Scaling Hartree-Fock Calculations

A Hartree-Fock (HF) calculation is often the first step in many electronic structure methods as it provides a basis for wavefunction-based electron correlation methods such as CC and MP2. The HF orbitals, which are the canonical molecular orbitals, are optimized in an iterative procedure called the self-consistent field (SCF) method. Briefly, the SCF algorithm begins with a guess at the one-particle density matrix which is used to calculate the Fock operator. The Fock operator is then involved in a matrix-eigenvalue problem which is conventionally solved through diagonalization and a better density matrix is obtained. This process is repeated until the calculation converges. The two most expensive steps in the canonical SCF algorithm are the formation of the two-electron contribution to the Fock matrix and the diagonalization of the Fock matrix. This discussion will focus on the construction of the Fock matrix, the two-electron direct and exchange contributions in particular, as this is related to the work in this thesis. The SCF algorithm can be accelerated by directly solving for the one-particle density matrix using density matrix based SCF methods[47–58] and the convergence of the method can be improved with accelerators such as direct inversion of the iterative subspace (DIIS)[59–62].

In the 1980's it was suggested by Almlöf[63] that the two-electron integrals should be recomputed at every SCF iteration to avoid the storage issues, this lead to the direct-SCF method which is staple in modern SCF codes. Integral pre-screening is essential for direct SCF methods as negligible contributions to the Fock matrix can be recognized and avoided, which reduces the algorithm to quadratic scaling. The Schwartz-inequality is the simplest of the screening methods[64], it provides an estimate on the integral that is an upper bound but not necessary a tight bound. Over the decades more efficient screening methods have been developed which are more efficient and use tighter bounds for screening[65]. These screening methods are also highly advantageous for correlated calculations[66]. Almlöf's second suggestion was to calculate the direct Coulomb

and exchange contributions individually using the most efficient methods for each[67,68]. This is counter intuitive at first because the direct (J) and exchange (K) contributions involve the same quantities, the density matrix, $D_{\gamma\beta}$, and the two-electron repulsion integrals, $(\alpha\delta|\gamma\beta)$.

$$J_{\alpha\beta} = \sum_{\gamma\delta} (\alpha\beta|\gamma\delta) D_{\gamma\delta} \tag{1.3}$$

$$K_{\alpha\beta} = \sum_{\gamma\delta} (\alpha\delta|\gamma\beta) D_{\gamma\delta} \tag{1.4}$$

The Coulomb contribution can be formed via fast multipole method (FMM)[69–72], the first linear scaling method to evaluate the Coulomb potential is the continuous fast multipole method (CFMM) by White *et. al.*[70]. In this method Coulomb contribution is partitioned into near field and far-field contributions where the near-field contributions are calculated through conventional methods and the far-field contributions are treated with a multipole expansion. The CFMM is efficient for the treatment of far-field interactions but the near-field contributions, which scale linearly, dominate the computational time. The J-engine method[73] greatly improves the efficiency of calculating the near-field contributions by directly summing the density matrix into the underlying Gaussian integral. This method avoids the calculation of two-electron integral intermediates. Similar techniques have also been developed by Ahmadi and Almlo$\ddot{o}$f[68] and Neese[74] which rely on the resolution of the identity, commonly referred to as density fitting (DF). In this thesis DF is essential for the treatment of our integrals. An alternative scheme for calculating J is through a discrete Fourier transform of the density, called the Fourier-transform Coulomb (FTC) method[75–78]. This method scales linearly with respect to system size and uses plane waves.

The natural scaling of the exchange contribution is almost linear for local electronic structures, this is partially due to the fact the at exchange phenomenon is predominately a local phenomenon. However, the exchange is still the more expensive of the two contributions. Improvements have been made to the calculation of the exchange contribution, the first linear scaling exchange algorithm is ONX by Schwegler *et. al.*[79] which utilizes a sophisticated loop structure and preordered integral estimates. The draw back to this method was that it lacked permutational symmetry in the integral calculation and contractions. The linear-in-K (LinK) method by Ochsenfeld *et. al.*[79,80] exploits the permutational symmetry and is linear scaling. The chain-of-spheres algorithm by Neese *et. al.*[81] is a semi-numerical algorithm which reduces the scaling with

respect to the size of basis set and is advantageous when large l-quantum number basis functions dominate the overall computational time. In recent years efforts have been made to implement linear scaling exchange algorithms for use with graphic processing units (GPU)[82–85].

1.3 Overview of Range-Separated Coulomb Interactions

The Coulomb potential is an inverse pairwise potential that decays slowly as a function of $\mathbf{r}_{12}$.

$$V_{Coul}(\mathbf{r}_{12}) = \frac{1}{|\mathbf{r}_{12}|} \tag{1.5}$$

This slow decay is partially why large scale calculations are so difficult, long-range interactions scale quadratically with respect to system size. Coulomb interactions can be evaluated linearly with respect to system size through the FMM[69,71]. In the FMM the system space is divided into smaller and smaller boxes and the well-separatedness of the boxes is used to partition the potential into near field (NF) and far field (FF) contributions where the FF contributions are treated through multipole expansions. The range-separated potentials in this thesis do not use this method, instead Ewald summation techniques[86,87] are used to explicitly partition the potential into short and long-range regions.

$$V_{Coul}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{lr}(\mathbf{r}_{12}) \tag{1.6}$$

$$V_{sr}(\mathbf{r}_{12}) = \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} \tag{1.7}$$

$$V_{lr}(\mathbf{r}_{12}) = \frac{1}{|\mathbf{r}_{12}|} - V_{sr}(\mathbf{r}_{12}) \tag{1.8}$$

The parameter α, called the range-separation parameter, is the control which governs the extent of the short-range region and the onset of the long-range region. In this form the long-range potential is not purely long range because $V_{lr}(\mathbf{r}_{12})$ is not zero at $\mathbf{r}_{12}=0$. To insist that the long-range potential is zero at $\mathbf{r}_{12}=0$, a Gaussian is added to $V_{sr}(\mathbf{r}_{12})$ and subtracted from $V_{lr}(\mathbf{r}_{12})$.

$$V_{sr}(\mathbf{r}_{12}) = \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_0 e^{-\gamma|\mathbf{r}_{12}|^2} \tag{1.9}$$

The Gaussian parameters X_0 and γ are chosen such that $V_{lr}(0)$ is zero and as flat as possible around $\mathbf{r}_{12}=0$. Both X_0 and γ are related to α such that α is the only controlling parameter for partitioning the potential. The value of α ranges from 0.1 to 0.8, the smaller the value of α the further the short-range region extends.

Short-range interactions scale linearly with respect to system size because the short-range potential quickly decays to zero as a function of $\mathbf{r}_{12}$. These interactions can be treated analytically using conventional methods.

The long-range potential is more complicated as it still exhibits the slow $\frac{1}{|\mathbf{r}_{12}|}$ decay of the Coulomb potential. The long-range potential can be evaluated through a multipole method or it can evaluated numerically through a Fourier transform method. The Fourier transform of $V_{lr}(\mathbf{r}_{12})$ is given by

$$V_{lr}(\mathbf{r}_{12}) = \int dg^3 \frac{1}{|\mathbf{g}|^2} \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{1.10}$$

$$\eta(\mathbf{g}) = \frac{4\pi}{(2\pi)^3} e^{-\omega|\mathbf{g}|^2} - \frac{|\mathbf{g}|^2}{(2\pi)^3(2\gamma)^{3/2}} e^{-\frac{1}{4\gamma}|\mathbf{g}|^2} \tag{1.11}$$

The Fourier transform introduces a singularity at $\mathbf{g}=0$. The singularity can be cancelled out by choosing to evaluate the integral in spherical coordinates.

$$V_{lr}(\mathbf{r}_{12}) = \int d\Omega \int dg_r \eta(g_r) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{1.12}$$

Equation 1.12 is evaluated numerically using points and weights. The grid points are comprised of an angular and radial grid. Lebedev[88] quadrature is used for the angular grid and a simple equidistant grid is used for the radial grid, other radial quadratures[89–92] can be used if desired. The numerical weights represent the function $\eta(g_r)$ which rapidly decays as a function of g_r limiting the extent of the radial grid. The numerical grid does not need to increase with increasing system sizes.

Evaluating the long-range potential using the Fourier method works well for short/medium-range interactions. The Fourier method does not work well for very far interactions because of the term $e^{i\mathbf{g}\cdot\mathbf{r}_{12}}$, which is essentially a rapid oscillator at large enough $\mathbf{r}_{12}$. The multipole

method exhibits the opposite behaviour, it is most accurate for interactions at large $\mathbf{r}_{12}$ while at smaller $\mathbf{r}_{12}$ high-level multipole expansions are needed to maintain accuracy. The preferred method for evaluating $V_{lr}(\mathbf{r}_{12})$ is to use the Fourier method for small $\mathbf{r}_{12}$ then switch to the multipole method for large $\mathbf{r}_{12}$.

The singularity in the Fourier transform of $V_{lr}(\mathbf{r}_{12})$ can be explicitly removed through a regularization procedure introduced in chapter 2. The regularization is performed by introducing a grid of fixed points called gauge centres $\mathbf{R}_i$ which are located near the coordinates $\mathbf{r}_i$. This definition of the gauge centres is ambiguous and a more rigorous definition is provided when atomic orbitals are introduced. The regularization procedure leads to a three part potential where the long-range potential is split into a regularized long-range potential, V_{lr}^R, and a one-body potential V_{1body}.

$$V_{Coul}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) + V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) \tag{1.13}$$

$$V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = \int d\Omega \int dg_r \eta(g_r)[e^{-i\mathbf{g}\cdot\mathbf{r}_1} - e^{-i\mathbf{g}\cdot\mathbf{R}_1}][e^{i\mathbf{g}\cdot\mathbf{r}_2} - e^{i\mathbf{g}\cdot\mathbf{R}_2}] \tag{1.14}$$

$$V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = V_{lr}(|\mathbf{r}_2 - \mathbf{R}_1|) + V_{lr}(|\mathbf{R}_2 - \mathbf{r}_1|) - V_{lr}(|\mathbf{R}_2 - \mathbf{R}_1|) \tag{1.15}$$

The one-body potential can be evaluated analytically while the regularized long-range potential is evaluated numerically. The one-body potential is the dominant contribution to the long-range potential and it is inexpensive to calculate. The regularized long-range potential, V_{lr}^R, is the significantly smaller contribution, it makes up about 1% of the total long-range potential. It is also the only contribution that needs to be evaluated numerically. The key advantage to the regularized method is that the most expensive term to calculate is also the smallest. Therefore one can choose to either ignore V_{lr}^R completely or to evaluate it using a small numerical grid without much loss in accuracy. The regularization procedure is general for any smooth radial potential, it is not limited to the potential described in equation 1.6.

Chapter 2 entitled "Range-separated Coulomb potentials" is devoted to the derivations and analysis of the bare range-separated potentials described above. The methods defined in chapter 2 establish the ground work needed to introduce the range-separated two-electron Coulomb integrals. Chapter 3 is the core of this thesis and a thorough investigation of the two-electron integrals is presented in this chapter with emphasis on data scaling and accuracy.

Using the potential defined in equation 1.6 the two-electron repulsion integrals are partitioned into short and long-range integrals.

$$(\alpha\beta|\gamma\delta) = (\alpha\beta|\gamma\delta)_{sr} + (\alpha\beta|\gamma\delta)_{lr} \tag{1.16}$$

The short-range integrals are well behaved and can be treated with conventional methods such as DF. The long-range integrals are treated in a similar manner as the long-range potential described above, using a numerical Fourier transform for short/medium-range interactions and switching to the multipole method for long-range interactions. The integrals evaluated through each method; DF, Fourier and multipole can all be represented compactly using three centre integrals.

$$\begin{aligned}(\alpha\beta|\gamma\delta) = &\sum_{xy}(\alpha\beta|x)_{sr}M^{-1}_{xy}(y|\gamma\delta)_{sr} \\ &+\sum_{\mathbf{g}}(\alpha\beta|\mathbf{g})\Theta(\mathbf{R}_{12})\eta(\mathbf{g})(\gamma\delta|\mathbf{g})^* \\ &+\sum_{mn}(\alpha\beta|m)f^{mn}(\mathbf{R}_{12})\bar{\Theta}(\mathbf{R}_{12})(\gamma\delta|n)\end{aligned} \tag{1.17}$$

The first line in equation 1.17 describes the short-range DF integrals, which uses the auxiliary fitting basis x and the inverse short-range Coulomb metric. The short-range three centre integrals $(\alpha\beta|x)_{sr}$ are sparse because there is a limited number of AO orbitals, $\alpha\beta$, in the vicinity of the auxiliary orbital x. The amount of significant integrals scale linearly with respect to system size because of this sparsity, which is not present in full-range DF integrals because those integrals use the Coulomb potential.

The second and third line in equation 1.17 represent the long-range interactions. The functions $\Theta(\mathbf{R}_{12})$ and $\bar{\Theta}(\mathbf{R}_{12})$ are switching functions used to determine which method is used for a specific long-range integral. The Fourier integrals $(\alpha\beta|\mathbf{g})$ are complex, sparse and also scale linearly with respect to system size. They scale linearly with respect to system size because the size of the numerical grid does not need to increase with respect to system size because $\eta(\mathbf{g})$ decays rapidly as a function of $\mathbf{g}$. The multipole integrals $(\alpha\beta|m)$ are Cartesian multipole moment integrals where m represents the multipole level and function $f^{mn}(\mathbf{R}_{12})$ denotes analytic derivatives of

the long-range potential. Correctly switching between the Fourier and multipole methods is one of the most challenging aspects to this project and is discussed thoroughly in chapter 3.

The regularization procedure used for the long-range potential is expanded in chapter 3, from this procedure regularized Fourier and regularized multipole methods are derived. The two-electron one-body integrals arise from this procedure.

$$(\alpha\beta|\gamma\delta)_{1body} = S_{\alpha\beta}V_{\gamma\delta}(\mathbf{R}_\alpha) + V_{\alpha\beta}(\mathbf{R}_\gamma)S_{\gamma\delta} - S_{\alpha\beta}S_{\gamma\delta}V(\mathbf{R}_\alpha - \mathbf{R}_\gamma) \tag{1.18}$$

Here $S_{\alpha\beta}$ is the overlap matrix and $V_{\gamma\delta}(\mathbf{R}_\alpha)$ are nuclear-electron like attraction integrals with a charge centred on $\mathbf{R}_\alpha$. $\mathbf{R}_\alpha$ is a gauge centre used in the regularization and an unambiguous definition of these centres is explained in chapter 3. The one-body contribution is evaluated analytically and these terms do not scale linearly with respect to system size but they are one-body contributions and will not become a computational bottleneck until very large systems are a used. The linear scaling of the integrals is due to the sparsity in the three centre representations, therefore integral pre-screening is essential to the evaluation of these integrals.

The accuracy of the methods is reported using the accuracy of the Hartree-Fock direct and exchange matrices because the four centre integrals do not fit in memory and are never fully assembled in practice. The accuracy of the exchange and direct term exhibit different behaviour with respect to the range-separation parameter α. The exchange is only assembled with the Fourier method, it was found that a large α and a medium numerical grid were sufficient to accurately represent the exchange. The long-range direct term is more complicated and requires switching between methods. The accuracy of the direct term is underwhelming, it can only be assembled accurately if a small α and a sizeable numerical grid are used. The smaller the value of α the further the short-range region extends implying the DF integrals are less sparse and this increases the short-range linear scaling prefactor.

Chapter 4 describes the range-separated HF method, where algorithms for assembling the long-range direct and exchange contributions are outlined. Approximate SCF schemes are also discussed in chapter 4 using the regularized representation.

Chapter 5 entitled "Conclusions and Outlook" provides the reader with insight into the future

direction of the project including how these range-separated Coulomb integrals can be used in correlated calculations.

Chapter 2

Range-separated Coulomb Potential

2.1 Introduction

The Coulomb potential is a pairwise potential representing the interaction between charged particles. The direct method for evaluating Coulombic interactions entering a molecular Hamiltonian scales as $\mathrm{O}(N^2)$ for nuclear-nuclear interactions and as $\mathrm{O}(N^4)$ for electron-electron interactions. Owing to the slow $\frac{1}{|\mathbf{r}_{12}|}$ decay and the lack of sparsity in long-range interactions of the coulomb potential, the evaluation of these interactions is delicate. In an effort to more efficiently treat Coulomb interactions in large systems, a range-separated Coulomb potential is explored.

$$V_{Coul}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{lr}(\mathbf{r}_{12}) \tag{2.1}$$

In this chapter, **boldface** is used to emphasizes vector quantities while non-bold typeface is used to represent scalars. The potentials are written as $V(\mathbf{r}_{12})$ even if they only depend on the magnitude of $\mathbf{r}_{12}$, written $|\mathbf{r}_{12}|$. The reason is that many of the representations and partitionings used in this chapter may depend on $\mathbf{r}_{12}$ and we will not reduce the expressions to a radial form as this is not how they would be used in practice when integrals over orbital products are introduced later in the thesis.

The potential is partitioned into short and long-range regions. The short-range region is well behaved because any long-range interactions are zero which leads to sparsity resulting in natural low order scaling with respect to system size. The long-range region still exhibits the challenging $\frac{1}{|\mathbf{r}_{12}|}$ decay however, it can be approximated accurately using the multipole expansion.

Ewald partitioning[86,87] is used to separate the potential, a Gaussian is subtracted from the long-range partition with optimal parameters to ensure that V_{lr} is zero at $\mathbf{r}_{12}=0$ and that it

is flattest near small $\mathbf{r}_{12}$.

$$V_{sr}(\mathbf{r}_{12}) = \frac{\text{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_0 e^{-\gamma|\mathbf{r}_{12}|^2} \tag{2.2}$$

$$V_{lr}(\mathbf{r}_{12}) = \frac{1}{|\mathbf{r}_{12}|} - V_{sr}(|\mathbf{r}_{12}|) \tag{2.3}$$

The parameter α is the range separation parameter which controls the extent of the short-range potential, while the parameters X_0 and γ are determined by α. The larger the value of α the shorter the short-range potential. For visual aid, a range-separated potential with $\alpha = 0.6$ is shown in figure 2.1.

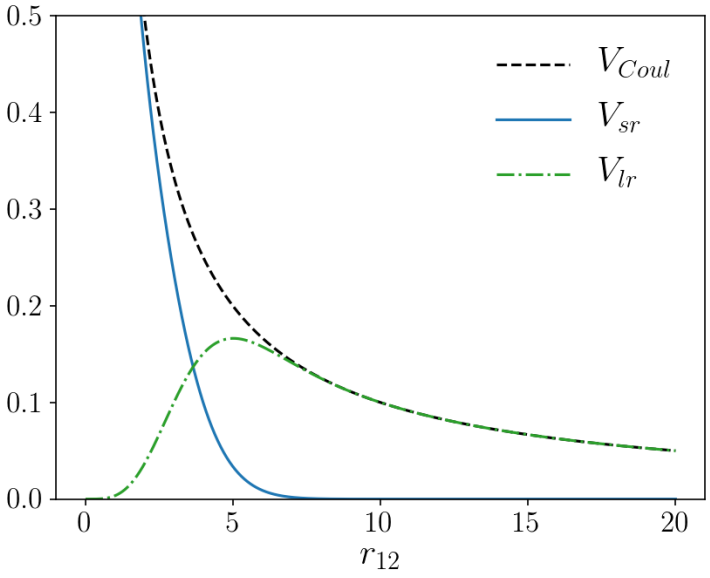

Figure 2.1: The Coulomb potential partitioned into long and short-range regions with $\alpha = 0.6$

The simplest method to evaluate the range-separated potential is to evaluate the short-range potential analytically and to approximate the long-range region of the potential with the multipole expansion. The multipole expansion is most accurate in the long-range region of V_{lr}, however, the short-range region of V_{lr} requires very high multipole levels to obtain accurate results. Figure 2.2 depicts V_{lr} approximated at increasing multipole levels, the analytical V_{lr} is shown for reference. It is clear that near $\mathbf{r}_{12} = 0$ the multipole expansion is inaccurate, but at large r_{12} the approximation is accurate even for the lowest level of expansion. It is possible to increase the overall accuracy of the short-range region of V_{lr} by choosing a small α. In this case however, V_{sr} has a large extent which is not ideal.

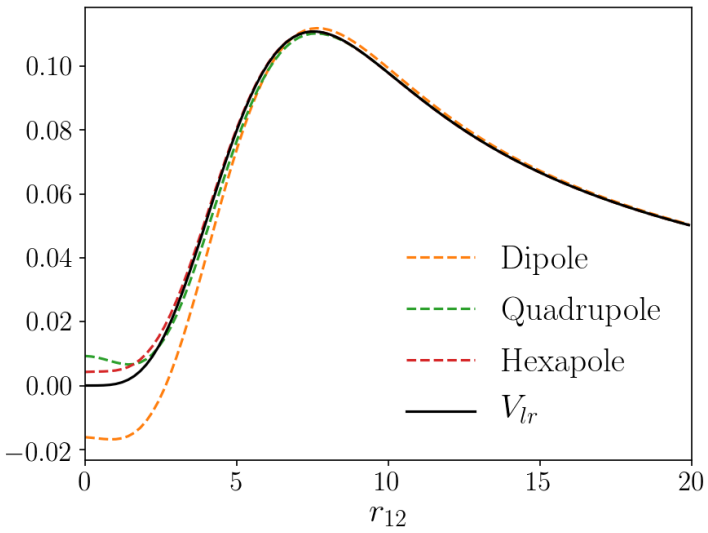

Figure 2.2: The long-range potential approximated with increasing orders of the multpiole expansion

Alternatively one could evaluate V_{lr} numerically through a Fourier transform.

$$V_{lr}(\mathbf{r}_{12}) = \int dg^3 \frac{1}{|\mathbf{g}|^2} \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.4}$$

$$\eta(\mathbf{g}) = \frac{4\pi}{(2\pi)^3} e^{-\omega|\mathbf{g}|^2} - \frac{|\mathbf{g}|^2}{(2\pi)^3 (2\gamma)^{3/2}} e^{-\frac{1}{4\gamma}|\mathbf{g}|^2} \tag{2.5}$$

The Fourier transform introduces a singularity at $g=0$ which is easily avoided by transforming to spherical coordinates.

$$V_{lr}(\mathbf{r}_{12}) = \int d\Omega \int dg_r \eta(g_r) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.6}$$

The Fourier representation of V_{lr} is evaluated numerically, using points and weights, which has stability issues at large $\mathbf{g}\cdot\mathbf{r}_{12}$. The stability issues do not come from $\eta(g_r)$ which has structure at small g_r but decays to zero at large g_r. Owing to the limited range of g_r a fixed integration grid can be used independent of system size. The stability issues arise at large $\mathbf{r}_{12}$ as depicted in figure 2.3.

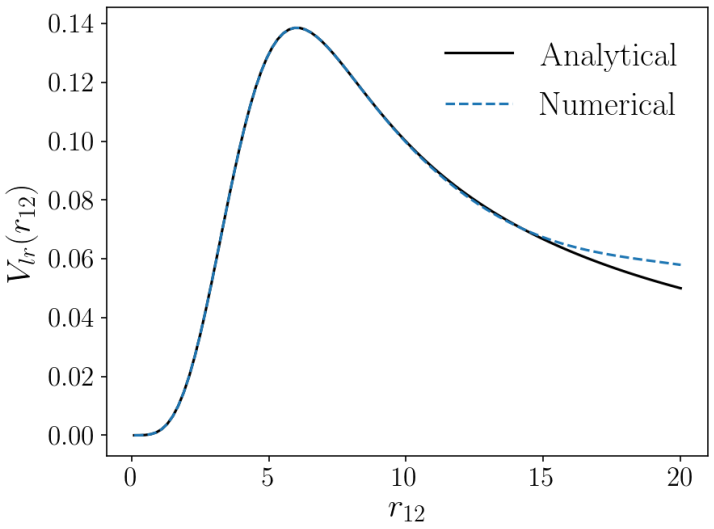

Figure 2.3: The function V_{lr} evaluated analytically and numerically. The numerical integration becomes unstable at large $\boldsymbol{r}_{12}$

The numerical integration can be viewed as the summation of many oscillators with frequencies $\mathbf{g}\cdot\mathbf{r}_{12}$. In order to increase the stability of the numerical integration a finer integration grid can be used, but eventually $\mathbf{r}_{12}$ will become large enough to destabilize the numerical integration.

The multipole representation works well at large $\mathbf{r}_{12}$ but struggles at smaller $\mathbf{r}_{12}$, the Fourier representation exhibits the converse stability. The simplest solution would be to combine the two methods and partition V_{lr} once more.

$$V_{Coul}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{inter}(\mathbf{r}_{12}) + V_{ulr}(\mathbf{r}_{12}) \tag{2.7}$$

$$V_{lr}(\mathbf{r}_{12}) = V_{inter}(\mathbf{r}_{12}) + V_{ulr}(\mathbf{r}_{12}) \tag{2.8}$$

Here V_{inter} is the intermediate range of the potential represented through the Fourier transform and V_{ulr} is the ultra long-range potential represented through the multipole expansion. To represent this potential mathematically, V_{mr} is introduced which is a medium-range potential

with α_m

$$V_{sr}(\mathbf{r}_{12}) = \frac{\text{erfc}(\alpha_s|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_s e^{-\gamma_s|\mathbf{r}_{12}|^2} \tag{2.9}$$

$$V_{mr}(\mathbf{r}_{12}) = \frac{\text{erfc}(\alpha_m|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_m e^{-\gamma_m|\mathbf{r}_{12}|^2} \tag{2.10}$$

$$V_{inter}(\mathbf{r}_{12}) = V_{mr}(\mathbf{r}_{12}) - V_{sr}(\mathbf{r}_{12}) \tag{2.11}$$

$$V_{ulr}(\mathbf{r}_{12}) = \frac{1}{\mathbf{r}_{12}} - V_{mr}(\mathbf{r}_{12}) \tag{2.12}$$

Visually the three-range potential is shown below in figure 2.4

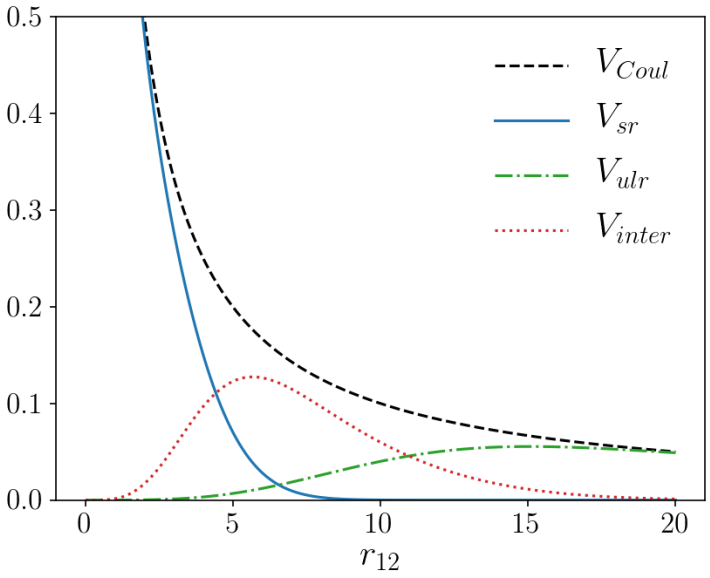

Figure 2.4: The three-range potential with $\alpha_s = 0.5$ and $\alpha_m = 0.2$

The value of α_m is small enough that the multipole expansion is accurate across the whole $\mathbf{r}_{12}$ region and α_s is large enough that V_{sr} has reasonably short extent. The medium-range of the potential is represented by V_{inter} which still exhibits instability at large $\mathbf{r}_{12}$ shown in figure 2.5. At large $\mathbf{r}_{12}$, V_{inter} is zero but the numerical integration does not evaluate to zero. The solution is to introduce a cut off threshold (T_{cut}) such that if $\mathbf{r}_{12} > T_{cut}$ the value of V_{inter} is set to zero.

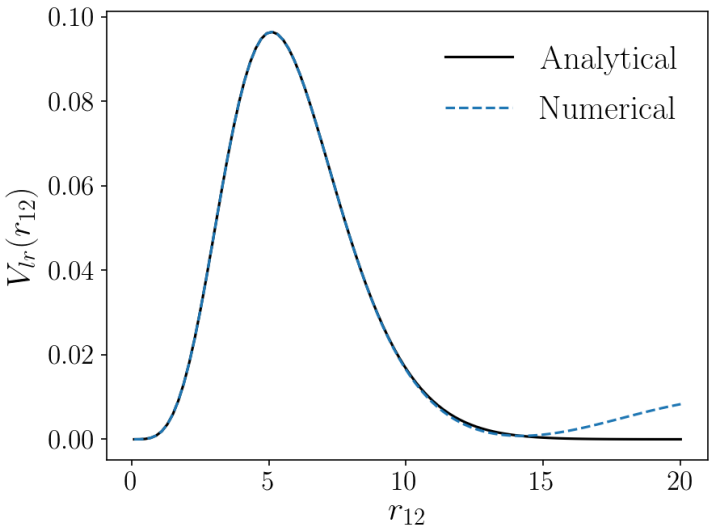

Figure 2.5: The function V_{inter} evaluated analytically and numerically. The numerical integration becomes unstable at large $\boldsymbol{r}_{12}$ where V_{inter} should be zero.

Accurate results of the Fourier transform can be obtained through the "*bandaid*" method of T_{cut}. However, is it possible to further improve the stability and accuracy of the Fourier transform? To answer this question we shall explore a gauge invariant regularized long-range potential. A grid of fixed gauge points $\mathbf{R}_i$ is introduced as points near the coordinates $\mathbf{r}_i$ and then $e^{i\mathbf{g}\cdot\mathbf{R}_i}$ is subtracted from $e^{i\mathbf{g}\cdot\mathbf{r}_i}$ in reciprocal space. Using V_{lr} from the two-range potential as an example, the regularized long-range potential can be introduced as.

$$\begin{aligned} V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) &= \int dg^3\eta(g)(e^{i\mathbf{g}\cdot\mathbf{r}_2}-e^{i\mathbf{g}\cdot\mathbf{R}_2})(e^{-i\mathbf{g}\cdot\mathbf{r}_1}-e^{-i\mathbf{g}\cdot\mathbf{R}_1}) \\ &= V_{lr}(|\mathbf{r}_2-\mathbf{r}_1|)-V_{lr}(|\mathbf{r}_2-\mathbf{R}_1|)-V_{lr}(|\mathbf{R}_2-\mathbf{r}_1|)+V_{lr}(|\mathbf{R}_2-\mathbf{R}_1|) \end{aligned} \tag{2.13}$$

If $\mathbf{R}_i$ is chosen to always be close to $\mathbf{r}_i$ the regularized potential will always be small. Introducing the vector $\mathbf{x}_i = \mathbf{r}_i - \mathbf{R}_i$ the oscillators in the regularized potential can be written as $e^{i\mathbf{g}\cdot\mathbf{R}}(e^{i\mathbf{g}\cdot x}-1)$ which is the product of a fast and slow oscillator. Equation 2.13 can be rearranged into V_{lr}^R and a group of one-body terms called V_{1body}

$$V_{lr}(\mathbf{r}_{12}) = V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2)+V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) \tag{2.14}$$

$$V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = V_{lr}(|\mathbf{r}_2-\mathbf{R}_1|)+V_{lr}(|\mathbf{R}_2-\mathbf{r}_1|)-V_{lr}(|\mathbf{R}_2-\mathbf{R}_1|) \tag{2.15}$$

The advantage here is that V_{lr}^R is the only term that needs to be evaluated numerically, V_{1body} is evaluated analytically and is inexpensive by comparison. The regularized contribution is the smallest contribution by a significant amount only contributing about one percent of the total long-range potential as seen in 2.6. The regularized potential does have some structure at small $\mathbf{r}_{12}$ and in this case has a small negative contribution. The orientation of $\mathbf{R}_i$ with respect to $\mathbf{r}_i$ can adjust the shape of V_{lr}^R.

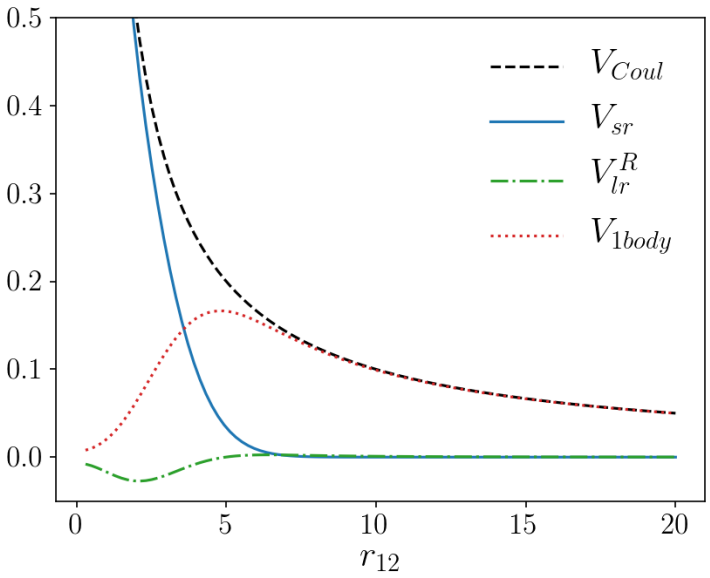

Figure 2.6: The two-range potential with V_{lr} regularized into V_{lr}^R and V_{1body}. The points $\boldsymbol{R}_i$ are placed 1 unit orthogonally away from $\boldsymbol{r}_i$.

A regularized multipole expansion also exists and the regularization procedure can be performed on the three-range potential. The hope was that if V_{lr}^R is small, the accuracy of the Fourier transform would increase. The data shows that this does not appear to be the case. Using the regularized potential in general does not appear to improve the accuracy. However, when accuracy is not the greatest concern V_{lr} can be approximated as V_{1body} by either fully neglecting V_{lr}^R or neglecting V_{lr}^R for $\mathbf{r}_{12}$ larger than a threshold. This approximation would be a fairly cheap approximation to the long-range potential. Cases where this might be advantageous would be in early stages of geometry optimizations or in transition state searches.

The disadvantages to the regularization technique are that it is fairly complicated and doesn't really provide a clear gain in accuracy. The cost of evaluating V_{lr}^R is similar to the cost of evaluating V_{lr}. The definition of the gauge points $\mathbf{R}_i$ is ambiguous, which could potentially

lead to small discrepancies every time the gauge centres are defined. Any algorithm that uses this potential should have a clear algorithm to define the gauge centres.

In summary the Coulomb potential will be explored through the two range-separated potentials, two-range and three-range, which are further divided into eight evaluation schemes. The schemes are summarized in the table 2.1 below. Several schemes require a threshold T_{cut} which occurs at a specific $\mathbf{r}_{12}$ value. Every scheme involving a Fourier transform requires T_{cut} to turn off the numerical integration at $\mathbf{r}_{12}=T_{cut}$. In Scheme 2, T_{cut} is used to turn off the Fourier transform while simultaneously turning on the multipole approximation. Schemes four through eight are regularized schemes requiring gauge centres.

Scheme	Long-Range Potential	Multipole	Fourier	One-Body	Notes
1	V_{lr}	✓			Requires large α
2	V_{lr}	✓	✓		Switch between FT and MP at T_{cut}
3	$V_{inter}+V_{ulr}$	✓	✓		Turn off FT at T_{cut}
4	$V_{lr}^{R}+V_{1body}$	✓		✓	
5	$V_{lr}^{R}+V_{1body}$		✓	✓	Turn off FT at T_{cut}
6	V_{1body}			✓	Use $V_{lr}^{R}=0$
7	$V_{inter}^{R}+V_{ulr}^{R}+V_{1body}$	✓	✓	✓	Regularize with α_s
8	$V_{inter}+V_{ulr}^{R}+V_{1body}$	✓	✓	✓	Regularize with α_m

Table 2.1: Eight schemes to evaluate the long-range Coulomb potential.

2.2 Derivations

This section is devoted to deriving many of the equations introduced in the introductory section of this chapter. For ease of access some equations have been repeated. **Boldface** is used to emphasizes vector quantities while non-bold typeface is used to represent scalars. Repeated indices are assumed to be summed over.

2.2.1 Partitioning the Coulomb Potential

Range separation and long-range regularization is most easily understood through the Fourier transform. Let $V(\mathbf{r}_{12})$ be the Coulomb potential between points $\mathbf{r}_1$ and $\mathbf{r}_2$ ($\mathbf{r}_{12}=\mathbf{r}_2-\mathbf{r}_1$) such

that the Fourier transform is written as:

$$\frac{1}{|\mathbf{r}_{12}|}=\frac{4\pi}{(2\pi)^3}\int d\mathbf{g}^3\frac{1}{|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.16}$$

this function can be partitioned into long and short-range parts by introducing the function $\theta(\mathbf{g})=e^{-\omega|\mathbf{g}|^2}$, as employed in Ewald summation[86,87].

$$\frac{1}{|\mathbf{r}_{12}|}=\frac{4\pi}{(2\pi)^3}\int d\mathbf{g}^3(1-\theta(\mathbf{g}))\frac{1}{|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}}+\frac{4\pi}{(2\pi)^3}\int d\mathbf{g}^3\theta(\mathbf{g})\frac{1}{|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.17}$$

$$=V_{sr}(\mathbf{r}_{12})+\frac{4\pi}{(2\pi)^3}\int d\mathbf{g}^3\theta(\mathbf{g})\frac{1}{|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.18}$$

The short-range potential can be evaluated in real space as:

$$V_{sr}(\mathbf{r}_{12})=\frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} \tag{2.19}$$

$$\alpha=\frac{1}{2\sqrt{\omega}} \tag{2.20}$$

The short-range potential is well behaved and is not the focus of this chapter. The long-range potential has some issues that need to be addressed and hence the remainder of the chapter will focus on the long-range potential. The first issue with the long-range potential is that it is not purely long range, and the second issue is that the Fourier representation has a $|\mathbf{g}|^{-2}$ singularity. A Gaussian ($X_0e^{-\gamma|\mathbf{r}_{12}|^2}$) is added and subtracted in real-space to set the long-range potential to zero at $r=0$[93].

$$V_{sr}(\mathbf{r}_{12})=\frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|}+X_0e^{-\gamma|\mathbf{r}_{12}|^2} \tag{2.21}$$

$$V_{lr}(\mathbf{r}_{12})=\frac{4\pi}{(2\pi)^3}\int d\mathbf{g}^3\theta(\mathbf{g})\frac{1}{|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}}-\frac{X_0}{(2\pi)^3}\frac{1}{(2\gamma)^{3/2}}\int d\mathbf{g}^3e^{-\frac{1}{4\gamma}|\mathbf{g}|^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}}$$

$$=\int d\mathbf{g}^3\frac{1}{|\mathbf{g}|^2}\eta(\mathbf{g})e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.22}$$

$$\eta(\mathbf{g})=\frac{4\pi}{(2\pi)^3}e^{-\omega|\mathbf{g}|^2}-\frac{|\mathbf{g}|^2}{(2\pi)^3(2\gamma)^{3/2}}e^{-\frac{1}{4\gamma}|\mathbf{g}|^2} \tag{2.23}$$

The parameters, X_0 and γ are chosen by setting the limit of the long-range potential and its

first derivative to zero at $r=0$.

$$X_0 = \lim_{r\to 0} \frac{\mathrm{erf}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} = \frac{2\alpha}{\sqrt{\pi}} \tag{2.24}$$

$$\frac{d}{dr} V_{lr}(r=0) = 2X_0(-\frac{\alpha^2}{3} + \gamma) \tag{2.25}$$

$$\gamma_{opt} = \frac{\alpha^2}{3} \tag{2.26}$$

Therefore the potential has only one controlling parameter α, such that the smaller the α the further the short-range region extends. Typically α is chosen between 0.1 and 0.6. It is possible to have a long-range potential with additional features such as a small short-range attractive region. This is done by choosing a γ value smaller than γ_{opt}. There maybe applications where this could be a useful potential, however, for this thesis this is not ideal and γ_{opt} will be used for all potentials. Visually this is depicted in figure 2.7

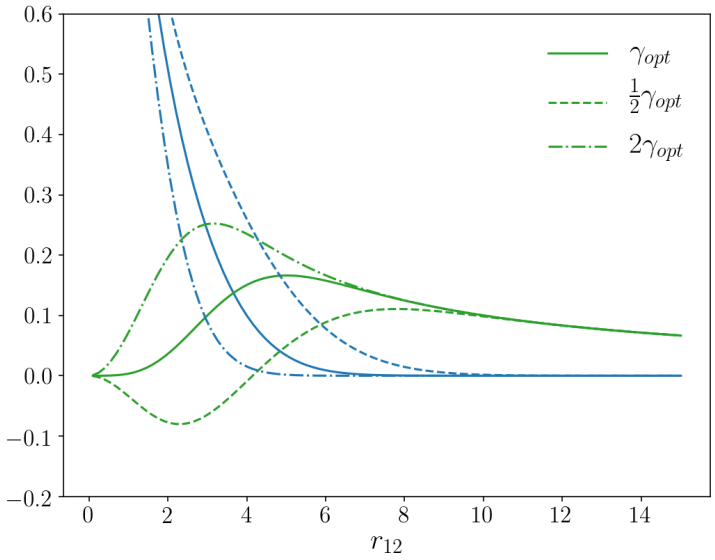

Figure 2.7: The Coulomb potential separated into two parts, short-range(blue) and long-range(green). The solid lines are for the optimal γ where the long-range is flattest for small $\mathbf{r}_{12}$. Choosing values larger or smaller than γ_{opt} can result in a steep slope (dash-dot) or a slightly negative potential (dashed). The shape of the short-range potential remains consistent with respect to γ

The three-range potential is obtained by further partitioning V_{lr} into medium-range and ultra long-range contributions. This is done though two α's, α_s which controls the extent of the short-range potential and α_m which controls the onset of the ultra long-range potential. A

medium-range potential V_{mr} is used to connect the three parts of the potential.

$$V_{sr}(\mathbf{r}_{12}) = \frac{\operatorname{erfc}(\alpha_s|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_s e^{-\gamma_s|\mathbf{r}_{12}|^2} \tag{2.27}$$

$$V_{mr}(\mathbf{r}_{12}) = \frac{\operatorname{erfc}(\alpha_m|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_m e^{-\gamma_m|\mathbf{r}_{12}|^2} \tag{2.28}$$

$$\begin{aligned} V_{inter}(\mathbf{r}_{12}) &= V_{mr}(|\mathbf{r}_{12}|) - V_{sr}(|\mathbf{r}_{12}|) \\ &= \frac{\operatorname{erfc}(\alpha_m|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} - \frac{\operatorname{erfc}(\alpha_s|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_m e^{-\gamma_m|\mathbf{r}_{12}|^2} - X_s e^{-\gamma_s|\mathbf{r}_{12}|^2} \\ &= \frac{\operatorname{erf}(\alpha_s|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} - \frac{\operatorname{erf}(\alpha_m|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} - X_s e^{-\gamma_s|\mathbf{r}_{12}|^2} + X_m e^{-\gamma_m|\mathbf{r}_{12}|^2} \end{aligned} \tag{2.29}$$

$$\begin{aligned} V_{ulr}(\mathbf{r}_{12}) &= \frac{1}{|\mathbf{r}_{12}|} - V_{mr}(\mathbf{r}_{12}) \\ &= \frac{\operatorname{erf}(\alpha_m|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} - X_m e^{-\gamma_m|\mathbf{r}_{12}|^2} \end{aligned} \tag{2.30}$$

The Fourier representation of V_{inter} is denoted as:

$$V_{inter}(\mathbf{r}_{12}) = \int dg^3 \frac{1}{|\mathbf{g}|^2} \eta_3(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{2.31}$$

$$\begin{aligned} \eta_3(\mathbf{g}) = &\frac{4\pi}{(2\pi)^3} e^{-\omega_s|\mathbf{g}|^2} - \frac{|\mathbf{g}|^2}{(2\pi)^3(2\gamma_s)^{3/2}} e^{-\frac{1}{4\gamma_s}|\mathbf{g}|^2} \\ &- \frac{4\pi}{(2\pi)^3} e^{-\omega_m|\mathbf{g}|^2} + \frac{|\mathbf{g}|^2}{(2\pi)^3(2\gamma_m)^{3/2}} e^{-\frac{1}{4\gamma_m}|\mathbf{g}|^2} \end{aligned} \tag{2.32}$$

The numerical integration of the two-range or three-range potential is mechanically the same. The functions $\eta(\mathbf{g})$ from equation 2.23 and $\eta_3(\mathbf{g})$ from equation 2.32 are the only difference between the numerical contributions of the potentials. Therefore the same code can be used to integrate both potentials.

Numerical Integration

The numerical evaluation of the long-range potential is performed in spherical coordinates to cancel the singularity in the Fourier transform. The long-range integral is written as a sum

of both angular and radial grid points.

$$\begin{aligned} V_{lr}(\mathbf{r}_{12}) &= \frac{4\pi}{(2\pi)^3}\int\int\int d\theta d\phi dg_r g_r^2 sin(\theta)\eta(g_r)\frac{1}{g_r^2}e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \\ &= \sum_{\Omega,r} w_\Omega w_r e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \end{aligned} \tag{2.33}$$

It may appear that the angular points are redundant, which would be true for a purely radial potential. However, the angular points are required when gauge centres are introduced as the orientation of $\mathbf{R}_i$ with respect $\mathbf{r}_i$ is important. In addition the numerical integration of atomic orbitals, introduced in chapter 3 requires an accurate angular grid.

The angular grid points are represented through Lebedev quadrature[88], which is readily available through the Python package QuadPy[92]. Lebedev grids are only available for specific polynomial orders as these quadratures are derived from octahedral symmetries. The Lebedev quadratures are the best choice for representing the angular part of V_{lr}.

Unlike the angular points, there are many quadratures available to represent the radial grids points. The radial points need to be well suited to represent the radial function $\eta(g_r)$ or $\eta_3(g_r)$ for V_{inter}. The η functions have structure at small g_r but quickly decay to zero at large g_r. Several common radial quadratures are shown in figure 2.8 along with the function $\eta(g_r)$. Here it can be seen that number of key integration points (points around small g_r) are more or less the same, and therefore an equidistant radial grid was found to be sufficient. Quadratures such as Gauss-Legendre[91] and Gauss-Chebyshev[90] were found to not have much significant improvement on accuracy.

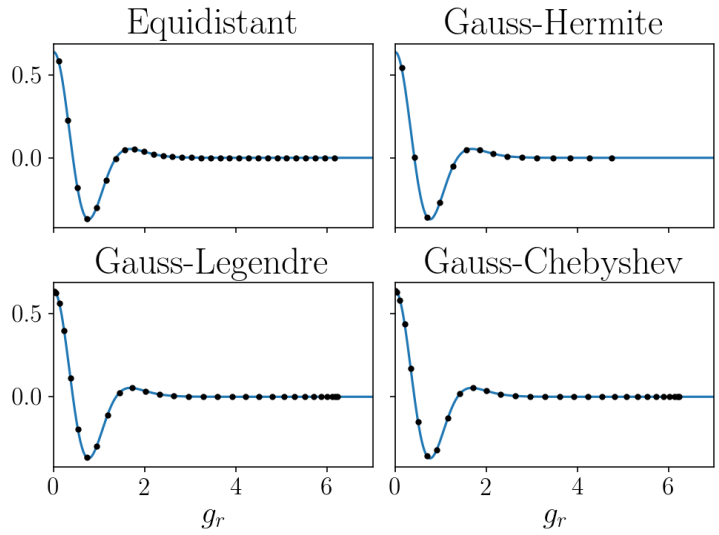

Figure 2.8: The radial function $\eta(g_r)$, with different numerical quadratures.

2.2.2 Regularization of the Long-Range Potential

The original motivation for regularizing the long-range potential was to explicitly remove $|\mathbf{g}|^{-2}$ singularity and improve the accuracy of the Fourier transform at large $\mathbf{g}\cdot\mathbf{r}$. The radial function $\eta(g_r)$ decays quickly for large g_r but when g_r is small and r is very large the numerical integration destabilizes. This is because we are essentially trying to represent a rapid oscillator with a small number of grid points. One solution is to use a larger numerical grid for larger systems, while this will work, the scaling with respect to system size is impractical.

Instead we introduce a grid of fixed points called gauge centres ($\mathbf{R}_i$) that are near to the points $\mathbf{r}_i$ and subtracting $e^{i\mathbf{g}\cdot\mathbf{R}_i}$. The Fourier transform can be rewritten as

$$V_{lr}^R(\mathbf{r},\mathbf{R}) = \int dg^3 \eta(\mathbf{g})(e^{i\mathbf{g}\cdot\mathbf{r}} - e^{i\mathbf{g}\cdot\mathbf{R}}) \tag{2.34}$$

This integrand will always be small as long as $\mathbf{r}$ is close to $\mathbf{R}$ and the $|\mathbf{g}|^{-2}$ singularity is also removed. For convenience, the vector $\mathbf{x}_i := \mathbf{r}_i - \mathbf{R}_i$ is introduced such that:

$$V_{lr}^R(\mathbf{r},\mathbf{R}) = \int dg^3 \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{R}}(e^{i\mathbf{g}\cdot\mathbf{x}} - 1) \tag{2.35}$$

Which can be viewed as the product of a fast oscillator and a slow oscillator, this should

improve the accuracy of the numerical integration over V_{lr}. In figure 2.9 the real part of the Fourier integrands are shown for both V_{lr} and V_{lr}^R. It is clear that V_{lr}^R is much smaller than V_{lr} however both do oscillate at similar frequencies as depicted in figure 2.9.

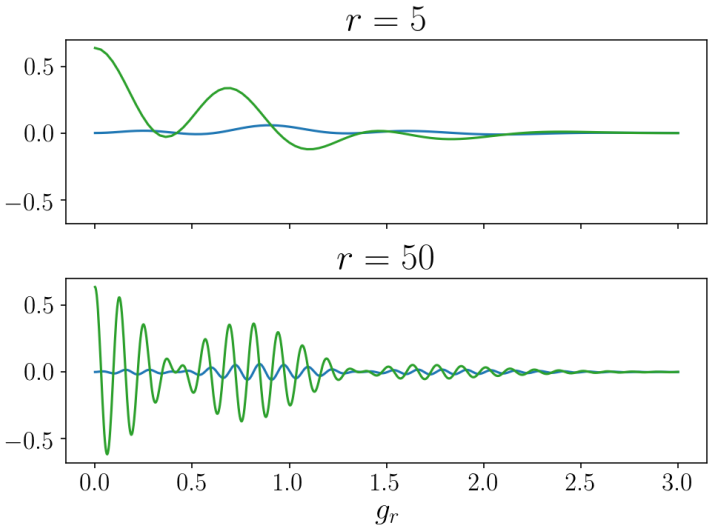

Figure 2.9: The real part of the Fourier integrands as a function of g_r for the V_{lr}^R (blue) and V_{lr} (green) for two different distances. The gauge centre $\boldsymbol{R}$ is 0.2 from $\boldsymbol{r}$.

Using the strategy introduced above, V_{lr}^R can be written for $\mathbf{r}_{12} := \mathbf{r}_2 - \mathbf{r}_1$.

$$\begin{aligned} V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) &= \int dg^3 \eta(\mathbf{g})(e^{i\mathbf{g}\cdot\mathbf{r}_2} - e^{ig\cdot\mathbf{R}_2})(e^{-i\mathbf{g}\cdot\mathbf{r}_1} - e^{-i\mathbf{g}\cdot\mathbf{R}_1}) \\ &= \int dg^3 \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_2} e^{-i\mathbf{g}\cdot\mathbf{r}_1} - \int dg^3 \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_2} e^{-i\mathbf{g}\cdot\mathbf{R}_1} \\ &\quad - \int dg^3 \eta(\mathbf{g}) e^{-i\mathbf{g}\cdot\mathbf{r}_1} e^{i\mathbf{g}\cdot\mathbf{R}_2} + \int dg^3 \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{R}_2} e^{-i\mathbf{g}\cdot\mathbf{R}_1} \\ &= V_{lr}(|\mathbf{r}_2 - \mathbf{r}_1|) - V_{lr}(|\mathbf{r}_2 - \mathbf{R}_1|) - V_{lr}(|\mathbf{R}_2 - \mathbf{r}_1|) + V_{lr}(|\mathbf{R}_2 - \mathbf{R}_1|) \end{aligned} \tag{2.36}$$

Rearranging the above equation the long-range potential can be written as the sum of the regularized potential plus the one-body potential.

$$V_{lr}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) + V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) \tag{2.37}$$

$$V_{1body}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = V_{lr}(|\mathbf{r}_2 - \mathbf{R}_1|) + V_{lr}(|\mathbf{R}_2 - \mathbf{r}_1|) - V_{lr}(|\mathbf{R}_2 - \mathbf{R}_1|) \tag{2.38}$$

In the three-range potential the intermediate-range contribution can be regularized as:

$$V^R_{inter}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = \int dg^3 \eta_3(\mathbf{g})(e^{i\mathbf{g}\cdot\mathbf{r}_2} - e^{ig\cdot\mathbf{R}_2})(e^{-i\mathbf{g}\cdot\mathbf{r}_1} - e^{-i\mathbf{g}\cdot\mathbf{R}_1}) \tag{2.39}$$

$$= V_{inter}(|\mathbf{r}_2-\mathbf{r}_1|) - V_{inter}(|\mathbf{r}_2-\mathbf{R}_1|)$$

$$- V_{inter}(|\mathbf{R}_2-\mathbf{r}_1|) + V_{inter}(|\mathbf{R}_2-\mathbf{R}_1|) \tag{2.40}$$

The ultra long-range potential can be regularized as well.

$$V^R_{ulr}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = \int dg^3 (\eta(\mathbf{g}) - \eta_3(\mathbf{g}))(e^{i\mathbf{g}\cdot\mathbf{r}_2} - e^{ig\cdot\mathbf{R}_2})(e^{-i\mathbf{g}\cdot\mathbf{r}_1} - e^{-i\mathbf{g}\cdot\mathbf{R}_1}) \tag{2.41}$$

$$= V_{lr}(|\mathbf{r}_2-\mathbf{r}_1|) - V_{lr}(|\mathbf{r}_2-\mathbf{R}_1|) - V_{lr}(|\mathbf{R}_2-\mathbf{r}_1|) + V_{lr}(|\mathbf{R}_2-\mathbf{R}_1|)$$

$$- V_{inter}(|\mathbf{r}_2-\mathbf{r}_1|) + V_{inter}(|\mathbf{r}_2-\mathbf{R}_1|)$$

$$+ V_{inter}(|\mathbf{R}_2-\mathbf{r}_1|) - V_{inter}(|\mathbf{R}_2-\mathbf{R}_1|) \tag{2.42}$$

There are two choices for the function $\eta(\mathbf{g})$ in the regularization of V_{ulr} using either α_s or α_m. If α_s is chosen $V_{lr} - V_{inter} = V_{ulr}$ such that equation 2.42 becomes:

$$V^R_{ulr}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2) = V_{ulr}(|\mathbf{r}_2-\mathbf{r}_1|) - V_{ulr}(|\mathbf{r}_2-\mathbf{R}_1|) - V_{ulr}(|\mathbf{R}_2-\mathbf{r}_1|) + V_{ulr}(|\mathbf{R}_2-\mathbf{R}_1|) \tag{2.43}$$

If α_m is chosen $V_{lr} - V_{inter} = V_{ulr} - V_{inter}$ which does not simplify things, therefore α_m is chosen. It is simpler to leave V_{inter} unregularized and regularize V_{ulr} only. To summarize the two choices;

- α_s : $V_{Coul} = V_{sr} + V^R_{inter} + V^R_{ulr} + V_{1body}(\alpha_s)$
- α_m : $V_{Coul} = V_{sr} + V_{inter} + V^R_{ulr} + V_{1body}(\alpha_m)$

2.2.3 Regularized Multipole Expansion

Similar to the Fourier representation the regularized long-range potential can be approximated through a multipole expansion. Cartesian coordinates are used for the multipole expansion as it is much simpler to understand the regularization procedure even though the Cartesian multipole expansion contains redundant information. First let $V(\mathbf{r})$ be a general radial potential.

A Taylor expansion about the point $\mathbf{r}=\mathbf{x}+\mathbf{R}$ is then performed. For convenience the variable $s=\frac{1}{2}\mathbf{r}\cdot\mathbf{r}=\frac{1}{2}(\mathbf{x}+\mathbf{R})\cdot(\mathbf{x}+\mathbf{R})$ is introduced, and it is assumed that the potential can be written in the form $V(s)$. The term V will be used for both and repeated indices are summed over. The third order Taylor expansion is given as:

$$V(\mathbf{x}+\mathbf{R})\approx V(\mathbf{R})+\frac{\partial V}{\partial x_i}x_i+\frac{1}{2}\frac{\partial^2 V}{\partial x_i\partial x_j}x_ix_j+\frac{1}{3!}\frac{\partial^3 V}{\partial x_i\partial x_j\partial x_k}x_ix_jx_k \tag{2.44}$$

changing variables from $\mathbf{r}$ to $\mathbf{s}$ and using the chain rule.

$$\begin{aligned}
\frac{\partial V}{\partial x_i}&=\frac{\partial V}{\partial s}\frac{\partial s}{\partial x_i}\\
&=\frac{\partial V}{\partial s}(x_i+R_i) &(2.45)\\
\frac{\partial^2 V}{\partial x_i\partial x_j}&=\frac{\partial}{\partial x_j}\frac{\partial V}{\partial x_i}\\
&=\frac{\partial^2 V}{\partial s^2}(x_i+R_i)(x_j+R_j)+\delta_{ij}\frac{\partial V}{\partial s} &(2.46)\\
\frac{\partial^3 V}{\partial x_i\partial x_j\partial x_k}&=\frac{\partial}{\partial x_k}\frac{\partial^2 V}{\partial x_i\partial x_j}\\
&=\frac{\partial^3 V}{\partial s^3}(x_i+R_i)(x_j+R_j)(x_k+R_k)\\
&+\frac{\partial^2 V}{\partial s^2}\big(\delta_{ik}(x_j+R_j)+\delta_{jk}(x_i+R_i)+\delta_{ij}(x_k+R_k)\big) &(2.47)
\end{aligned}$$

Note that derivatives of s with respect to x with order greater than two are zero, which simplifies the expansion. Evaluating the derivatives at $x=0$:

$$\begin{aligned}
V(\mathbf{x}+\mathbf{R})\approx V(\mathbf{R})&+\frac{\partial V}{\partial s}R^ix_i+\frac{1}{2}\big(\frac{\partial^2 V}{\partial s^2}R^iR^j+\delta_{ij}\frac{\partial V}{\partial s}\big)x_ix_j\\
&+\frac{1}{3!}\big[\frac{\partial^3 V}{\partial s^3}R^iR^jR^k+\frac{\partial^2 V}{\partial s^2}\big(\delta_{ij}R^k+\delta_{ik}R^j+\delta_{jk}R^i\big)\big]x_ix_jx_k
\end{aligned} \tag{2.48}$$

Above is the expansion for a general potential in Cartesian coordinates, as long as it is differentiable. One of the main advantages of this representation is that the derivatives of the potential are only dependent on the expansion centres $\mathbf{R}$ and not the moment coordinates $\mathbf{x}$. This means that when atomic orbitals are introduced only the Cartesian moment integrals are required.

Similar to the one sided expansion the radial potential $V(\mathbf{r}_2-\mathbf{r}_1)$ can be expanded about points $\mathbf{x},\mathbf{y}=0$ using the substitution $\mathbf{r}_1=\mathbf{x}+\mathbf{R}_1$ and $\mathbf{r}_2=\mathbf{y}+\mathbf{R}_2$.

$$V(\mathbf{r}_2-\mathbf{r}_1)\equiv V\big[(\mathbf{y}-\mathbf{x})+(\mathbf{R}_y-\mathbf{R}_x)\big]\equiv V\big[(\mathbf{y}-\mathbf{x})+\mathbf{R}\big] \tag{2.49}$$

Sparing some details.

$$\begin{aligned}V(\mathbf{y}-\mathbf{x}+\mathbf{R})\approx & V(\mathbf{R})+\frac{\partial V}{\partial s}R^i(y_i-x_i)\\ &+\frac{1}{2}\Big(\frac{\partial^2 V}{\partial s^2}R^iR^j+\delta_{ij}\frac{\partial V}{\partial s}\Big)(x_ix_j-(y_ix_j+x_iy_j)+y_iy_j)\\ &+\frac{1}{3!}\Big[\frac{\partial^3 V}{\partial s^3}R^iR^jR^k+\frac{\partial^2 V}{\partial s^2}\big(\delta_{ij}R^k+\delta_{ik}R^j+\delta_{jk}R^i\big)\Big]\\ &(y_iy_jy_k-(x_iy_jy_k+y_iy_jx_k+y_ix_jy_k)+(y_ix_jx_k+x_iy_jx_k+x_ix_jy_k)+x_ix_jx_k)\end{aligned} \tag{2.50}$$

A more compact notation can be introduced for convenience.

$$V(\mathbf{y}-\mathbf{x}+\mathbf{R})=\sum_{k,l}m_y^k f^{(k,l)}(\mathbf{R})m_x^l \tag{2.51}$$

Here m terms are the Cartesian moments for x and y, respectively, with orders k and l, while f is the derivatives of the potential. See appendix B for the two-range potential $f^{(k,l)}$ terms.

The regularized multipole potential can be derived by subtracting the one-body potential, equation 2.38 from V.

$$\begin{aligned}V^R(\mathbf{y}-\mathbf{x}+\mathbf{R})&=V(\mathbf{y}-\mathbf{x}+\mathbf{R})-V(\mathbf{r}_2-\mathbf{R}_1)-V(\mathbf{R}_2-\mathbf{r}_1)+V(\mathbf{R}_2-\mathbf{R}_1)\\ &=V(\mathbf{y}-\mathbf{x}+\mathbf{R})-V(\mathbf{y}+\mathbf{R})-V(\mathbf{R}-\mathbf{x})+V(\mathbf{R})\\ &=\sum_{k,l}m_y^k f^{(k,l)}(\mathbf{R})m_x^l-\sum_k m_y^k f^{(k,0)}(\mathbf{R})-\sum_l f^{(0,l)}(\mathbf{R})m_x^l+V(\mathbf{R})\\ &=\sum_{k,l}m_y^k f^{(k,l)}(\mathbf{R})m_x^l-\sum_{k\neq 0}m_y^k f^{(k,0)}(\mathbf{R})-\sum_{l\neq 0}f^{(0,l)}(\mathbf{R})m_x^l-f^{(0,0)}\\ &=\sum_{k,l\neq 0}m_y^k f^{(k,l)}(\mathbf{R})m_x^l\end{aligned} \tag{2.52}$$

The terms in the one-body potential can be expanded as one sided multipole expansions. In other words, the regularized multipole expansion is the multipole expansion without the k,$l=0$ terms.

A very important feature to the regularized multipole expansion is that the required Cartesian moment integrals are one order lower than that of the expansion coefficients. For example this is very useful when Gaussian basis sets are added because octupole accuracy can be achieved with only the hexapole moments.

2.3 Analysis of the Potential

The evaluation schemes presented in table 2.1 are analysed in this section. Since it is possible to analytically calculate all contributions through subtraction, errors are reported as the log of the absolute errors. Log errors around negative five or smaller are considered manageable, whereas log errors larger than negative three are considered too large. Typically we would want the errors to be as small as possible.

$$\text{Log Error} = \log_{10}(|V_{\text{analytical}} - V_{\text{numerical}}|) \tag{2.53}$$

The schemes involving regularization (schemes 4-8) require additional vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ which are defined as points near $\mathbf{r}_1$ and $\mathbf{r}_2$, typically 0.2 to 2 units away. To limit the amount of data, we limit ourselves to a collinear configuration where all points lie on a line unless specified as depicted in figure 2.10.

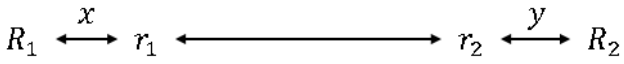

Figure 2.10: The orientation of the gauge centres with respect to the points $\mathbf{r}_1$ and $\mathbf{r}_2$.

When the points $\mathbf{r}_1$ and $\mathbf{r}_2$ are stretched the points $\mathbf{R}_1$ and $\mathbf{R}_2$ are adjusted such that $\mathbf{x}$ and $\mathbf{y}$ remain fixed.

2.3.1 Unregularized Potentials

The range separation parameter α has a large effect on the accuracy of both the multipole and Fourier contributions to the long-range potential. The smaller the value of α the further the short-range potential extends which reduces the toll on the long-range potential. Large α values result in a shorter short-range potential increasing the demand on the long-range potential, specifically in the short-range region of V_{lr}. The effect of α on the two-range potential is depicted in figure 2.11 where each column represents a different α. The bottom row of figure 2.11 displays the accuracy of both the Fourier and multipole methods. The Fourier method is evaluated with a radial grid of 30 points and the multipole method is evaluated with up to hexapole contributions.

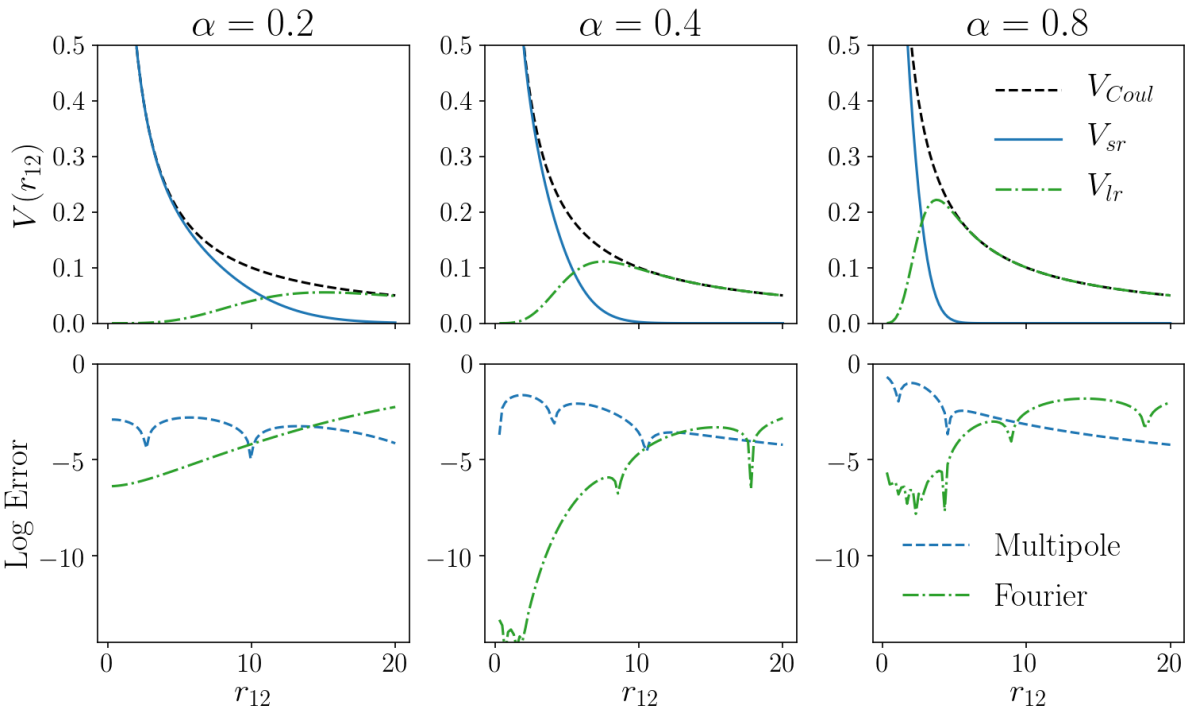

Figure 2.11: The effect of α on accuracy of the two-range potential. Each column represents a different α value while the top row depicts the range-separated potential and the bottom row displays the accuracy of V_{lr} evaluated through either the multipole or Fourier representations.

The multipole expansion (blue) has poor accuracy for small $\mathbf{r}_{12}$, but the accuracy increases with increasing $\mathbf{r}_{12}$. The multipole expansion is reasonable at small $\mathbf{r}_{12}$ for only the smallest values of α; at large $\mathbf{r}_{12}$ the multipole expansion is always viable regardless of α. The Fourier representation is only viable for small $\mathbf{r}_{12}$ and it is a poor choice once $V_{lr} \approx V_{Coul}$.

Scheme 1 in table 2.1 is to evaluate V_{lr} using only the multipole expansion, which is only viable for $\alpha \leq 0.2$. Scheme 1 is not ideal for large systems because V_{sr} extends too far. Scheme 2 evaluates V_{lr} using the Fourier representation at small $\mathbf{r}_{12}$ and switches to the multipole representation when it becomes more accurate than the Fourier method. The switch occurs at the point where the curves cross in the bottom row of figure 2.11. The switching scheme is reasonable for point charges, however, once atomic orbitals are introduced the switching point becomes ambiguous due to the extents of the orbitals which is not ideal.

Evaluation scheme 3 uses the three-range potential which requires the evaluation of V_{inter} to be truncated at $\mathbf{r}_{12} > T_{cut}$. The accuracy of this scheme is shown as the bottom half of figure 2.12. The dashed line in bottom half of figure 2.12 represents the accuracy of the three-range potential if V_{inter} was calculated analytically, depicting the multipole error, while the dashed dot line represents the accuracy if V_{ulr} was a calculated analytically depicting the Fourier error. The overall accuracy of the three-range potential is reasonable provided that V_{inter} is truncated at $\mathbf{r}_{12} = T_{cut}$.

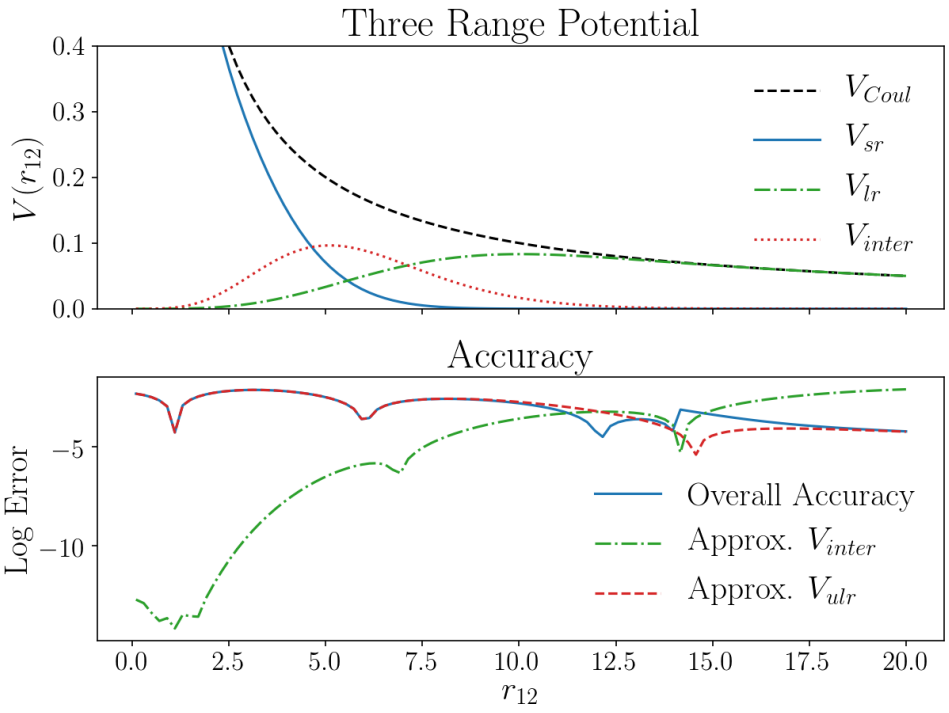

Figure 2.12: Top half: The three-range potential with $\alpha_s = 0.5$ and $\alpha_m = 0.3$. Bottom half: The overall accuracy (solid blue) of the three-range potential with the numerical evaluation of V_{inter} truncated at 14. The accuracy using approximated V_{ulr} and analytical V_{inter} is shown as the dashed line. The accuracy using approximated V_{inter} and analytical V_{ulr} is shown as the dashed dot line.

2.3.2 Regularized Potentials

The accuracy of the regularized two-range potential with respect to $\mathbf{r}_{12}$ is depicted in figure 2.13. The error for this potential solely originates from the evaluation of $V_{lr}^R(\mathbf{r}_1,\mathbf{r}_2,\mathbf{R}_1,\mathbf{R}_2)$ as the one-body and short-range contributions are evaluated analytically. The three methods to evaluate V_{lr}^R gives rise to the evaluation schemes: 4, 5 and 6 which are to evaluate V_{lr}^R through the mutipole expansion, Fourier transform or to approximate V_{lr}^R as zero. The data in figure 2.13 shows us that the regularized Fourier method is not good at long-range and that the multipole method is the best choice for the long-range region. The more interesting conclusion is that approximating V_{lr}^R as zero is not that much worse at long range than evaluating V_{lr}^R through either method.

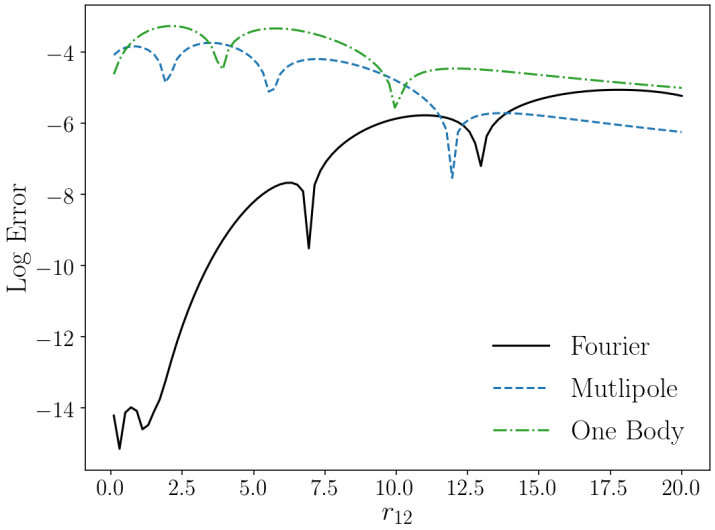

Figure 2.13: The log errors of the regularized two-range potential with α=0.4. The points $\boldsymbol{R}_i$ were located 0.2 from the points $\boldsymbol{r}_i$. The short-range and one-body contributions are evaluated analytically while the regularized long-range is evaluated through the multipole method, blue dashed line, the Fourier method, solid black or not at all, green dash dot.

There are two choices on how to regularize the three-range potential. The first choice is to regularize this potential with α_s which results in a regularized V_{inter} and a regularized V_{ulr}. The other option is to regularized with α_m which results in unregularized V_{inter} and regularized V_{ulr}. The accuracy of these two options is shown in figure 2.14 along with the unregularized three-range potential. In this case there is a gain in accuracy using the α_s regularized potential labelled as scheme 7. The α_m regularized potential is approximately the same accuracy as the unregularized potential. The dotted line labelled α_s switched is scheme 7 but V_{ulr} is evaluated

with the Fourier method for $\mathbf{r}_{12}<7.5$ then evaluated with the multipole method for $\mathbf{r}_{12}>7.5$ this is the most accurate evaluation scheme.

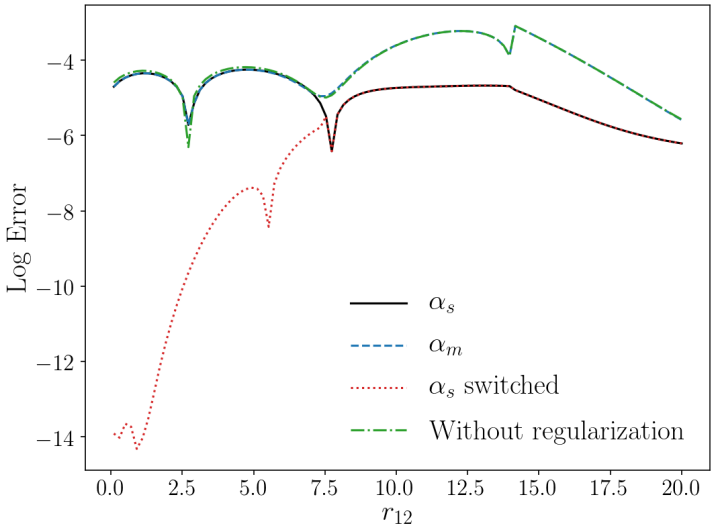

Figure 2.14: The log errors of the regularized three-range potential with $\alpha_s=0.5$ and $\alpha_m=0.3$. The point $\boldsymbol{R}_i$ were located 0.2 from the points $\boldsymbol{r}_i$. The short-range and one-body contributions are evaluated analytically while V_{inter} and V_{ulr} are evaluated through the Fourier and multipole methods respectively. The accuracy of the potentials without regularization is also shown. Note that V_{inter} as truncated at $\boldsymbol{r}_{12}=14$. The line labelled α_s switched has V_{ulr} is evaluated with the Fourier method for $\boldsymbol{r}_{12}<7.5$ then evaluated with the multipole method for $\boldsymbol{r}_{12}>7.5$

Orientation of the Gauge Centres

The orientation of the gauge centres $\mathbf{R}_i$ with respect to the coordinates $\mathbf{r}_i$ is important, specifically the distance between the vectors. Figure 2.15 depicts the average accuracy of the two-range potential when the point $\mathbf{R}_2$ is placed on a sphere of distance $\mathbf{y}$ from the point $\mathbf{r}_2$. The points $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{R}_1$ are fixed in a collinear fashion as in figure 2.10. From this simulation one can see that the average shape of the accuracy curve is similar to the collinear orientation figure 2.13. The distance vector $\mathbf{y}$ has a much larger effect on the accuracy than the spherical orientation.

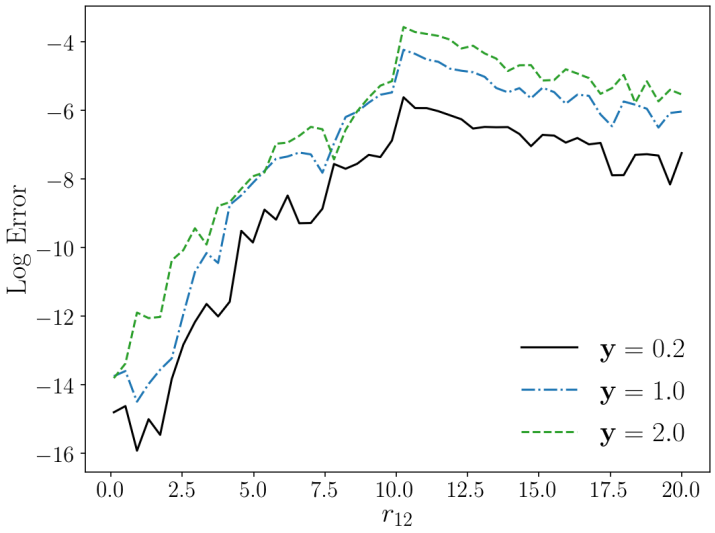

Figure 2.15: The log errors of the regularized two-range potential evaluated with the Fourier method $r_{12} < T_{cut}$ and the multipole method $r_{12} > T_{cut}$ with $\alpha = 0.4$ and $T_{cut} = 10.0$. The error for each value of r_{12} is the average error of 100 random orientations of point $\boldsymbol{R}_2$ on the sphere $\boldsymbol{y}$ from the point $\boldsymbol{r}_2$.

To further investigate the error associated with the angular orientation of $\mathbf{R}_2$, a contour plot is presented in figure 2.16. In this figure the log error is show for the gauge centre $\mathbf{R}_2$ located at $(0.2,\theta,\phi)$ from $\mathbf{r}_2$. In this figure it is clear there are some angles where the accuracy spikes, i.e. the error dips sharply, but most of the angular surface is dominated by larger error.

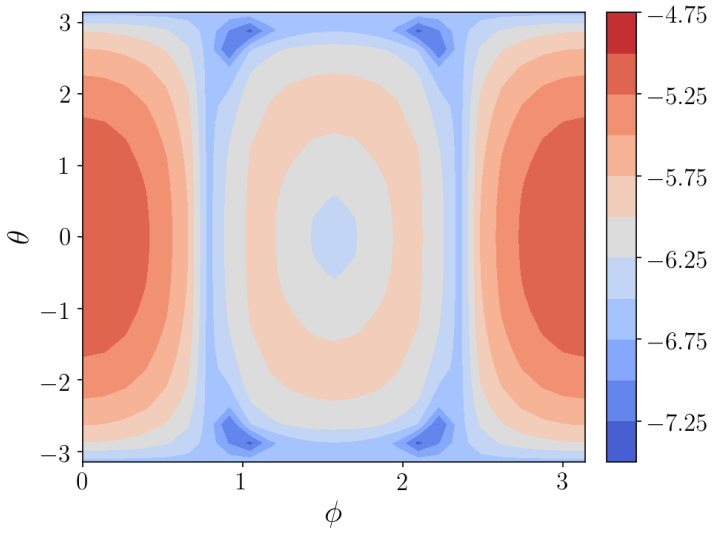

Figure 2.16: A contour plot of the log error for $\boldsymbol{R}_2$ located at $(0.2,\theta,\phi)$ from the point $\boldsymbol{r}_2$. The two-range potential with $\alpha = 0.4$ was used at $r_{12} = 9$. The points $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{R}_1$ are fixed in a collinear fashion.

2.4 Additional Potentials

One of the key strengths to the range-separated method is that the technique will work for any radial potential so long as it is smooth and differentiable. This method is not limited to the two- and three-range potentials. For example, it is possible to change the values of γ in the two-range potential, currently it is determined by setting the first derivative of the potential to zero at $\mathbf{r}_{12}=0$. The method is also not limited to a single Gaussian, it is possible to add a second Gaussian to the long-range potential.

$$V_{lr}(\mathbf{r}_{12})=\frac{erf(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|}-\frac{1}{2}X_0e^{-\gamma_1|\mathbf{r}_{12}|^2}-\frac{1}{2}X_0e^{-\gamma_2|\mathbf{r}_{12}|^2} \tag{2.54}$$

The second Gaussian provides the opportunity to tune the values of γ_1 and γ_2 such that both the first and second derivatives of the potential can be set to zero. The parameter X_0 is chosen such that $V_{lr}(0)=0$. The values of γ_1 and γ_2 are related to the range separation parameter through.

$$\frac{2}{3}\alpha^2-\gamma_1-\gamma_2=0 \tag{2.55}$$

$$\frac{4}{9}\alpha^4-2\gamma_1^2-2\gamma_2^2=0 \tag{2.56}$$

For $\alpha=0.2$ this potential has a very flat long-range region as shown in figure 2.17. The long-range potential does have a brief negative region, which may or may not be beneficial. The regularization of this potential would lead to a $V_{lr}^R(\mathbf{r}_{12})$ that is even smaller than the two-range potential's $V_{lr}^R(\mathbf{r}_{12})$.

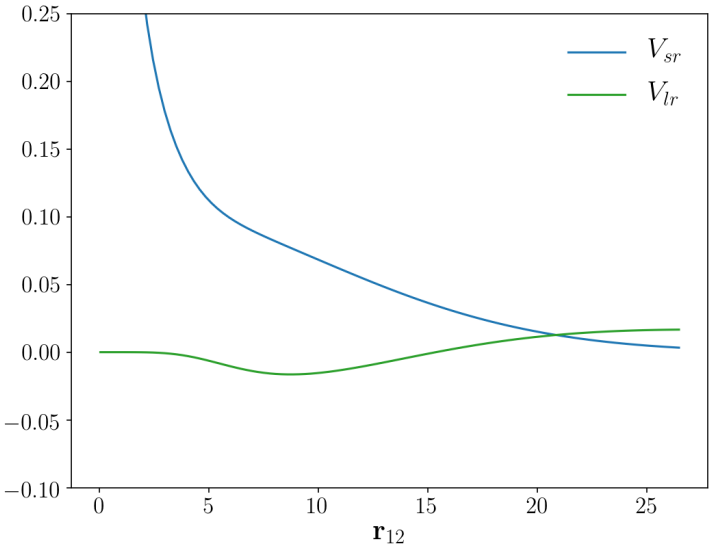

Figure 2.17: The double Gaussian potential with α=0.2.

There are many range-separated potentials that could be explored, the hard part of developing the engine to calculate the range-separated integrals has been achieved in this thesis and it is fairly easy to implement new potentials.

2.5 Conclusion

The evaluation of the range-separated potential is complicated, specifically the evaluation of the long-range region. Contributions to the long-range potential evaluated using the numerical Fourier transform are not accurate enough at large distances irrespective of the grid size. The multipole expansion is only accurate enough when a small range separation parameter α is chosen. A small α implies a far extending short-range potential defeating the purpose of separating the potential in the first place. A three-range potential is the best choice to represent the Coulomb potential out of the schemes presented. In the three-range potential V_{sr} is evaluated analytically, V_{inter} is evaluated through a truncated Fourier transform and V_{ulr} is evaluated through the multipole expansion. In practice the two-range potential is the better choice once atomic orbitals are introduced as discussed in chapter 3.

In an attempt to further improve the numerical accuracy of the long-range potential a regularization scheme was introduced. The regularized potentials are much smaller in magnitude than the one-body potentials which can always be evaluated analytically. Owing to the high

demand on accuracy in electronic structure calculations the regularized contributions must be evaluated accurately. The most accurate evaluation scheme is to regularize the three-range potential using α_s, which is slightly more accurate than the unregularized three-range potential. However, the comparison and accuracy data presented in this chapter is for point charges which can only tell us so much. It is critical to evaluate the numerical schemes in the context of electronic structure calculations presented in the next chapter.

Chapter 3

Two-electron Integrals

3.1 Introduction

The two-electron Coulomb integrals entering a molecular Hamiltonian can be written in terms of the atomic basis functions ϕ_α in a compact (Mulliken) notation as:

$$(\alpha\beta|\gamma\delta) = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_\alpha(\mathbf{r}_1)\phi_\beta(\mathbf{r}_1)\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|}\phi_\gamma(\mathbf{r}_2)\phi_\delta(\mathbf{r}_2) \tag{3.1}$$

The evaluation and representation of the four centre two-electron integrals (equation 3.1) is a critical aspect to the efficiency of quantum chemistry programs. In practice the two-electron integrals are too numerous to store or even fully calculate for large systems. To combat this large computational cost most large scale packages employ a combination of integral pre-screening and the resolution of the identity approximation, often referred to as density fitting (DF)[94–97]. In the DF approximation auxiliary basis functions are used to represent the two-electron integrals as:

$$(\alpha\beta|\gamma\delta) = \sum_{X,Y}(\alpha\beta|X)M^{-1}_{X,Y}(Y|\gamma\delta) \tag{3.2}$$

Where A, B represent the fitting basis and $M^{-1}_{X,Y}$ is the inverse Coulomb metric. In most applications the Coulomb metric only needs to be calculated once and hence the expense of calculating the two-electron integrals is reduced to the calculation of the three centre integrals $(\alpha\beta|X)$. DF with efficient integral pre-screening is an essential tool for many linear scaling electronic structure methods[42,45,98–104]. While the DF approximation is efficient, the three centre integrals are not sparse since the $\mathbf{r}^{-1}$ decay is so slow which can still lead to expensive calculations.

In this chapter the two-electron integrals are explored through the range-separated Coulomb

potentials introduced in the previous chapter, namely the two-range potential.

$$V_{2range}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{lr}(\mathbf{r}_{12}) \tag{3.3}$$

$$V_{sr}(\mathbf{r}_{12}) = \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_0 e^{-\gamma|\mathbf{r}_{12}|^2} \tag{3.4}$$

$$V_{lr}(\mathbf{r}_{12}) = \frac{1}{\mathbf{r}_{12}} - V_{sr}(\mathbf{r}_{12}) \tag{3.5}$$

Integrating the atomic basis functions over the two-range potential allows the two-electron integrals to be partitioned into short and long-range contributions.

$$(\alpha\beta|\gamma\delta)_{2range} = (\alpha\beta|\gamma\delta)_{sr} + (\alpha\beta|\gamma\delta)_{lr} \tag{3.6}$$

The short-range integrals can be approximated through density fitting analogous to the full-range integrals. The long-range integrals are more complicated and are evaluated either numerically through a Fourier transform or they are evaluated through a multipole expansion. Effectively, the four centre integrals are partitioned into three parts.

$$\begin{aligned}(\alpha\beta|\gamma\delta)_{2range} = &\sum_{xy}(\alpha\beta|x)_{sr} M^{-1}_{xy} (y|\gamma\delta)_{sr} \\ &+ \sum_{\mathbf{g}} (\alpha\beta|\mathbf{g})\Theta(\mathbf{R}_{12})\eta(\mathbf{g})(\gamma\delta|\mathbf{g})^* \\ &+ \sum_{mn}(\alpha\beta|m) f^{mn}(\mathbf{R}_{12})\bar{\Theta}(\mathbf{R}_{12})(\gamma\delta|n)\end{aligned} \tag{3.7}$$

The first line of equation 5.4 represents the short-range contribution to the two-electron integrals which is approximated through DF using the auxiliary basis x and the inverse short-range Coulomb metric M^{-1}_{xy}. The short-range three centre integrals $(\alpha\beta|x)$ are sparse because the size of the integral decays exponentially with increasing distance between the orbital's α and β and for each orbital pair there are only a limited number of atomic orbital (AO) centres within the vicinity of each auxiliary centre x. The short-range DF integrals behave essentially the same as full-range DF integrals with the added sparsity, this added sparsity leads to linear scaling with respect to system size because the range of α,β and the auxiliary index is limited.

The second and third line in equation 5.4 represent the long-range contribution of the potential,

the second line is the evaluation of the potential through the Fourier transform method and the third line is the evaluation of the potential through a multipole expansion. The Fourier method is most accurate when the distance between the orbital densities $\rho_{\alpha\beta}$ and $\rho_{\gamma\delta}$ ($\rho_{\alpha\beta}(\mathbf{r}) = \phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})$) is small, and the multipole method is most accurate when the distance between the orbital densities is large. The function $\Theta(\mathbf{R}_{12})$ and its converse $\bar{\Theta}(\mathbf{R}_{12})$ are used as switch between the Fourier and multipole methods, which ever method is most accurate is the method that is used.

The Fourier integrals $(\alpha\beta|\mathbf{g})$ are complex and sparse, $\eta(\mathbf{g})$ denotes the numerical weights of the long-range potential and $\mathbf{g}$ is the numerical grid points. The size of the numerical grid does not need to increase with respect to system size because the function $\eta(\mathbf{g})$ in spherical coordinates decays rapidly with respect to $\mathbf{g}$ and the Fourier method is never used for large $\mathbf{R}_{12}$.

$$\eta(\mathbf{g}) = \frac{4\pi}{(2\pi)^3}\mathbf{g}^2 e^{-\omega|\mathbf{g}|^2} - \frac{1}{(2\pi)^3(2\gamma)^{3/2}} e^{-\frac{1}{4\gamma}|\mathbf{g}|^2} \tag{3.8}$$

For small systems the evaluation of the Fourier integrals is expensive; the number of grid points is larger than the short-range auxiliary basis, but for large systems the size of the auxiliary grows with respect to systems size while the Fourier grid can remain constant.

The multipole integrals are $(\alpha\beta|m)$ and are Cartesian multipole moment integrals expanded about a centre $\mathbf{R}_1$. The multipole method is assembled in Cartesian coordinates for simplicity even though there is some redundant information in this representation. The multipole level m can take on values 0,1,2,3 and these integrals are independent of the potential, the potential information is represented in the function $f^{mn}(\mathbf{R}_{12})$ which is the analytic derivatives of the potential. These quantities are inexpensive with respect to the DF and Fourier integrals.

The challenge in assembling the two-electron integrals through the two-range potential is correctly choosing either the Fourier or multipole method to evaluate the long-range contribution. The accuracy of each method largely depends on the distance between the orbital densities $\rho_{\alpha\beta}$ and $\rho_{\gamma\delta}$ which can be estimated from $|\mathbf{R}_{12}| := |\mathbf{R}_2 - \mathbf{R}_1|$. The centres $\mathbf{R}_1$ and $\mathbf{R}_2$ are gauge centres representing the centres of the orbital densities $\rho_{\alpha\beta}$ and $\rho_{\gamma\delta}$ respectively. There is some ambiguity in the definition of these centres; should the centre $\mathbf{R}_1$ sit closer to the centre of ϕ_α or closer to ϕ_β? This problem is addressed in more detail in the the subsection "Assigning gauge centres" where a solution using local orbitals (LO) is introduced to removed the ambiguity.

The local orbitals μ have well defined centres $\mathbf{R}_\mu$ which are used as the gauge centres.

A generic three centre integral $(\alpha\beta|G)$ can be half transformed into a mixed AO/LO representation $(\alpha\mu|G)$ which has a well defined centre $\mathbf{R}_\mu$ leading to half transformed four centre long-range integrals which can be assembled as:

$$\begin{aligned}(\alpha\mu|\gamma\nu)_{lr} = &\sum_{\mathbf{g}}(\alpha\mu|\mathbf{g})\Theta(\mathbf{R}_{\mu\nu})\eta(\mathbf{g})(\gamma\nu|\mathbf{g})^* \\ &+\sum_{mn}(\alpha\mu|m)f^{mn}(\mathbf{R}_{\mu\nu})\bar{\Theta}(\mathbf{R}_{\mu\nu})(\gamma\nu|n)\end{aligned} \tag{3.9}$$

Here the well defined vector $\mathbf{R}_{\mu\nu}$ is used to accurately choose between the Fourier and multipole methods. Once assembled in the half transformed representation the integrals can be transformed back into the AO basis if needed.

The gauge centres $\mathbf{R}_\mu$ are convenient expansion centres for the multipole method and they can be used to regularize the long-range potential, a procedure introduced in chapter 2. Briefly a regularization scheme for the long-range potential was introduced to explicitly remove the singularity in the Fourier transform in an effort to improve the numerical accuracy of the method. The regularization procedure involves the gauge centre $\mathbf{R}$ near the point $\mathbf{r}$.

$$V_{lr}^R(\mathbf{r}) = \int dg^3\eta(\mathbf{g})(e^{i\mathbf{g}\cdot\mathbf{r}} - e^{i\mathbf{g}\cdot\mathbf{R}}) \tag{3.10}$$

The half transformed Fourier integrals then become:

$$(\alpha\mu|g)^R = \int d^3r\phi_\alpha(\mathbf{r})\phi_\mu(\mathbf{r})e^{i\mathbf{g}\cdot\mathbf{r}} - S_{\alpha\mu}e^{i\mathbf{g}\cdot\mathbf{R}_\mu} \tag{3.11}$$

A consequence of the regularization is the introduction of analytical one-body integrals.

$$V_{lr}(\mathbf{R})_{\alpha\beta} = \int d^3r\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})V_{lr}(|\mathbf{r}-\mathbf{R}|) \tag{3.12}$$

The half transformed regularized two-range potential can then be written as:

$$
\begin{aligned}
(\alpha\mu|\gamma\nu)^{R}_{2range} = & \sum_{xy}(\alpha\mu|x)M^{-1}_{xy}(y|\gamma\nu) \\
& + \sum_{g}(\alpha\nu|\mathbf{g})^{R}\Theta(\mathbf{R}_{\mu\nu})\eta(\mathbf{g})(\gamma\nu|\mathbf{g})^{R*} \\
& + \sum_{mn\neq 0}(\alpha\mu|m)f^{mn}(\mathbf{R}_{\mu\nu})\bar{\Theta}(\mathbf{R}_{\mu\nu})(\gamma\nu|n) \\
& + S_{\alpha\mu}V_{lr}(\mathbf{R}_{\mu})_{\gamma\nu} + V_{lr}(\mathbf{R}_{\nu})_{\alpha\mu}S_{\gamma\nu} - S_{\alpha\mu}S_{\gamma\nu}V_{lr}(\mathbf{R}_{\mu\nu})
\end{aligned}
\tag{3.13}
$$

The short-range DF integrals do not need to be assembled in the mixed representation unless it is convenient to do so. The Fourier weights are the same for the regularized and unregularized potentials, only the three centre integrals, $(\alpha\nu|\mathbf{g})^{R}$ are regularized. The regularized multipole expansion can be derived by subtracting the one-body potential from the multipole expansion, effectively removing the the $m,n=0$ terms from the multipole expansion. The fourth line in equation 3.13 is the one-body contribution which is assembled from nuclear-electron like attraction integrals with a charge of one centred at $\mathbf{R}$ and the overlap matrix $S_{\alpha\beta}$. These integrals do not scale linearly with respect to system size but are evaluated analytically and are two index quantities which is relatively inexpensive compared to the three centre integrals. For very large systems these integrals may become a bottleneck.

The potential benefit to using the regularized potential is that contribution from the Fourier/Multipole methods are very small and could perhaps be evaluated with a smaller numerical grid/lower level expansion than their unregularized counterparts, while maintaining similar accuracy increasing the overall efficiency of a calculation. In preliminary geometry optimizations steps the regularized potential might even be neglected completely.

The accuracy of these methods can be studied by comparing the approximated integrals to the analytical four centre integrals. However, in practice the four centre two-electron integrals are never assembled, instead quantities that contribute to the molecular Hamiltonian are assembled through the three centre representations. Therefore to study the accuracy and validity of the range-separated method a four centre integral analysis will only be used for the smallest systems and the Hartree-Fock (HF) method[4,105–107] will be used as the benchmark method for larger systems.

The HF wave function is the cornerstone of *ab initio* electronic structure theory and the optimization of molecular orbitals by the HF method is often the first step in many electronic structure methods. In quantum chemistry, molecular orbitals $|\psi_i\rangle$ are represented as a linear combination of atomic orbitals, or basis functions $|\phi_\alpha\rangle$.

$$|\psi_i\rangle = \sum_\alpha C_{\alpha,i}|\phi_\alpha\rangle \tag{3.14}$$

Here $C_{\alpha,i}$ are the expansion coefficients, these coefficients can be used to form the one electron density matrix.

$$D_{\alpha\beta} = \sum_i C_{\alpha,i}C_{\beta,i} \tag{3.15}$$

The molecular orbitals are obtained by solving the Hartree-Fock-Roothaan equations.

$$FC = SC\epsilon \tag{3.16}$$

Here F is the Fock operator and S is overlap integrals. These non-linear equations are solved iteratively though a procedure called the self-consistent field (SCF) method. The Fock matrix is composed of one and two-electron contributions and the focus will be on the accurate construction of the two-electron contribution to the Fock matrix. The Fock matrix for closed shelled systems in the atomic orbital basis is shown in equation 3.17.

$$F_{\alpha\beta} = H_{\alpha\beta}^{core} + 2J_{\alpha\beta} - K_{\alpha\beta} \tag{3.17}$$

The term $H_{\alpha\beta}^{core}$ denotes the one electron integrals describing the kinetic energy and nuclear attraction of an electron, the second and third terms are the direct and exchange Coulomb integrals contracted with the density matrix.

$$J_{\alpha\beta} = \sum_{\gamma\delta}(\alpha\beta|\gamma\delta)D_{\gamma\delta} \tag{3.18}$$

$$K_{\alpha\beta} = \sum_{\gamma\delta}(\alpha\gamma|\beta\delta)D_{\gamma\delta} \tag{3.19}$$

To tie everything together, the density matrix can be factorized through a pivoted-Cholesky routine to yield local occupied orbital transformation coefficients.

$$D_{\gamma\delta} = \sum_{\mu} C_{\gamma,\mu} C_{\delta,\mu} \tag{3.20}$$

Which can be used to simplify 3.18 and 3.19.

$$J_{\alpha\nu} = \sum_{\mu} (\alpha\nu|\mu\mu) \tag{3.21}$$

$$K_{\alpha\beta} = \sum_{\mu} (\alpha\mu|\beta\mu) \tag{3.22}$$

The exchange is more straightforward to analyse. Using equation 3.9 one can see that for the exchange $\Theta(\mathbf{R}_{\mu\mu})$ is evaluated at 0 always, such that Fourier method is always the method of choice. The direct term uses $\Theta(\mathbf{R}_{\nu\mu})$ as the switch between the Fourier and multipole methods and the half transformed direct term is the result. The direct term ($J_{\alpha\nu}$) must be projected back into the AO basis to be assembled into the AO Fock matrix, which has some additional complications associated with it discussed in the "Analysis of the Integral Accuracy" section.

To summarize the goals of this chapter, the range-separated potentials introduced in chapter 2 are used to partition the two-electron Coulomb integrals into short and long-range contributions. The range-separated integrals are all compact and sparse, the expensive integrals scale linearly with respect to system size. The accuracy of the methods are validated using the four centre integrals for small systems and the HF direct and exchange contributions for larger systems.

3.2 Methods

3.2.1 Notation

The notation used throughout this chapter is summarized below in table 3.1. **Boldface** is used indicate vectors $\mathbf{r}$ is used for electron coordinates and $\mathbf{R}$ is used for gauge centres.

Label	Definition
α,β,γ,δ	General atomic orbitals
i,j,k,l	Canonical occupied orbitals
p,q,r,s	Canonical virtual orbitals
μ,ν,σ,τ	Local occupied orbitals
$\bar{\mu}$,$\bar{\nu}$,$\bar{\sigma}$,$\bar{\tau}$	Local virtual orbitals

Table 3.1: Index Notation.

3.2.2 Computational Details

This subsection is devoted to clarifying the computational details of this chapter. All calculations were carried out in a custom written Python program interfaced with Python-based simulations of chemistry framework (PySCF)[108]. Primitive three centre integrals were calculated in PySCF which uses the Libcint integral library[109], the range-separated integrals and two-electron quantities such as the Fock matrix are assembled in the custom code. At this stage a converged SCF calculation from PySCF is used to check the final results in the new code. Once appropriate procedures have been determined the convergence of Hartree-Fock calculations can be tested. This will be discussed in Chapter 4.

The two properties that we wish to demonstrate for the range-separated potential are the accuracy and the scaling of the integrals with respect to system size. To present the data three different molecular examples are chosen to highlight the key feature of this method.

A simple water dimer is used to demonstrate many of the challenges faced when working with the range-separated potential without overcomplicating the analysis. The distance between the oxygen atoms is increased such that the long-range potential can be studied. The water dimer maybe too simple at times, in this case a similar approach is taken with a stretched glycine dimer. These two stretched dimers are used to show the strengths and weaknesses of the potentials.

To demonstrate integral scaling with respect to system size polyacetylene chains with repeating unit $(C_2H_2)_n$ are used as test molecules. By using a molecule comprised of repeating chains it is easier to demonstrate the scaling of three centre integrals.

To demonstrate general properties four molecules were chosen and are depicted in figure 3.1.

The molecules are glycine, toluene, benzophenone and 12-amino dodecahexanone. Glycine and toluene were chosen to demonstrate how the integrals behaves with small molecules. Benzophenone and 12-amino dodecahexanone are larger molecules to demonstrate how the long-range integrals behave.

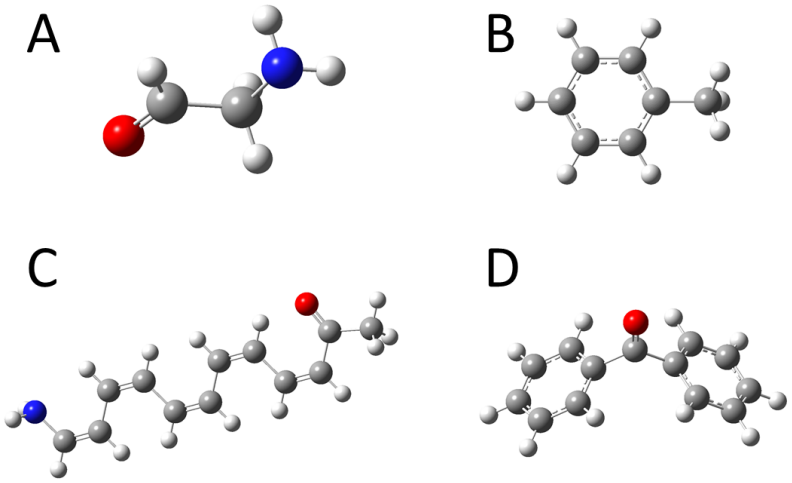

Figure 3.1: Four molecules used to demonstrate features of the range-separated method. A. Glycine. B. toluene. C. Benzophenone. D. 12-amino dodecahexanone.

A more general study on a larger varriety of molecules is presented in chapter 4 where molecules from the GMTKN55 database[110–112] were investigated. The GMTKN55 database is a database of small to medium sized molecules designed to benchmark general main group thermochemistry, kinetics and noncovalent interactions. It consists of some 2400 plus single point calculations which make up approximately 1500 relative energies and it is often used to benchmark the validity of density functionals used in density functional theory. This thesis uses a restricted HF code so only the restricted single point calculations from the GMTKN55 database were performed.

The Python programming language is excellent for fast development of code and testing of new algorithms, but it is very slow compared to other languages and in this case the overhead between the custom code, PySCF and Libcint can be expensive. However, our goal was to create a code to test the validity of the range-separated method and it is not optimized for large scale calculations, therefore for the smaller molecules triple zeta basis sets such as ccpvtz were used with the associated auxiliary fitting set ccpvtz-jkfit but for larger molecules such

as the polychains smaller Pople basis sets such as 6-31g had to be used.

3.2.3 Sparse Matrices

It is important for efficiency and scaling to take advantage of the sparsity of the AO integrals. All three index integrals used throughout this thesis are stored as sparse matrices to reduce amount of memory required for matrix-matrix operations. A sparse matrix is a matrix whose elements are dominated by zeros with only a few non-zero elements. There are several ways to format a sparse matrix. The coordinate (COO) format is the most human readable format where the sparse matrix is sorted in three arrays: data, row and column indices. The COO format is not efficient for matrix-matrix operations but it is easy to understand and work with. There are other formats such as compressed sparse row (CSR) or compressed sparse column (CSC) which can be efficient for sparse matrix-matrix operations if the sparsity structure is random or unknown. In quantum chemistry there is structure to the sparsity and several groups exploit this sparsity to further optimize matrix operations[113,114]. In this thesis COO and CSR formats are used, the COO format is used to make the code more readable and the CSR format is used for matrix operations. Simple transformations are available in the Scipy[115] python libraries to switch between formats.

One of the technical challenges faced in this thesis is that Scipy sparse matrix routines only work for square matrices (two index) but our primitive integrals are three index quantities. The dimensions mismatch is a challenge which is solved by compressing indices in the three index integrals. For example the three index integral $(\alpha\beta|\gamma)$ can be stored as $(\alpha+\beta*n_\alpha|\gamma)$ where n_α is the size of α index. Index compression is used where needed throughout the code.

3.2.4 Local Orbitals

Localised molecular orbitals (LO) are molecular orbitals that occupy a limited spatial region. LO's are an important tool in many electronic structure methods as contributions between spatially distant orbitals can be treated at a lower (cheaper) level of theory, which can drastically increase the efficiency of post Hartree-Fock correlation calculations. For example, the popular local correlation approaches include domain-based local pair natural orbitals (DLPNO)[7–13] and cluster in molecules (CIM)[22–36,116] which all rely on the use of local molecular orbitals.

There are many flavours of orbital localisation schemes available; two of the more popular schemes are the Pipek-Mezey and Boys localisation schemes[117–119]. Boy's localization minimizes the spatial extent of the orbitals while Pipek-Mezey localisation maximizes the orbital dependent partial charges on the molecules nuclei. Both are readily available in many software packages, including PySCF.

In this thesis, the preferred method for obtaining the localisation coefficients is through a pivoted-Cholesky decomposition[120] of the Hartree-Fock density matrix.

$$D_{\alpha\beta} = \sum_{\mu} C_{\alpha\mu} C_{\beta\mu} \tag{3.23}$$

the coefficients can be orthonormalized through Löwdin orthonormalization if required.

$$C_{\alpha\mu} = \sum_{\mu'} C_{\alpha\mu'} M^{1/2}_{\mu'\mu} \tag{3.24}$$

$$M = CSC^{\dagger} \tag{3.25}$$

This is the preferred localisation method because we can introduce the half-transformed three centre integral, $(\alpha\mu|g)$ which is important in construction of the HF direct and exchange contributions.

The rank of the density matrix is the number of electrons (number of electron pairs in restricted calculations) and hence the range of μ is limited to the occupied labels. For the virtual orbitals, the same pivoted-Cholesky procedure can be performed on $\bar{D}_{\alpha\beta}$ which would yield the local virtual orbitals labelled $\bar{\mu}$. Orthogonalizing the local virtual orbitals from Cholesky can lead to small eigenvalues in the inverse metric, therefore orthogonalizing the virtual orbitals should be avoided.

The local orbitals μ have a well defined centre $\mathbf{R}_{\mu}$ which can be obtained via the dipole moments.

$$\mathbf{R}_{\mu} = \frac{(\mu|\mathbf{r}|\mu)}{(\mu|\mu)} \tag{3.26}$$

The well defined centre is essential for assigning the correct gauge centre to a half-transformed integral block $(\alpha\mu|g)$, a problem which is discussed in the "Assigning Gauge Centres" subsection.

The "locality" of an orbital can be quantified by calculating the standard deviation of the orbitals.

$$\begin{aligned}\sigma_x^2 &= (\mu|x^2|\mu) - (\mu|x|\mu)^2 \\ \sigma_y^2 &= (\mu|y^2|\mu) - (\mu|y|\mu)^2 \\ \sigma_z^2 &= (\mu|z^2|\mu) - (\mu|z|\mu)^2 \end{aligned} \tag{3.27}$$

3.2.5 Assigning Tile and Gauge Centres

In chapter 2 gauge centres are introduced as an essential part of the treatment of the long-range potential. The gauge centres have multiple uses, they are used to regularize the Fourier integrals, as expansion centres for the multipole representation, distance metric for switching between multipole and Fourier methods and finally the gauge centres are used as part of a coarse grained pre-screening described in the "Integral pre-screening" section. In terms of integral pre-screening the gauge centres are referred to as tile centres. The definition of a gauge centre ($\mathbf{R}_i$) in chapter 2 is a point near the point $\mathbf{r}_i$. In this chapter we are concerned about atomic basis functions which are not points and therefore the gauge centres need to be redefined as points near the centre of these atomic orbitals. Atomic orbitals are atom centred orbitals and the centre is always the nuclear coordinate, therefore gauge centres can be defined as points near the nuclear coordinates.

It is not required to have a gauge centre on each nuclear coordinate if atoms are close enough one gauge centre would be sufficient to represent all of the nearby orbitals. For example, in a water molecule it is sufficient to place a gauge centre at coordinate of the oxygen atom as the hydrogens are close enough. However, for more complicated systems such as polyethylene it may be sufficient to place a gauge centre between each carbon pair or on each carbon atom themselves. The placement of gauge centres is still ambiguous.

Figure 3.2 depicts a simple algorithm for assigning AO gauge centres. The first step (panel A) is to introduce gauge centres as a three dimensional grid of points equidistantly separated

by $1-2$ angstroms (adjustable parameter T_{sep}). The second step is to place the molecule of interest in this grid then assign each nucleus nearest gauge centre. If no atoms are assigned to a gauge centre that centre is removed as shown in panel C. If multiple atoms are assigned to the same gauge centre, that gauge centre is migrated to the centre of charge of that particular group of atoms shown in panel D which is the resulting AO gauge centres.

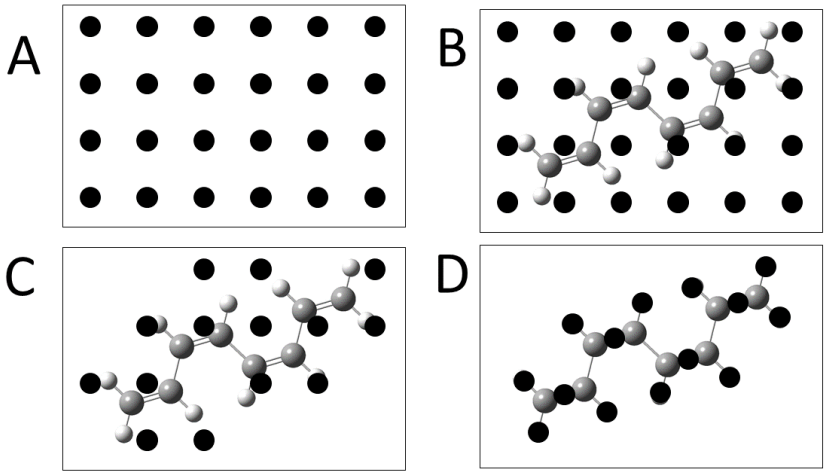

Figure 3.2: A pictorial representation of the algorithm used to assign atomic orbital gauge centres.

Each AO label can be assigned to a gauge centre. Capital letters are used to identify the gauge centre in the Mulliken notation.

$$(\alpha_A \beta_B | \gamma_C \delta_D) \tag{3.28}$$

In the evaluation of the four centre two-electron integrals there are potentially four unique gauge centres based on the current definition, but there can only be two centres used for the regularization. For the orbital pair $\alpha\beta$, what vector should be used for $\mathbf{R}_1$? There are three possibilities to consider, $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$ or $\mathbf{R}_{\alpha\beta} := \frac{\mathbf{R}_\alpha + \mathbf{R}_\beta}{2}$, which leads to three situations.

- If α,β are close: $\mathbf{R}_\alpha$,$\mathbf{R}_\beta$ and $\mathbf{R}_{\alpha\beta}$ are all sufficient.
- If α,β are far apart: The integral is zero because α,β decays exponentially due to the Gaussian product theorem.
- If α,β are far but not too far: $\mathbf{R}_\alpha$ is OK for ϕ_α but poor for $\mathbf{R}_\beta$. $\mathbf{R}_{\alpha\beta}$ is poor choice

for both α,β.

Integrals that fall in the third situation are the hardest to represent accurately as there is no good solution to the placement of the gauge centre. To keep choices simple $\mathbf{R}_\alpha$ will always be used. Note that this breaks some of the symmetry in the four centre integrals as $(\alpha\beta_A|\gamma\delta_C) \approx (\beta\alpha_B|\gamma\delta_C)$. This error can be used as a metric to judge the accuracy of the method.

A solution to the ambiguity of gauge centre placement that works very well with the assembly of the Fock matrix is to introduce local orbitals and the half transformed integral, written in a mixed AO/LO representation.

$$(\alpha\mu|\gamma\nu) = \sum_{\beta\delta} (\alpha\beta|\gamma\delta) C_{\beta\mu} C_{\delta\nu} \tag{3.29}$$

Here the gauge centres will be defined as $\mathbf{R}_\mu$ and $\mathbf{R}_\nu$, their centres are defined by equation 3.26.

For the direct term the vector $\mathbf{R}_{\mu\nu} := \mathbf{R}_\mu - \mathbf{R}_\nu$ can be used to determine the switching between multipole and Fourier methods. The switch can be improved by including the orbital extents, equation 3.27. The orientation of the extents that minimizes $\mathbf{R}_{\mu\nu}$ is the orientation used for determining switching.

To summarize, there are two sets of gauge centres calculated for each molecule $\mathbf{R}_\alpha$ and $\mathbf{R}_\mu$. The AO gauge centres $\mathbf{R}_\alpha$ are defined once at the start of the calculation and are used to partition the integrals into sizeable blocks that are easily handled in memory. The local gauge centres $\mathbf{R}_\mu$ are calculated once per SCF iteration and are used in the assembly of the long-range integrals as the distance metric for $\Theta(\mathbf{R}_{\mu\nu})$. They are also used as expansion centres for the multipole method and are essential for regularizing the Fourier integrals.

3.2.6 Integral Pre-screening

All integral calculations should be pre-screened to avoid the calculation of integrals that evaluate to zero. Integral pre-screening is imperative for the calculation of the primitive three centre integrals because they are all sparse objects and their low order scaling is a result of their sparsity. In practice it would be ideal to screen all AO integrals by their basis function

shells. However, in our code, there is too much overhead between the Libcint integral package (written in C) and the python code. Instead, a course grained pre-screening algorithm is used focused on the AO gauge centres. In the context of integral pre-screening tile centres are better a name for the AO gauge centres.

The AO indices and hence their basis function shells are mapped to their nearest tile centres based off a tile separation threshold. The smaller the threshold the more tile centres used, maxing out at one tile centre for each nuclear coordinate. The three centre integrals can then be written with a tile centre subscript, $(\alpha_A\beta_B|x_X)$, this is referred to as a block of integrals. In the screening algorithm, any quantities referencing a block are referencing the maximum value on any basis shell in that block.

The screening that is implemented is a modification to the tight distance dependent estimator proposed by Hollman *et al.* in 2015[65]. The reason for the modification is due to the block screening and the range-separated long-range potential.

The simplest estimator invokes the Schwarz inequality[121].

$$(\alpha\beta|\gamma\delta) \leq |(\alpha\beta|\alpha\beta)|^{1/2}|(\gamma\delta|\gamma\delta)|^{1/2} \tag{3.30}$$

This inequality holds true because equation 3.30 is positive-definite. However, while this is an upper bound, the Schwarz screening is not a tight bound. A better screening algorithm would use a reliable tight bound with conservative thresholds and the bound doesn't necessarily need to be an upper bound[122]. Screening the short-range three centre integrals with the Schwarz inequality would yield:

$$(\alpha_A\beta_B|x_X) \leq |(\alpha_A\beta_B|\alpha_A\beta_B)|^{1/2}|(x_X|x_X)|^{1/2}$$

$$(\alpha_A\beta_B|x_X) \leq Q_{AB}Q_X \tag{3.31}$$

$$Q_{AB} = \mathrm{Max}(|(\alpha_A\beta_B|\alpha_A\beta_B)|^{1/2}) \tag{3.32}$$

$$Q_X = \mathrm{Max}(|(x_X|x_X)|^{1/2}) \tag{3.33}$$

The above inequality is lacking information about the distance between blocks AB and blocks

X. To add in information about the radial distance between the blocks the vector $\mathbf{r}_{ABX}$ is introduced.

$$\begin{aligned} \mathbf{r}_{ABX} &= \min(r_{AX}, r_{BX}) \\ \mathbf{r}_{AX} &= |R_A - R_X| - \text{ext}_A - \text{ext}_B \end{aligned} \tag{3.34}$$

The maximum extent of an orbital ext_A is included in the definition of $\mathbf{r}_{ABX}$. The screening estimator is then:

$$\begin{aligned} &\text{if } \mathbf{r}_{ABX} > 0\text{:} \\ &\qquad (\alpha_A \beta_B | x_X) \approx Q_{AB} Q_X V_{sr}(r_{ABX}) \\ &\text{if } \mathbf{r}_{ABX} < 0\text{:} \\ &\qquad (\alpha_A \beta_B | x_X) \approx Q_{AB} Q_X \end{aligned} \tag{3.35}$$

The estimator in equation 3.35 results in a tight screening that is not an upper bound. If a conservative screening threshold is used the block screening can be fairly efficient. To demonstrate the tight bound the short-range three centre block integrals $(\alpha_A \beta_B | x_x)$ were calculated for the polyacetlyene molecule with 16 carbon atoms in the def2tzvp basis (def2tzvp-jk auxiliary basis). A log-log plot of the actual verse estimated block value is depicted in figure 3.3. The features to note in figure 3.3 is that the actual values are fairly close to the one-one ratio (solid black line), with a slight bias towards the conservative side.

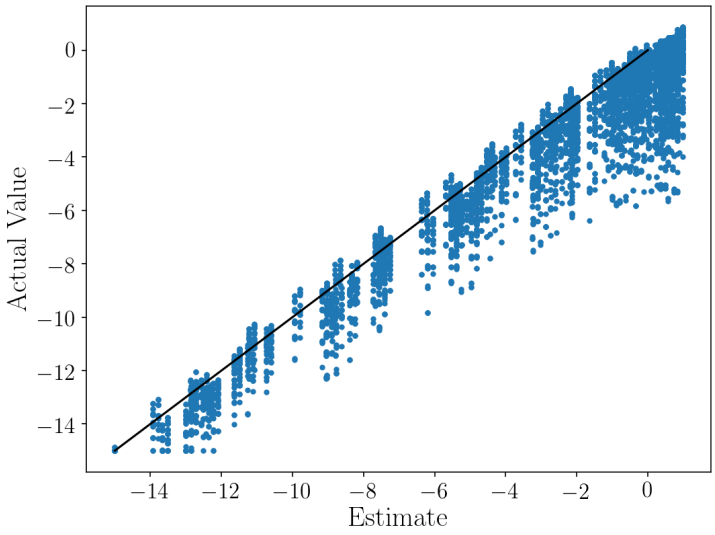

Figure 3.3: Log-Log plot of the estimated value and actual short-range integral value for the C16 polyacetylene chain in the def2tzvp basis.

The tile separation threshold (T_{sep}) is the control between the number of integral blocks and the rough size of each integral block, this parameter is used to find a compromise between memory usage and run time for the calculation of the short-range three centre integrals. Data for the polyacetlyene molecule with 20 carbon atoms are tabulated in table 3.2. The numerical quantities required for pre-screening are calculated once at the beginning of the calculation and stored in memory. These values are per block which totals to the number of tile centres cubed. The timings reported in table 3.2 are the time it takes to evaluate equation 3.35 and call the C library for a block of integrals if the estimator is larger than the pre-screen threshold (10^{-8} in this case). It can be seen that it is expensive to call a block of integrals which is mainly due to the python-C overhead. The most efficient block screening is with a tile centre on each atom (42 gauge centres) which screens out about 54 percent of all three centre integrals. However this calculation takes about 13 times longer than when 18 tile centres are used which only screens about 45 percent of all integrals. The average size of an integral block is also reported, ideally the integral blocks should not be too large as the code starts to run into memory issues. As a compromise between speed and memory a tile separation of somewhere between 2.5 and 5$\mathring{A}$ is used. The parameter T_{sep} may need further tuning for genuine 3d molecules rather than the 2d structures used here.

T_{sep} ($\AA$)	# of Tile Centres	% of Integrals Screened	Average Ints./Block	Time (s)
0	42	54.2	588	2125.6
2.5	18	45.8	8845	170.0
5	5	24.5	574417	11.4
10	3	15.5	2979107	9.7
100	1	0.0	95133696	9.8

Table 3.2: Tabulated data on the screening of the C20 molecule with α=0.4 in the def2tzvp basis.

A similar technique is used to screen the long-range Fourier integrals, resulting in the following estimator.

$$(\alpha_A \beta_B | g) \leq Q_{AB} \sqrt{w_g} \tag{3.36}$$

Figure 3.4 depicts log-log plots comparing the estimated integral value to the actual integral value for the C16 polyacetlyene chain in the def2tzvp basis. When the numerical weights are not included the estimator is not a tight bound but it is an upper bound. Using the full numerical weights for screening yields an estimate that is slightly too large. The square root of the weights is the best estimator because it is the tightest bound. Conservative pre-screening values of 10^{-8} to 10^{-11} should be used to ensure integrals are accurately screen to approximately 10^{-7} in magnitude.

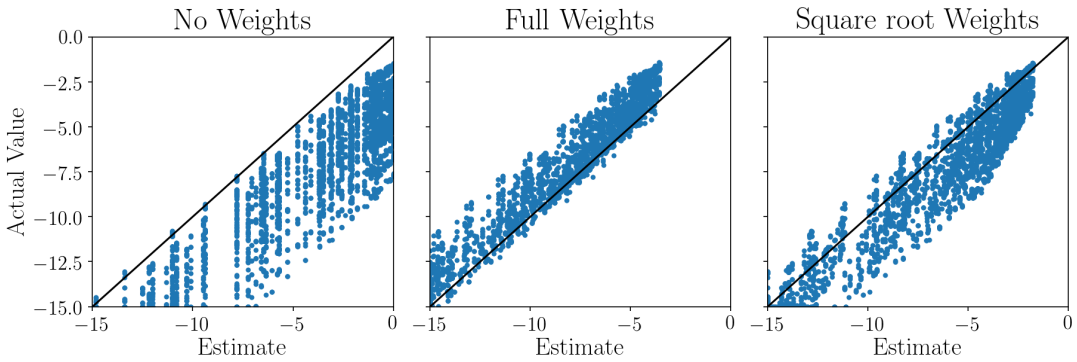

Figure 3.4: Log-Log plot of the estimated value and actual integral value, for the C16 polyacetlyene chain in the def2tzvp basis.

3.3 Integral Definitions

3.3.1 Short-Range Analytical Integrals

The short-range integrals are represented through density fitting, which is commonly used technique in many quantum chemistry packages. To briefly summarise, an auxiliary basis set is chosen such that equation 3.1 can be written as:

$$(\alpha\beta|\gamma\delta) = \sum_{xy} (\alpha\beta|x)(x|y)^{-1}(y|\gamma\delta) \tag{3.37}$$

Where

$$(\alpha\beta|x) = \int dr_1 dr' \phi_\alpha(\mathbf{r}_1)\phi_\beta(\mathbf{r}_1)V_{sr}(\mathbf{r}_1,\mathbf{r}')\chi_x(\mathbf{r}') \tag{3.38}$$

$$(x|y) = \int dr' dr'' \chi_x(\mathbf{r}')V_{sr}(\mathbf{r}',\mathbf{r}'')\chi_y(\mathbf{r}'') \tag{3.39}$$

The metric matrix $(x|y)$ shall be relabelled as M_{xy}, and for density fitting the inverse metric matrix is required. It is also beneficial to decompose the metric matrix into two symmetric halves ($M^{-1}_{xy} = m_{xl}m_{yl}$) such that equation 3.37 can be written as:

$$(\alpha\beta|\gamma\delta) = \sum_{x,y,l} (\alpha\beta|x)m_{xl}m^\dagger_{yl}(y|\gamma\delta) \tag{3.40}$$

$$= \sum_{l} (\alpha\beta|l)(l|\gamma\delta) \tag{3.41}$$

Which simplifies the evaluation of the short-range integrals. There are a few ways to perform the decomposition of the metric matrix, such as using the matrix square root or Cholesky decomposition. The Cholesky decomposition is the preferred method because it is more easily implemented and for tight thresholds more sparse than the square root method. However, for looser thresholds the opposite is true.

Data for the total number of nonzero $(\alpha\beta|l)$ integrals using both the square root and Cholesky methods for factoring the metric matrix are presented in figure 3.5. A linear polyacetylene chain was used as the example molecule in the 6-31g basis, the integrals are stored in a sparse

matrix, the data plotted is the length of the data array. A threshold of 10^{-15} (solid lines) and a threshold of 10^{-8} (dashed lines) were used to prune the sparse matrices. There is a sizable difference in the amount of data between the methods with no notable loss in accuracy, therefore pruning the sparse matrix data is essential.

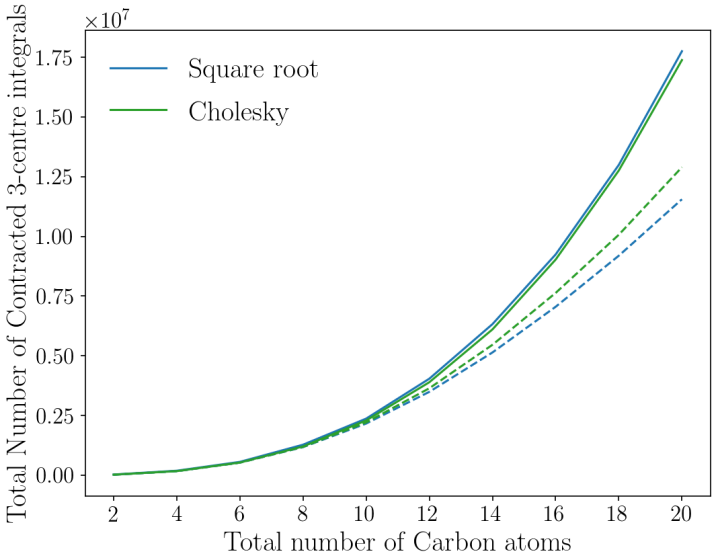

Figure 3.5: Total number of non zero $(\alpha\beta|l)$ integrals. The solids lines are with a pruning thresh of 10^{-15} and the dashed lines are with a pruning thresh of 10^{-8}.

The short-range primitive integrals scale linearly with respect to system size because the $\alpha\beta$ range is limited, as it decays exponentially due to the Gaussian product theorem. The auxiliary basis centre has to be close to the orbital density $\rho_{\alpha\beta}$ because of the short-range potential used. The linear scaling of the three centre integrals is not clear in figure 3.5 because the system has to reach a certain size for linear scaling to become clear, this is shown in the "Integral Scaling" subsection.

The DF Fock matrix terms used in the accuracy analysis are:

$$J_{\alpha\beta} = \sum_{\gamma,\delta,l} (\alpha\beta|l)(l|\gamma\delta) D_{\gamma\delta} \tag{3.42}$$

$$K_{\alpha\beta} = \sum_{\gamma,\delta,l} (\alpha\gamma|l)(l|\beta\delta) D_{\gamma\delta} \tag{3.43}$$

3.3.2 Long-Range Fourier Integrals

The two-electron integrals can be evaluated using the Fourier representation of the Coulomb potential as:

$$(\alpha\beta|\gamma\delta) = \int dg^3 (\alpha\beta|\mathbf{g}) f(\mathbf{g}) (\mathbf{g}|\gamma\delta) \tag{3.44}$$

$$f(\mathbf{g}) = \frac{4\pi}{(2\pi)^3} \frac{1}{|\mathbf{g}|^2} \tag{3.45}$$

$$(\alpha\beta|\mathbf{g}) = \int d^3 r \phi_\alpha^*(\mathbf{r}) \phi_\beta(\mathbf{r}) e^{i\mathbf{g}\cdot\mathbf{r}} \tag{3.46}$$

Where equation 3.46 denotes the Fourier transform atomic orbitals (FTAO) integrals, which are complex integrals. Let us first replace the Coulomb potential with the long-range half of the two-range potential. The two-electron integral becomes:

$$(\alpha\beta|\gamma\delta)_{lr} = \int dg^3 (\alpha\beta|\mathbf{g}) f(\mathbf{g}) (\mathbf{g}|\gamma\delta) \tag{3.47}$$

$$f(\mathbf{g}) = \frac{4\pi}{(2\pi)^3} \frac{1}{|\mathbf{g}|^2} e^{-\gamma|\mathbf{g}|^2} \tag{3.48}$$

The only difference between the equations 3.44 and 3.47 is the function $f(\mathbf{g})$. The integrals over the atomic orbital basis functions, equation 3.46, are identical in both equations. The function $f(\mathbf{g})$ that contains all the information about the long-range potential. This is very convenient because the numerical evaluation of equation 3.47 requires the continuous vector $\mathbf{g}$ and function $f(\mathbf{g})$ to be represented as numerical points and weights, the FTAO integrals are evaluated analytically (integral over r). The FTAO integrals are also readily available in many integral codes such that evaluating these integrals over new potentials is not a difficult development task. To evaluate new potentials one simply needs to calculate new points and weights to adjust the numerical quadrature.

The numerical representation of equation 3.47 using spherical grid points $\mathbf{g}_i$ is shown in equation 3.49.

$$(\alpha\beta|\gamma\delta)_{lr} = \sum_i (\alpha\beta|\mathbf{g}_i) w_i (\mathbf{g}_i|\gamma\delta) \tag{3.49}$$

In practice one would split the weights in two and attach them to either side of the three centre AO integrals. This is usually very straight forward and a simple square root of the weights will work for most potentials as AO integrals are complex numbers but the weights are semi-definite positive. However, when the regularized potential from chapter 2 is introduced, the function $\eta(\mathbf{g})$ might have negative values. This is a technical problem that is easily solved by factoring the weights into magnitude and phase.

$$(\alpha\beta|\gamma\delta)_{lr} = \sum_i (\alpha\beta|\mathbf{g}_i)|w_i|p_i(\mathbf{g}_i|\gamma\delta) \tag{3.50}$$

$$= \sum_i (\alpha\beta|\mathbf{g}_i)\sqrt{|w_i|}p_i\sqrt{|w_i|}(\mathbf{g}_i|\gamma\delta) \tag{3.51}$$

$$= \sum_i I_{\alpha\beta}(\mathbf{g}_i)p_i I^*_{\gamma\delta}(\mathbf{g}_i) \tag{3.52}$$

Here the phase p_i has only three values -1,0 or $+1$ and through presorting the grid points according the the sign of the phase the above equation becomes:

$$(\alpha\beta|\gamma\delta)_{lr} = \sum_{i,p_i=1} I_{\alpha\beta}(\mathbf{g}_i)I^*_{\gamma\delta}(\mathbf{g}_i) - \sum_{i,p_i=-1} I_{\alpha\beta}(\mathbf{g}_i)I^*_{\gamma\delta}(\mathbf{g}_i) \tag{3.53}$$

To maximize efficiency only unique FTAO integrals are calculated. These integrals are $I_{\alpha\beta}(\mathbf{g}_i)$ for $\alpha \geq \beta$, and they are pre-screened using the magnitude of the square root of the weights in conjunction with the overlap matrix.

Regularizing the Fourier transform involves introducing the gauge centre $\mathbf{R}_\alpha$ defined as a point near $\mathbf{r}$. Similar to how the potentials are regularized in the previous chapter the regularized three centre FTAO integrals are:

$$\begin{aligned}(\alpha\beta|\mathbf{g})^R &= \int d^3r\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})(e^{i\mathbf{g}\cdot\mathbf{r}} - e^{i\mathbf{g}\cdot\mathbf{R}_\alpha}) \\ &= \int d^3r\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})e^{i\mathbf{g}\cdot\mathbf{r}} - \int d^3r\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})e^{i\mathbf{g}\cdot\mathbf{R}_\alpha} \\ &= (\alpha\beta|\mathbf{g}) - S_{\alpha\beta}e^{i\mathbf{g}\cdot\mathbf{R}_\alpha}\end{aligned} \tag{3.54}$$

Therefore equation 3.49 becomes:

$$(\alpha\beta|\gamma\delta)_{lr} = \sum_i (\alpha\beta|\mathbf{g}_i)^R w_i (\mathbf{g}_i|\gamma\delta)^{R*} \tag{3.55}$$

In summary, the regularized FTAO's are obtained by subtracting phase shifted overlap integrals. The pure Fourier integrals and regularized Fourier integrals are both written in the same compact notation. Therefore, it very easy to switch between pure FT and regularized FT in a calculation. This can be done for the whole system or for individual orbital pairs.

The Fourier representation is only good when the orbital densities $\rho_{\alpha\beta}$ and $\rho_{\gamma\delta}$ are not very far apart, up to about $10\mathring{A}$. The accuracy of the Fourier integrals breaks down for large $\mathbf{g}\cdot\mathbf{r}_{12}$, which will always occur if the system is large enough. The breakdown arises from the phase factor $e^{i\mathbf{g}\cdot\mathbf{r}_{12}}$ which oscillates very rapidly and is hard to integrate accurately for large $\mathbf{r}_{12}$. Therefore the Fourier integrals need to be turned off and the long-range multipole integrals need to be turned on at some point.

Introducing the half transformed representation the Fock contributions can be written down with a well defined distance metric $\mathbf{R}_{\mu\nu}$ which is used even if regularization is not.

$$J_{\alpha\mu} = \sum_{\nu,g} (\alpha\mu|\mathbf{g})(\mathbf{g}|\nu\nu) \tag{3.56}$$

$$K_{\alpha\beta} = \sum_{\nu,g} (\alpha\nu|\mathbf{g})(\mathbf{g}|\beta\nu) \tag{3.57}$$

The key feature to the exchange is that the distance vector $\mathbf{R}_{\nu\nu}$ is zero always. Therefore the construction of the Fourier exchange will always be accurate and the numerical grid will not need to increase with respect to system size.

3.3.3 Long-Range Multipole Integrals

In chapter 2 we derived an expression for the multipole expansion of a radial potential in Cartesian coordinates as:

$$V\big[(\mathbf{y}-\mathbf{x})+(\mathbf{R}_y-\mathbf{R}_x)\big] = \sum_{k,l} m_y^k f^{(k,l)}(\mathbf{R}) m_x^l \tag{3.58}$$

Atomic basis functions can now be introduced such that equation 3.58 can be written as:

$$(\alpha\beta|\gamma\delta) = \sum_{k,l} (\alpha\beta|k,\mathbf{R}_y) f^{(k,l)}(\mathbf{R}_{xy}) (\alpha\beta|l,\mathbf{R}_x) \tag{3.59}$$

$$(\alpha\beta|k,\mathbf{R}_y) = \int (d^3y) \phi^*_\alpha(\mathbf{R}_y) \phi_\beta(\mathbf{R}_y) \mathbf{y}^k \tag{3.60}$$

The integrals $(\alpha\beta|k,\mathbf{R}_y)$ are the local Cartesian moment integrals and these are easily obtained by shifting the global moment integrals which are:

$$\tilde{S}_{\alpha\beta} = (\alpha|\beta) \tag{3.61}$$

$$\tilde{D}_{\alpha\beta} = (\alpha|\mathbf{r}|\beta) \tag{3.62}$$

$$\tilde{Q}_{\alpha\beta} = (\alpha|\mathbf{r}^2|\beta) \tag{3.63}$$

$$\tilde{O}_{\alpha\beta} = (\alpha|\mathbf{r}^3|\beta) \tag{3.64}$$

Using the tilde to identify global moments and substituting $\mathbf{r} = \mathbf{x} + \mathbf{R}_x$ the local moments can be obtained.

$$S_{\alpha\beta} = \tilde{S}_{\alpha\beta} \tag{3.65}$$

$$D_{\alpha\beta} = \tilde{D}_{\alpha\beta} - \mathbf{R}_x \tilde{S}_{\alpha\beta} \tag{3.66}$$

$$Q_{\alpha\beta} = \tilde{Q}_{\alpha\beta} - \sum_{mn} \left[(\mathbf{R}_x \tilde{D}_{\alpha\beta})_{mn} + (\mathbf{R}_x \tilde{D}_{\alpha\beta})_{nm} \right] + \mathbf{R}_x \mathbf{R}_x \tilde{S}_{\alpha\beta} \tag{3.67}$$

$$\begin{aligned} O_{\alpha\beta} = \tilde{O}_{\alpha\beta} &- \sum_{lmn} \left[(\mathbf{R}_x \tilde{Q}_{\alpha\beta})_{lmn} + (\mathbf{R}_x \tilde{Q}_{\alpha\beta})_{nlm} + (\mathbf{R}_x \tilde{Q}_{\alpha\beta})_{mnl} \right] \\ &+ \sum_{lmn} \left[(\mathbf{R}_x \mathbf{R}_x \tilde{D}_{\alpha\beta})_{lmn} + (\mathbf{R}_x \mathbf{R}_x \tilde{D}_{\alpha\beta})_{nlm} + (\mathbf{R}_x \mathbf{R}_x \tilde{D}_{\alpha\beta})_{mnl} \right] + \mathbf{R}_x \mathbf{R}_x \mathbf{R}_x \tilde{S}_{\alpha\beta} \end{aligned} \tag{3.68}$$

The function $f^{(k,l)}$ denotes the derivatives of the potential. In the regularized multipole the values of k,l are restricted such that they are not zero. Note this leads to gain of one free multipole order. For example, the regularized multipole can be calculated at hexapole accuracy using only the octupole AO moment integrals and the 4th derivatives of the potential. This is an advantage because the derivatives are cheap to calculate but the multipole moment integrals can be sizeable.

The above method can be used for the assembly of any radial potential that has well behaved derivatives. From here on, the long-range part of the two-range potential will be used. The derivatives of the long-range potentials are well behaved even at $\mathbf{r}=0$. Here the full Cartesian expansion is presented, but in practice only the unique degrees of freedom would be calculated.

The contributions to the Fock matrix are:

$$J_{\alpha\mu}=\sum_{\nu,k,l}(\alpha\mu|k,\mathbf{R}_\mu)f^{k,l}(\mathbf{R}_{\mu\nu})(l,\mathbf{R}_\nu|\nu\nu) \tag{3.69}$$

$$K_{\alpha\beta}=\sum_{\nu,k,l}(\alpha\nu|k,\mathbf{R}_\mu)f^{k,l}(0)(l,\mathbf{R}_\nu|\beta\nu) \tag{3.70}$$

The multipole method is never very accurate at small $\mathbf{R}_{\mu\nu}$ therefore the exchange is always calculated using the Fourier method. The multipole contribution to the direct term is only used when $\mathbf{R}_{\mu\nu}$ is large.

3.3.4 Analytical One-Body Integrals

The one-body integrals are a consequence of the regularization procedure and are only used when regularized long-range integrals are used. The one-body potential from chapter 2 is defined as:

$$V_{1body}=V_{lr}(|\mathbf{r}_2-\mathbf{R}_1|)+V_{lr}(|\mathbf{R}_2-\mathbf{r}_1|)-V_{lr}(|\mathbf{R}_2-\mathbf{R}_1|) \tag{3.71}$$

Including the AO basis functions the one-body contribution can be written as an outer product of one-body terms.

$$(\alpha\beta|\gamma\delta)_{1body}=S_{\alpha\beta}V_{\gamma\delta}(\mathbf{R}_\alpha)+V_{\alpha\beta}(\mathbf{R}_\gamma)S_{\gamma\delta}-S_{\alpha\beta}S_{\gamma\delta}V(\mathbf{R}_\alpha-\mathbf{R}_\gamma) \tag{3.72}$$

Physically, the integrals $V_{\alpha\beta}(\mathbf{R}_\gamma)$ are nuclear-electron attraction integrals with a charge centred at $\mathbf{R}_\gamma$ and as always $S_{\alpha\beta}$ is the AO overlap matrix.

$$V_{lr}(\mathbf{R}_\gamma)_{\alpha\beta}=\int d^3r\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})V_{lr}(|\mathbf{r}-\mathbf{R}_\gamma|) \tag{3.73}$$

The direct and exchange contributions are assembled using local gauge centres $\mathbf{R}_\mu$ and $\mathbf{R}_\nu$.

$$J_{\alpha\mu} = \sum_{\nu} S_{\alpha\mu} V_{\nu\nu}(\mathbf{R}_\mu) + V_{\alpha\mu}(\mathbf{R}_\nu) S_{\nu\nu} - S_{\alpha\mu} S_{\nu\nu} V(\mathbf{R}_\mu - \mathbf{R}_\nu) \tag{3.74}$$

$$K_{\alpha\beta} = \sum_{\nu,k,l} S_{\alpha\nu} V_{\beta\nu}(\mathbf{R}_\nu) + V_{\alpha\nu}(\mathbf{R}_\nu) S_{\beta\nu} - S_{\alpha\nu} S_{\beta\nu} V(0) \tag{3.75}$$

3.4 Integral Scaling

The scaling of the data with respect to systems size is one of the motivations behind this research. In particular the scaling of the short-range density fit and long-range Fourier integrals. To convince the reader that the data scales as advertised, polyacetylene, which has repeating units of $(C_2H_2)_n$, is chosen as the test molecule. Owing to the code limitations a chain length of 12 units or 24 carbon atoms is the largest chain available using the ccpvtz basis with ccpvtz-jkfit auxiliary basis. In some cases the 6-311g basis needed to be used.

3.4.1 Three Centre Integrals

There are three types of three center integrals that are considered: short-range DF, long-range FT and multipole integrals. The DF and FT integrals are further separated into two types: primitive and contracted. Primitive integrals are the bare three centre AO integrals that are returned from the integral library. Contracted integrals are primitive integrals that have been contracted with the middle term, for DF integrals this is the square root of the inverse metric and for FT integrals it is the square root of the weights. It is important to consider both types of integrals because their sparsity is different and expensive contractions like the half-transformation are more efficient to perform on sparser matrices.

All integrals are stored as sparse matrices and the data plotted in this section is the total number of non-zero integrals. In practice many of the integrals are non-zero but extremely small and do not significantly contribute to the accuracy of the method. Essentially they can be set to zero. The three centre integrals are pruned according to a threshold T_{sparse} which typically is set to a small value between 10^{-8} and 10^{-15}.

Two parameters that control the linear scaling or the onset of linear scaling are T_{sparse} and

α, the range separation parameter. Figure 3.6 depicts how the total number of short-range integrals scales with the sparsity threshold for the polyacetylene chain in the 6-31g basis. One can see how large of an effect the sparse threshold has on the system's scaling with respect to number of carbon atoms. It should be noted that all three thresholds do not effect the accuracy of the Hartee-Fock direct and exchange contributions in any meaningful way. The accuracy analysis is more thoroughly covered in the next section and all data presented from here on are screened with a sparsity threshold of 10^{-9}.

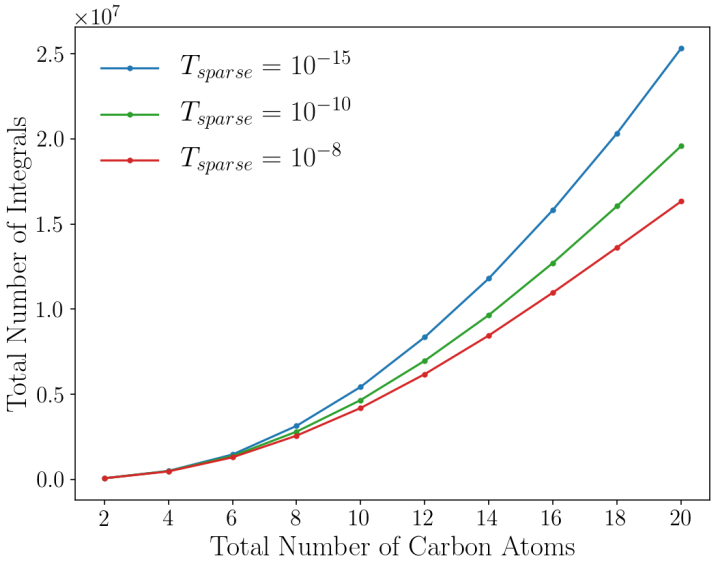

Figure 3.6: Total number of integrals above the sparsity thresholds for the ployacetylene chain system in the 6-31g basis with α=0.2.

The range separation parameter is essential to control the onset of linear scaling for the short-range integrals. Although the parameter does not affect the amount of primitive long-range Fourier or multipole integrals, it does effect the weighted Fourier integrals and regularized Fourier integrals. Figure 3.7 depicts the total number of primitive short-range integrals (left) and the total number of weighted long-range Fourier integrals (right). The amount of short-range integrals can be reduced significantly with increasing α values. Only the weights of the numerical grid are effected by α such that the effect of α on the Fourier integrals is less significant. In the next section "Analysis of the Integral Accuracy" we will learn that the accuracy of the Fourier integrals decreases with increasing α. A trade off exists between total number of integrals and accuracy which can be controlled through α.

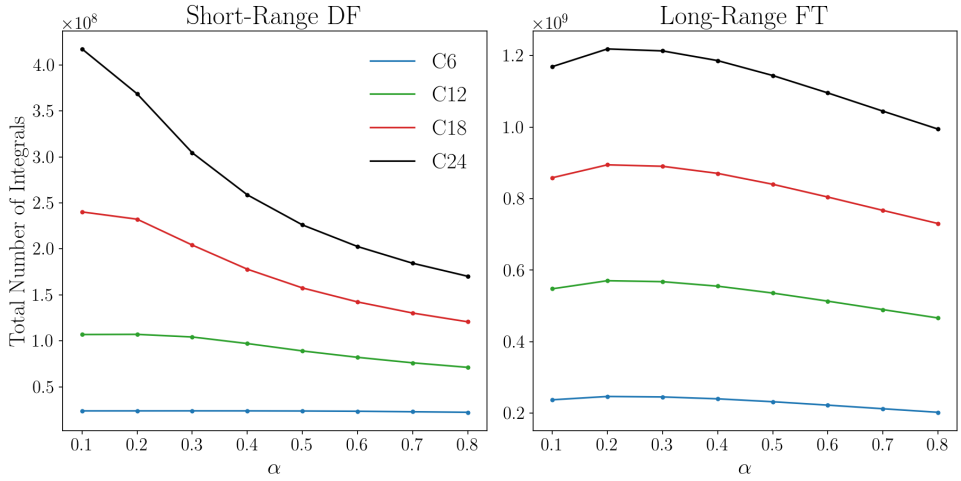

Figure 3.7: Total number of integrals above the sparsity thresholds for the ployacetylene chain system in the ccpvtz-basis basis. The short-range integrals are primitive integrals and the Fourier integrals are weighted integrals. A sparsity threshold of 10^{-9} *was used.*

Linear scaling for the short-range integrals can be seen in figure 3.8 which is dependent on the value of α. For $\alpha = 0.1$ the linear scaling begins around 8 chain lengths and for $\alpha = 0.5$ the onset of linear scaling is closer to 4 chain lengths.

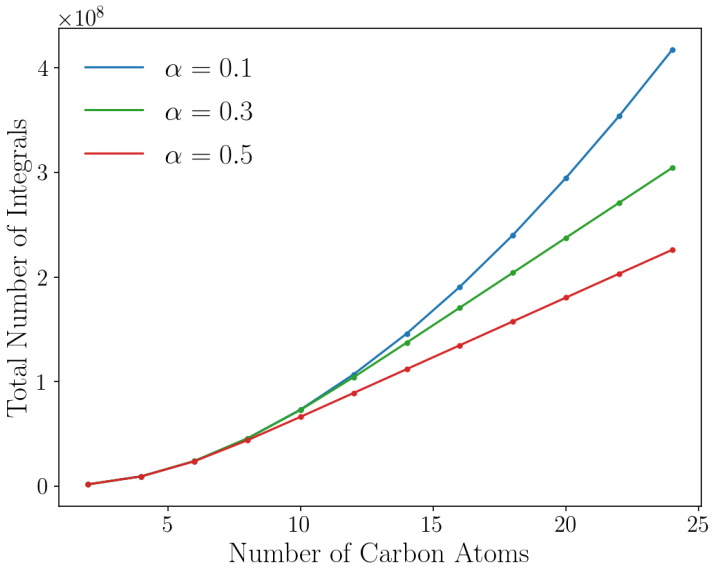

Figure 3.8: Total number of short-range integrals for the ployacetylene chain system in the ccpvtz-basis basis. One can see how the onset of linear scaling can be controlled through the value of α*. A sparsity threshold of* 10^{-9} *was used.*

The difference between the primitive and contracted short-range integrals is significant. For small α values the contracted integrals are more sparse than the primitive integrals, the opposite

is true for large α values and larger systems as shown in figure 3.9.

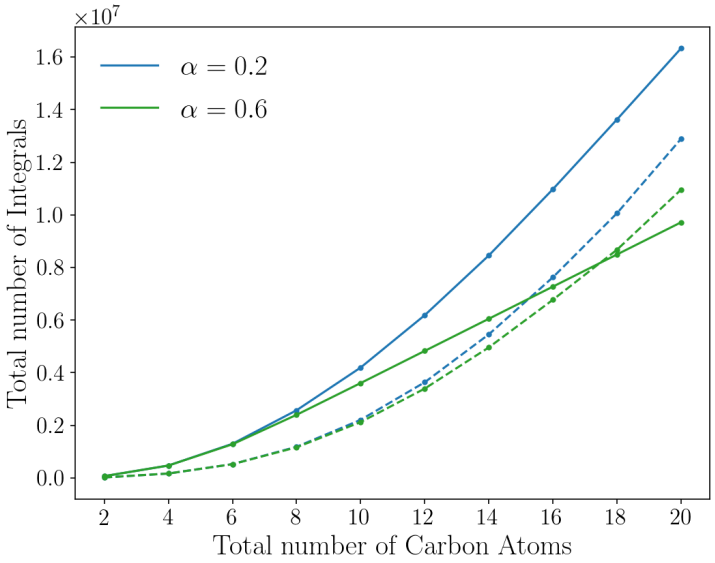

Figure 3.9: Total number of short-range integrals above the 10^{-9} sparsity threshold for polyacetylene in 6-31g basis. The solid lines represent the primitive integrals and the dashed lines represent the contracted integrals.

All Fourier integrals scale naturally with respect to system size, as shown in figure 3.10. The three types of Fourier integrals are primitive, weighted and regularized. The regularized integrals are significantly more sparse than the primitive and weighted integrals which is advantageous if efficient screening and sparse matrix operations are used.

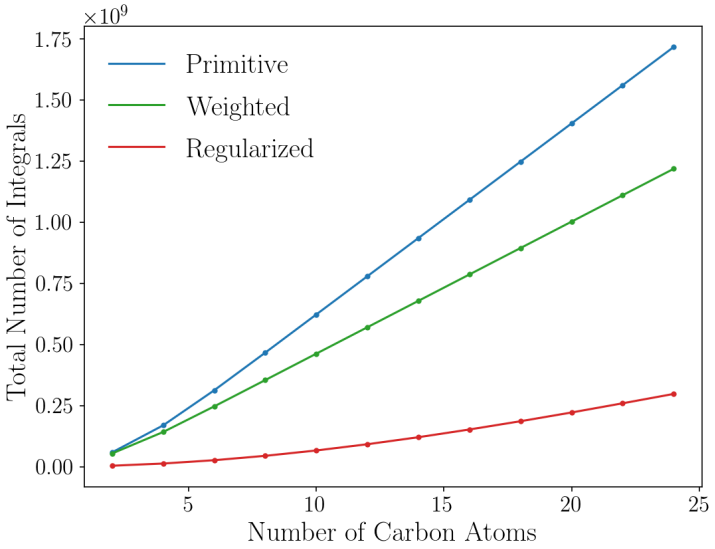

Figure 3.10: Total number of Fourier integrals in the ccpvtz basis with a numerical grid of size (23,22).

The multipole integrals do not scale linearly with respect to system size but are considerably

less numerous than the DF and Fourier integrals. The multipole integrals are also independent of α.

3.4.2 One-Body integrals

The number of one-body integrals is significantly less than the number three centre integrals for small to medium sized systems. For very large systems these integrals become a bottleneck because they do not scale linearly with respect to system size. The scaling of the one-body nuclear electron like integrals are shown below. The local occupied gauge centres are used for $V_{\alpha\mu}$ and the local virtual gauge centres are used for $V_{\alpha\bar{\mu}}$.

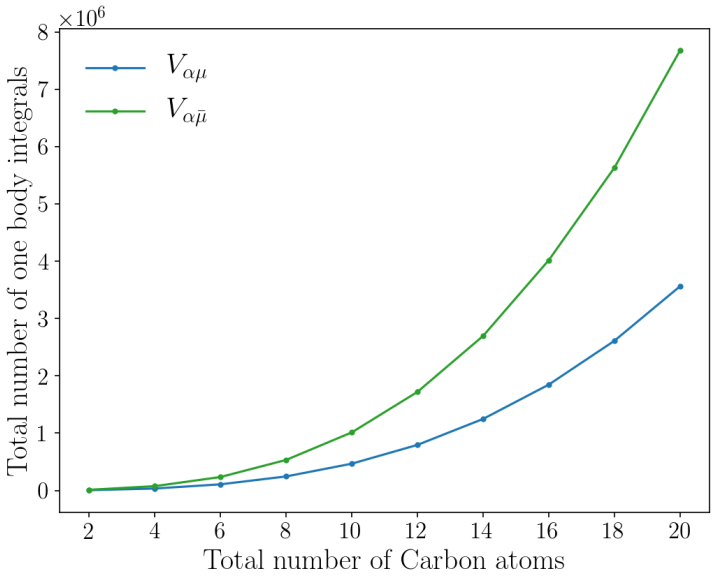

Figure 3.11: The scaling of the one-body terms for the polyacetylene chain in 6-31g basis with a sparse threshold of 10^{-9}.

For reference figure 3.12 depicts the scaling of several important quantities needed for a HF calculation. It can be seen that the density matrix and the exchange contribution do not scale linearly with respect to system size. The overlap and direct contributions do scale linearly with respect to system size. The short-range exchange does scale linearly with respect to system size.

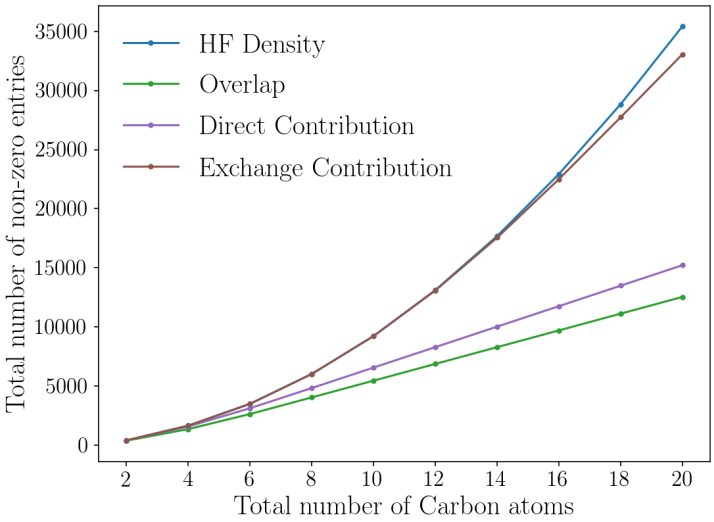

Figure 3.12: The scaling of several common one-body quantities for the polyacetylene chain in 6-31g basis with a sparse threshold of 10^{-9}.

3.5 Analysis of the Integral Accuracy

The accuracy of the direct and exchange contributions to the Fock matrix are analysed as a means to judge the accuracy of the integrals. The two-electron integrals are only analysed for the short-range contribution, the memory demand is too high to analyse the long-range four centre integrals. PySCF is used to calculate the "answer" which is compared to the result from the range-separated code. All errors are then reported as the log of the absolute error to quantify the accuracy, for quantities other than the HF energy the maximum absolute error of the matrix is reported.

$$\text{Log(Error)} = \text{Log}_{10}(|\text{PySCF} - \text{Range-separated}|) \tag{3.76}$$

To analyse Fock contributions, a restricted HF calculation is first performed in PySCF to obtain the converged density matrix and the Fock matrix answer. The converged PySCF density matrix is used to assemble the range-separated two-electron Fock contributions which are:

$$J = J_{DF} + J_{FT} + J_{MP} \tag{3.77}$$

$$K = K_{DF} + K_{FT} \tag{3.78}$$

The PySCF HF calculation uses density fitting such that the short-range contributions J_{DF} and K_{DF} can be subtracted from the answer to isolate long-range errors. There is a slight problem with this method in that the long-range answer is the long-range DF answer which is slightly different from the analytic long-range answer which cannot be calculated in the current code as it would require the analytic long-range four centre integrals. The values from the range-separated code are precise enough that this is not a large concern.

3.5.1 Short-Range Contributions

The short-range contributions are assembled using density fitting. Figure 3.13 depicts the maximum error of the short-range four centre integrals as a function of the range separation parameter α. The maximum error is independent of the range separation parameter. It can be assumed that the short-range integrals are precise to around the fourth decimal place, small fluctuations from this value would depend on the basis and fitting basis.

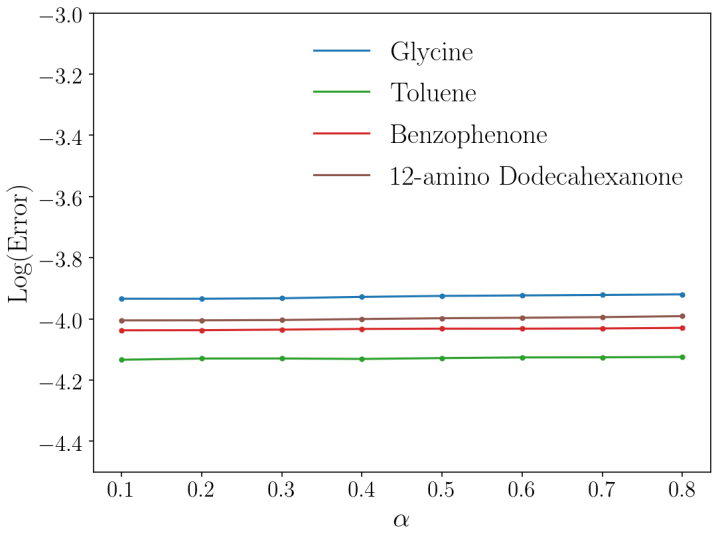

Figure 3.13: The maximum log error in the short-range four centre integrals assembled through density fitting as a function of range separation parameter α. The ccpvtz basis was chosen with the ccpvtz-jkfit auxiliary basis.

The total number of short-range integrals decreases with increasing α as the potential is more long-range. The smaller molecules glycine and toluene are small enough that the short-range potential extends the full length of the molecule even for $\alpha = 0.8$. For the two larger molecules the total number of non-zero integrals decreases with increasing α. It is important to point out that the DF error is dependent on the atoms and basis set used rather than the magnitude of

the integral. Figure 3.14 depicts the total number of non-zero integrals as a function of α.

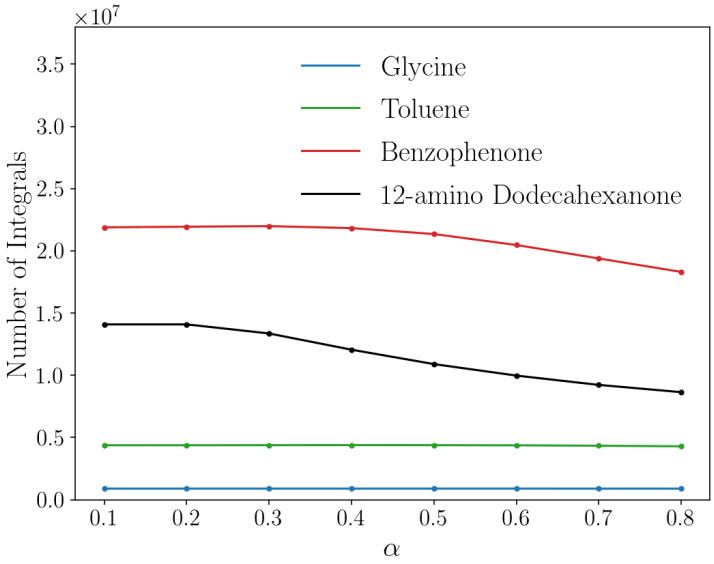

Figure 3.14: The total number of short-range four centre integrals with magnitude greater than 10^{-9}. *The ccpvtz basis was chosen with the ccpvtz-jkfit auxiliary basis.*

3.5.2 Long-Range Exchange Contributions

The long-range exchange contribution is assembled in the half transformed representation using the Fourier transform method or the regularized Fourier transform method.

$$K_{\alpha\beta} = \sum_{\mu,g} (\alpha\mu|g) w_g (g|\beta\mu)^* \tag{3.79}$$

$$K^R_{\alpha\beta} = \sum_{\mu,g} (\alpha\mu|g)^R w_g (g|\beta\mu)^{R*} + \sum_{\mu} S_{\alpha\mu} V_{\beta\mu}(\mathbf{R}_\mu) + V_{\alpha\mu}(\mathbf{R}_\mu) S_{\beta\mu} \tag{3.80}$$

Where w_g are the numerical weights of the Fourier transform. The associated gauge centres for the half transformed integrals is R_μ and the distance between the gauge centres is $R_{\mu\mu}$ which is always zero such that the Fourier method is always chosen. The exchange can be evaluated through the multipole method but it is much less accurate.

Figure 3.15 depicts the log error of the HF exchange energy for both the Fourier transform (solid lines) and the regularized Fourier transform (dashed lines) for changing α values. The accuracy of the HF exchange energy is better than the DF error (dotted red line) for most values of α . The larger the value of α the more the function $\eta(g)$, represented by the numerical

weights, oscillates. This is why the accuracy diminishes for larger values of α. For small α the regularized Fourier method is approximately the same accuracy as the Fourier method, there is some variance at larger α values.

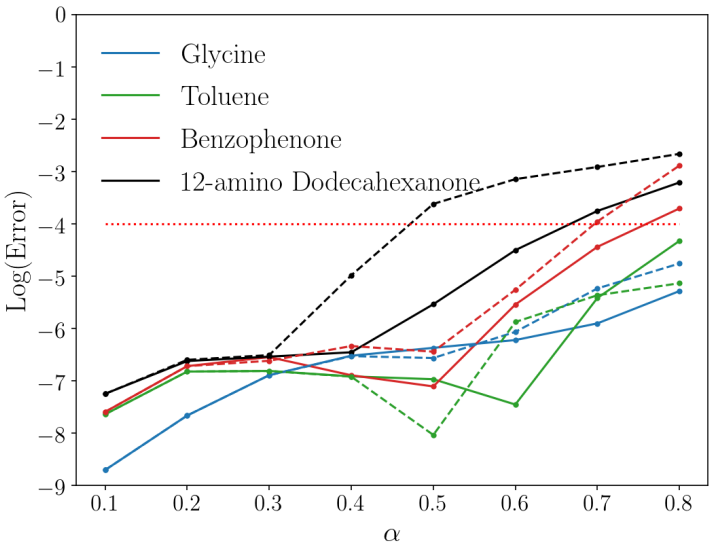

Figure 3.15: The log error of the HF exchange energy as a function of α at the ccpvtz level of theory with the (23,22) numerical grid. The solid lines represent the Fourier transform method, the dashed lines represent the regularized method. The dotted red line is the DF error for reference.

The accuracy of the Fourier method depends on the numerical grid size which is a combination of angular points and radial points. Figure 3.16 depicts the exchange energy error for a fixed angular grid (left) and a fixed radial grid (right). There is no significant gain in accuracy by increasing the numerical grid. If the grid is too small (23,14) the error can be improved. To be conservative the (23,22) grid or larger should be used to calculate the exchange contribution.

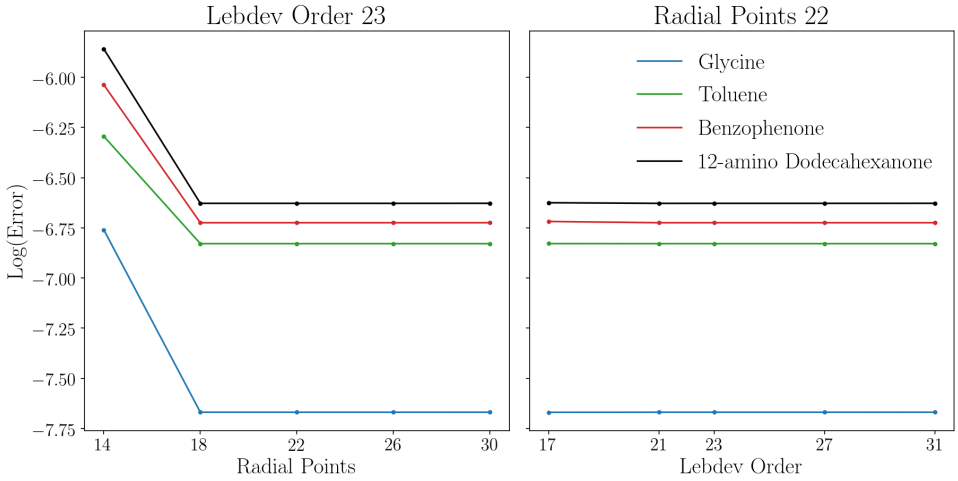

Figure 3.16: Log (Errors) of the HF exchange energy at the ccpvtz level of theory for chaining numerical grids. Left is a fixed angular grid of order 23. Right is a fixed radial grid of 22 equidistant points.

An important feature to the scaling of the Fourier integrals is that numerical grid does not need to increase with respect to system size to maintain the same accuracy. Figure 3.17 demonstrates this feature. The log(error) of the HF exchange energy is plotted for three different grid sizes (17,18), (23,22) and (31,30) for the polyacetylene system. The error grows with respect to system size which is expected for a systematic error. The error begins to scale linearly with respect to system size for the larger chains and all three grids evaluate to the same error. For the largest chain C26 the DF fitting set contains 1180 basis functions while the (17,18) grid consists of 1980 grid points. If the system was to double in size, not possible with the current code, the number of DF integrals will be larger than the number of Fourier integrals.

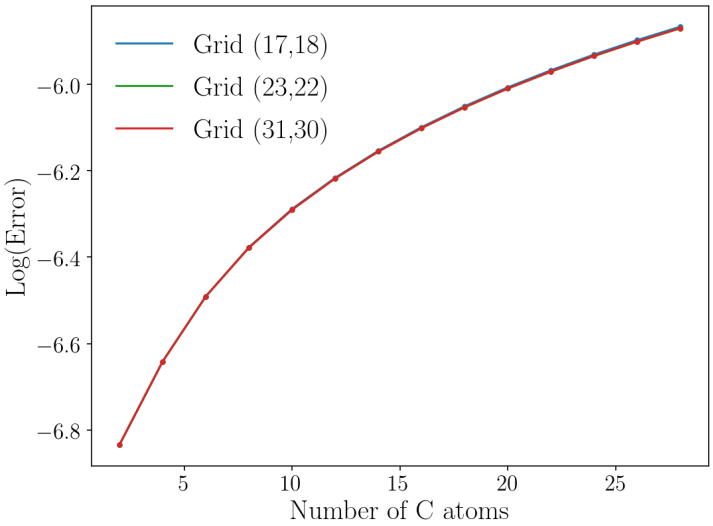

Figure 3.17: Log (Errors) of the HF exchange energy at the 6-31g level of theory for the polyacetylene system for α=0.2. All three numerical grids evaluate to the same numerical error independent of chain length.

It is well known that the exchange is predominantly a short-range phenomenon and one would expect the long-range energy contribution to be small. The HF exchange energy values are tabulated in 3.3 for both the Fourier and regularized Fourier methods. The regularized Fourier method contains two values, the regularized Fourier value and the one-body value which is calculated analytically. The first feature to note is that the long-range exchange contributions are small and they are accurate to roughly 6 decimal places which is better than the DF error.

The regularized Fourier exchange contributions are small and negative which is strange, the one-body contribution is larger and positive such that the sum of the two is correct. The negative energy contributions can be adjusted by deviating from γ_{opt} in the range-separated potential. In applications where accuracy is not the largest concern, K_{1body} or perhaps a scaled value of K_{1body} can be used to approximate the long-range exchange.

	α 0.1			α 0.3			α 0.5		
	FT Method	Regularized FT		FT Method	Regularized FT		FT Method	Regularized FT	
	K_{FT}	K^R_{FT}	K_{1body}	K_{FT}	K^R_{FT}	K_{1body}	K_{FT}	K^R_{FT}	K_{1body}
Glycine	0.0010	-0.0011	0.0021	0.1380	-0.1619	0.2999	0.8602	-1.0778	1.9380
Toluene	0.0027	-0.0049	0.0076	0.3024	-0.5931	0.8955	1.6540	-3.1816	4.8356
Benzophenone	0.0051	-0.0175	0.0226	0.5560	-1.5395	2.0955	3.0250	-6.8783	9.9033
12-amino dodecahexanone	0.0055	-0.0087	0.0141	0.5715	-0.9748	1.5463	3.1806	-5.5234	8.7043

Table 3.3: Tabulated long-range HF exchange energy contributions (ccpvtz basis) for the Fourier (K_{FT}) and regularized Fourier methods (K^R_{FT} + K_{1body}). The one-body contribution is calculated analytically.

To summarize this section the HF exchange contribution works very well for a decent range

of α values (0.1-0.5). Increasing the numerical grid does not significantly increase the accuracy and as long as the grid is large enough the exchange is calculated accurately. There is no clear advantage to the regularized method other than the one-body contribution might be a valid approximation if accuracy is not a large concern. For very large systems the numerical Fourier method will contain less integrals than the short-range density fit contribution. Conservative recommended parameters for calculating the exchange would be a grid size of (23,22) with an α value from 0.1 to 0.3.

3.5.3 Long-Range Direct Contributions

The long-range direct contribution should be simple, however it is the most challenging term to assemble. The direct term is challenging to assemble because it requires a switch between the long-range Fourier and long-range multipole methods. This switch (T_{switch}) is determined by the distance between the gauge centres $\mathbf{R}_{\mu\nu}$ which are the local gauge centres. The AO gauge centres do not work well for the switching as seen in figure 3.18.

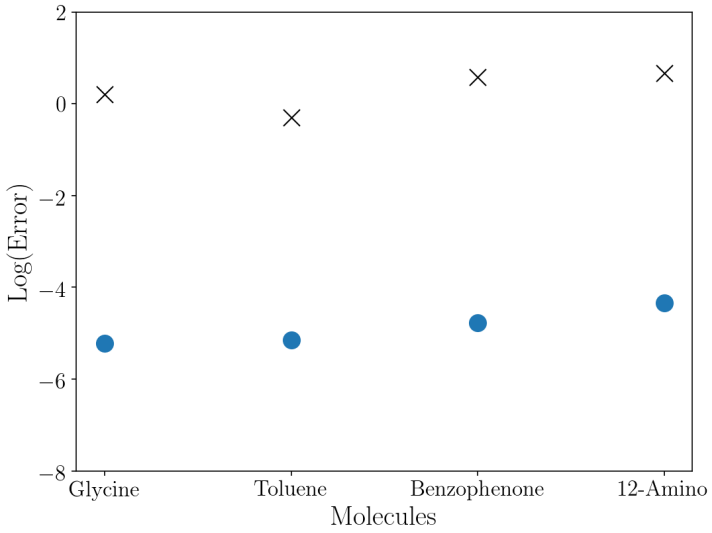

Figure 3.18: Log(error) of the direct HF energy using the AO gauge centres (x) and the local gauge centres (circles). The AO gauge centres should not be used.

The multipole method is essential to maintain accuracy in the direct term which is demonstrated in figure 3.19 using a stretched water dimer. The green dashed line is the error of the Fourier method without a switch, it breaks down for largely separated integrals. The blue curve is the switched method, with $T_{switch} = 8\mathring{A}$, the Fourier method is turned off and the

multipole method is turned on at this value..

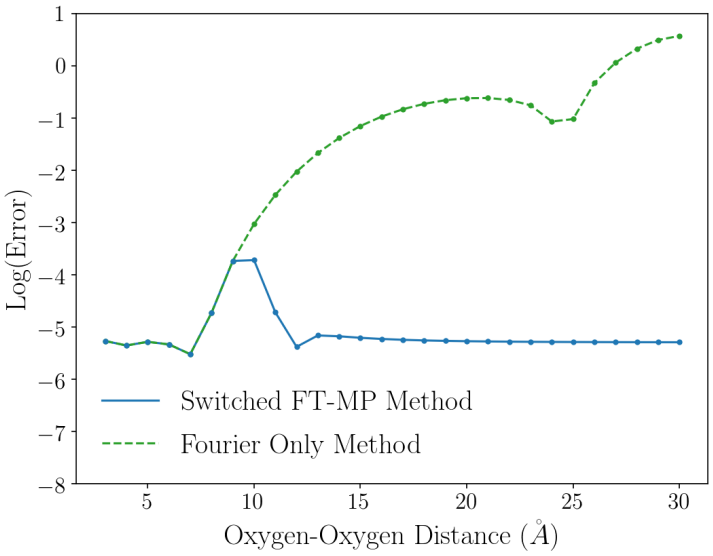

Figure 3.19: The maximum error of the direct matrix for the stretched water dimer in the ccpvtz basis. The green line is the pure Fourier method, the blue line is the switched method where contributions closer than 8Å are calculated using the Fourier method and contributions further than this are calculated using the multipole method.

The switch between Fourier and multipole methods is dependent on the range separation parameter. To find optimal T_{switch} values as a function of α, calculations were run using the stretched water dimer in the ccpvtz basis and the error of the direct term was monitored. For more complex systems the T_{switch} may need to be adjusted, the recommended values are presented below in table 3.4.

α	T_{switch} (Å)	α	T_{switch} (Å)
0.1	24	0.5	8
0.2	20	0.6	6.3
0.3	14	0.7	5.8
0.4	10	0.8	5

Table 3.4: Optimal T_{switch} values for a given range separation parameter. Values are determined from the stretched water dimer in the ccpvtz basis.

The accuracy of the direct terms are much more sensitive to α than the exchange terms. Figure 3.20 depicts the error of the HF direct energy. It is recommended to use small α values for calculation of the direct term. The stability of the numerical Fourier transform breaks down for large α values as the function $\eta(g)$ oscillated more rapidly.

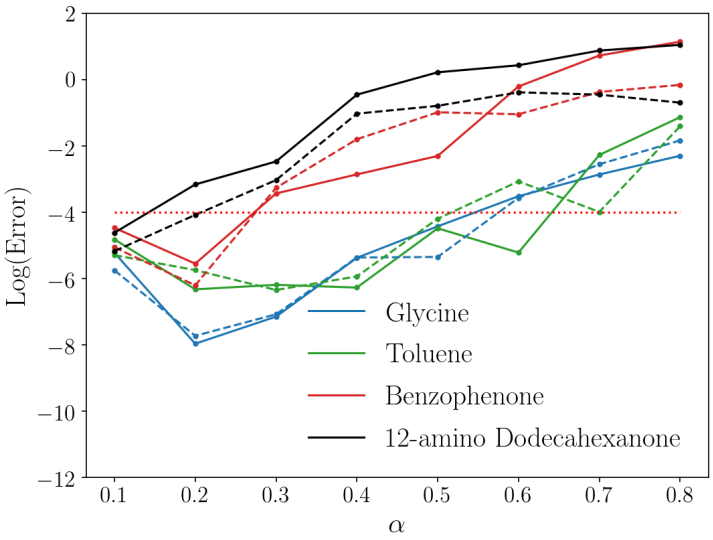

Figure 3.20: The log error of the direct long-range HF energy as a function of α. Solid lines are the Fourier-multipole method and dashed lines are the regularized method. The red dotted line is the density fit error for reference.

The size of the numerical grid is important for the accuracy of the method. In figure 3.21 log of the HF energy error is plotted as a function of N-Grid. N-grid is the both the angular order and the number of equidistant radial points. For the largest molecule 12-amino dodecahexanone (solid black line), a numerical grid of at least 27 is required to achieve numerical accuracy on the order of the DF error (dotted red line).

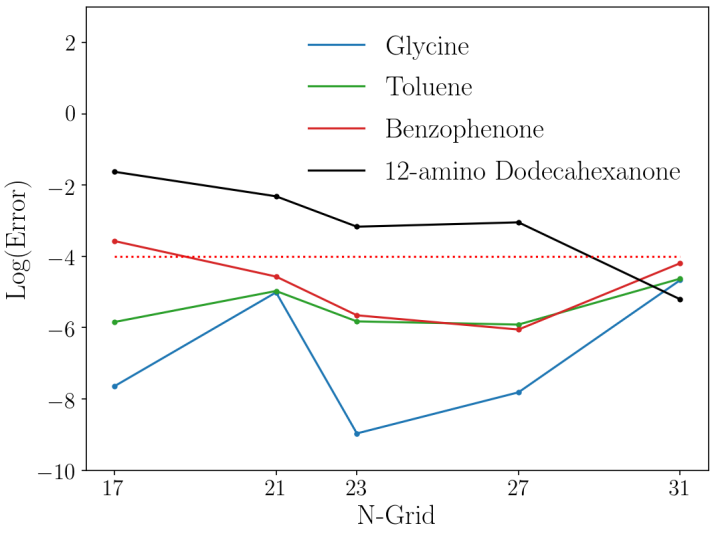

Figure 3.21: The log of the HF energy error as a function of N-Grid. N-grid is the both the angular order and the number of equidistant radial points. A range separation parameter of 0.2 and the ccvptz basis was used. The dotted red line is approximately the density fit error.

There is a small window in the parameters were the direct term is evaluated accurately. It is recommended to evaluate this term with $\alpha \leq 0.3$ and a numerical grid of at least 27.

3.6 Conclusion

Range-separated two-electron integrals are discussed in this chapter with the goal of convincing the reader that the amount data scales linearly with respect to system size and that the integrals are accurate enough to be used in quantum chemistry calculations. The two-range potential described in chapter 2 is used to partition the four centre integrals into short and long-range contributions. The short-range contribution is assembled using density fitting. The three centre DF integrals, $(\alpha\beta|x)$, scale linearly with respect to system size because the orbital β must be close to α (decays like the AO overlap matrix) and the auxiliary index x must be spatially close to $\alpha\beta$ because the integrals are short range.

The long-range contribution is evaluated through either a numerical Fourier transform or a multipole expansion. The Fourier method is used for close interactions while the multipole method is used for long-range interactions, both methods are essential to maintain high accuracy. Determining when to switch between methods is complicated. Local orbitals $|\mu\rangle$ are used to define local gauge centres $\mathbf{R}_\mu$ which are unambiguous in definition. The distance between gauge centres $\mathbf{R}_\mu$ and $\mathbf{R}_\nu$ is used to determine which method may be employed to evaluate the integral $(\alpha\mu|\gamma\nu)$, orbital extents are accounted for in this distance metric. The switching is controlled by T_{switch} and table 3.4 reports recommended T_{switch} values for a range of α's.

The Fourier integrals scale linearly with respect to system size because the numerical grid size does not need to increase with respect to system size. The Cartesian multipole moments do not scale linearly with respect to system size but are significantly less numerous then the DF and Fourier integrals.

The accuracy of the integrals is analysed through an analysis of the HF direct and exchange contributions. This is because it is not practical to fully assemble the four centre integrals to check accuracy in sizeable systems. The accuracy of the direct and exchange terms behave differently with respect to the range separation parameter and the numerical grid size. The exchange contribution can be evaluated with a large α value and a small numerical grid. The

direct term should be assembled with a small α value, we recommend $\alpha \leq 0.3$ and a numerical grid of at least N-grid 23. If the recommended parameters are chosen, the long-range error should be better than the error introduced from density fitting.

Regularized Fourier and mulitpole integrals are derived in this chapter following a similar derivation to the regularization of the long-range potential in chapter 2. The one-body contribution is a product of the regularization procedure. These one-body integrals do not scale linearly with respect to system size but they are comprised of nuclear-electron like attraction integrals and the overlap matrix which are significantly cheaper to calculate than the DF, Fourier and multipole integrals. In very large systems these integrals might become a bottleneck. The one-body integrals are evaluated analytically and the error in the regularized method comes from the error in the regularized Fourier and multipole methods. The regularized integrals are significantly more sparse than their unregularized counterpart. The accuracy of regularized integrals is on the same order as the accuracy of unregularized integrals.

Hopefully the data presented in this chapter convinces the reader that the range-separated molecular integrals and the regularized integrals are accurate enough to use in quantum chemistry calculations.

Chapter 4

Integral Direct Hartree-Fock

4.1 Introduction

Hartree-Fock theory[4,105–107] is the cornerstone of *ab initio* electronic structure theory and it is equivalent to the molecular orbital approximation. Optimizing the molecular orbitals through the HF method is often the first step in many electronic structure methods. The orbitals are optimized by solving the Hartree-Fock Roothaan equations. The Hartree-Fock Roothaan equations, see equation 4.1, are non-linear because the Fock operator F is dependent on the density matrix (the molecular orbital (MO) coefficients C) and is therefore solved iteratively.

$$F(C)C = SC\epsilon \tag{4.1}$$

The iterative self-consistent field (SCF) procedure is used to solve these equations. Briefly, the SCF procedure starts with a guess at the MO coefficients which are used to construct the Fock operator F which is then used to obtain a better guess at the MO coefficients. This procedure is repeated until the MO coefficients converge. The largest expense of the SCF procedure is calculating the direct (J) and exchange (K) contributions to the Fock matrix, which are assembled using the two-electron integrals and the atomic orbital (AO) density matrix $D_{\gamma\delta}$ as:

$$J_{\alpha\beta} = \sum_{\gamma,\delta} (\alpha\beta|\gamma\delta) D_{\gamma\delta} \tag{4.2}$$

$$K_{\alpha\beta} = \sum_{\gamma,\delta} (\alpha\gamma|\beta\delta) D_{\gamma\delta} \tag{4.3}$$

Since the two-electron integrals are too numerous to store, they are recalculated at each SCF iteration[66]; this procedure is called direct-SCF and is computationally expensive but the storage bottleneck caused by the two-electron integrals is avoided.

In this chapter, we present a range-separated Fock matrix that is assembled from range-separated integrals that scale linearly with respect to system size. The range-separated potentials used throughout this chapter are the two-range potential and the regularized two-range potential, both of which are explored in the previous chapters.

$$V_{2range} = V_{sr} + V_{lr} \tag{4.4}$$

$$V^{R}_{2range} = V_{sr} + V^{R}_{lr} + V_{1body} \tag{4.5}$$

The range-separated Fock matrix is assembled in two parts; short-range and long-range contributions. The short-range contributions are assembled through the density fitting (DF) approximation using the short-range three centre integrals $(\alpha\beta|x)$ which exhibit the same behaviour as conventional full-range integrals with the added sparsity along the auxiliary axis. In principal the short-range contribution to the Fock matrix can be assembled in modern J and K engines by swapping integral codes, the added sparsity can only improve the efficiency of these codes. Our own implementation tries to take full advantage of the sparsity of the 3-center integrals.

The equations for the short-range J and K are:

$$J_{\alpha\beta} = \sum_{\gamma,\delta,x,y} (\alpha\beta|x) M^{-1}_{xy} (y|\gamma\delta) D_{\gamma\delta} \tag{4.6}$$

$$K_{\alpha\beta} = \sum_{\gamma,\delta,x,y} (\alpha\gamma|x) M^{-1}_{xy} (y|\beta\delta) D_{\gamma\delta} \tag{4.7}$$

The inverse metric, M^{-1}_{xy}, is calculated once at the start of the calculation and does not need to be calculated at every SCF iteration. Since he integrals $(\alpha\beta|x)$ need to be calculated at each iteration, it is essential to have efficient integral pre-screening which can be enhanced by including information about the density matrix.

In order to assemble the long-range contribution, local orbitals (LO) $|\mu\rangle = L_{\alpha\mu}|\alpha\rangle$ and corresponding local gauge centres $\mathbf{R}_{\mu}$ must first be introduced. The transformation coefficients $L_{\alpha\mu}$ are obtained through a pivoted-Cholesky decomposition of the density matrix that needs to be done once at the start of each SCF iteration. The coefficients are orthonormalized through

Löwdin orthonormalization.

$$D_{\gamma\delta} = \sum_{\mu} L_{\gamma\mu} L_{\delta\mu} \tag{4.8}$$

The gauge centres are well-defined vectors that can be calculated from the dipole moments.

$$\mathbf{R}_{\mu} = \frac{\langle \mu | \mathbf{r} | \mu \rangle}{\langle \mu | \mu \rangle} \tag{4.9}$$

Using the local orbitals, the range-separated direct contribution can then be written down in a half transformed AO/LO representation.

$$\begin{aligned} J_{\alpha\nu} = & \sum_{\mu,g} \Theta(\mathbf{R}_{\mu\nu})(\alpha\nu|g)|w_g|P_g(g|\mu\mu)^* \\ & + \sum_{k,l,\mu} \bar{\Theta}(\mathbf{R}_{\mu\nu})(\alpha\nu|k) f^{(k,l)}(\mathbf{R}_{\mu\nu})(k|\mu\mu) \end{aligned} \tag{4.10}$$

The first line in equation 4.10 is the numerical Fourier transform and the second line is the multipole expansion of the same contribution. The function $\Theta(\mathbf{R}_{\mu\nu})$ and its converse $\bar{\Theta}(\mathbf{R}_{\mu\nu})$ are switching functions to decide which method is used for each $\mu\nu$ pair. The switch is based off the magnitude of $\mathbf{R}_{\mu\nu}$ which is an estimate of the distance between the three centre integrals $(\alpha\nu|g)$ and $(\mu\mu|g)^*$. The Fourier method is used when $\mathbf{R}_{\mu\nu}$ is small and the multipole expansion is used when $\mathbf{R}_{\mu\nu}$ is large, the size of $\mathbf{R}_{\mu\nu}$ is the magnitude of the vector minus the extents of orbitals, see Chapter 3 "Assigning Gauge Centres" for more information. The terms $|w_g|$ and P_g are the magnitudes and phase of the numerical grid points, they are split into two terms because it is possible to have negative weights. The function $f^{(k,l)}$ is the analytical derivatives of the long-range potential. The switching between Fourier and multipole methods is explained in more detail in the chapter 3.

To obtain the direct contribution in the AO representation, $J_{\alpha\mu}$ must be projected back into the AO space. The projector $L^{\intercal}S$ is used to project from the half transformed representation to the AO representation. However, this projector only projects out the occupied block of the Fock matrix such that equation 4.11 only contains information about the occupied-occupied

and virtual-occupied blocks since the α index is not projected.

$$J_{\alpha O} = \sum_{\mu\gamma} J_{\alpha\mu} L_{\gamma\mu} S_{\gamma\beta} \tag{4.11}$$

$$= J_{VO} + J_{OO} \tag{4.12}$$

The occupied-virtual block is easy to obtain from the transpose of $J_{\alpha O}$. To get the virtual-virtual block, an assembly similar to the assembly of $J_{\alpha\mu}$ is conducted using local virtual orbitals $|\bar{\mu}\rangle$ to assemble $J_{\alpha\bar{\mu}}$. The local virtual orbitals are obtained via pivoted-Cholesky decomposition of $\bar{D} = S^{-1} - D$ without orthonormalization as this can lead to numerical issues with small eigenvalues in large basis sets. The long-range direct contribution can be assembled as:

$$J = J_{\alpha O} + J_{O\alpha} - J_{OO} + J_{VV} \tag{4.13}$$

The long-range exchange is simpler to assemble as only the Fourier transform method is used and there is no need to project back into the AO representation.

$$K_{\alpha\beta} = \sum_{\mu,g} (\alpha\mu|g)|w_g|P_g(g|\beta\mu)^* \tag{4.14}$$

To assemble the regularized long-range direct and exchange contributions one only needs use the regularized integrals for the long-range J and K matrices and calculate the one-body contributions.

$$J_{\alpha\nu} = \sum_{\mu} V_{\alpha\nu}(\mathbf{R}_\mu) S_{\mu\mu} + S_{\alpha\nu} V_{\mu\mu}(\mathbf{R}_\nu) - S_{\alpha\nu} S_{\mu\mu} V(\mathbf{R}_\mu - \mathbf{R}_\nu) \tag{4.15}$$

$$K_{\alpha\beta} = \sum_{\mu} V_{\alpha\mu}(\mathbf{R}_\mu) S_{\beta\mu} + S_{\alpha\mu} V_{\beta\mu}(\mathbf{R}_\mu) - S_{\alpha\mu} S_{\mu\mu} V(0) \tag{4.16}$$

Again $J_{\alpha\bar{\nu}}$ must be assembled using local virtual orbitals to obtain the correct J_{VV} block.

The algorithms to assemble the two-electron contributions to the range-separated Fock matrix are presented in section "Algorithms", the algorithms make use of sparse-matrix multiplies and integral pre-screening. There are many techniques that have been developed over the decades

to aid in the efficiency of the construction of the Fock matrix and the SCF procedure[4,107], see the introduction chapter of this thesis for more details. Examples of such are the incremental Fock matrix approach, where the change in the density matrix is used to construct the change in the Fock matrix at each iteration.

$$F'_{\alpha\beta} = F_{\alpha\beta} + \Delta F_{\alpha\beta} \tag{4.17}$$

$$\Delta F_{\alpha\beta} = J(\Delta D_{\gamma\delta}) + K(\Delta D_{\gamma\delta}) \tag{4.18}$$

$$\Delta D_{\gamma\delta} = D'_{\gamma\delta} - D_{\gamma\delta} \tag{4.19}$$

The incremental Fock approach is advantageous because the $\Delta D_{\gamma\delta}$ can be used to pre-screen the two-electron integrals for large computational gain. The use of an incremental Fock approach for the long-range contributions is not obvious in our scheme because the incremental density is not positive definite.

The primary goal of this chapter is to convince the reader that the range-separated Fock matrix can be assembled and that there are no issues with the convergence of the SCF procedure. Therefore many of the "bells and whistles" found in optimized codes have not been implemented in our code mainly due to the required implementation time. The code is still in early stages of development and is not competitive in regards to computation time.

4.2 Algorithms

The following algorithms are called once for each SCF iteration to assemble the direct and exchange term to the Fock matrix. At the start of each iteration the local occupied and local virtual orbitals are obtained from pivoted-Cholesky decomposition of D and $\bar{D}$ along with the local gauge centres.

Integral pre-screening is essential for efficiency of the algorithms. The integrals are pre-screened in blocks defined by their associated nuclear coordinates. The AO's α are batched by grouping their associated atoms A into a group of local atoms centred around $\mathbf{R}_A$ termed tile centres. The molecule's coordinates are reordered to follow the tile centre ordering, this is done once at the start of the calculation. The integrals can then be batched into small batches $(\alpha_A \beta_B | \kappa_K)$

(κ is used for general 3 index integral) that can be recalculated as necessary. The tile centres $\mathbf{R}_A$ and $\mathbf{R}_B$ can be screened to identify which A,B blocks are non-zero indicated as $B \in A$. The batching of the label κ_K depends on the specific type of three centre integral, for DF integrals this is the auxiliary index, for Fourier integrals this the grid point g and for the multipole integrals this label is the multipole level l.

Similar to the AO labels the Cholesky labels μ can be grouped around a common center M. The coefficients $L_{\alpha\mu}$ can then be batched into smaller $L_{\alpha_A\mu_M}$ blocks and the A,M pairs that are non-zero (above a threshold) can be identified and labelled as $A \in M$.

To assemble the direct and exchange terms simultaneously we assume an integral block $(\alpha_A\beta|x)$ can be calculated and stored in memory as a sparse matrix. Screening is used such that A,B,K blocks that are essentially zero are ignored, greatly reducing the memory demand.

The long-range direct term is evaluated through either the Fourier or multipole method depending on the distance between the local orbitals $\mu\nu$. This choice is made at the start of each SCF iteration and depends on the value of the range separation parameter. In later iterations when the calculation is near convergence it is possible to fix these choices. The choices are stored in boolean arrays called FT_{map} and MP_{map}.

The current code is still is not optimized and it is not possible to report on the scaling of computational time versus system size.

4.2.1 Assembly of the Density Fit Contributions

The short-range direct term is assembled in three steps. First the inverse metric matrix (M^{-1}_{xy}) is contracted with the three centre integrals to form the intermediate $I_1(\gamma,\delta,y)$, which is later contracted with the density matrix to form $I_2(x)$. The final step is to contract the intermediate $I_2(x)$ with the three centre integrals, this requires the three centre integrals to be calculated twice per SCF iteration which can be costly but the integrals are short-range and the amount of data calculated at each call scales linearly with respect to system size. The efficiency of the calculation of $I_1(\gamma,\delta,y)$ and $I_2(x)$ can be improved by including the sparsity of the density

matrix in the integral pre-screening.

$$\begin{aligned} J_{\alpha\beta} &= \sum_{\gamma,\delta,x,y} (\alpha\beta|x) M^{-1}_{xy} (y|\gamma\delta) D_{\gamma\delta} \\ &= \sum_{\gamma,\delta,x} (\alpha\beta|x) I_1(\gamma,\delta,x) D_{\gamma\delta} \\ &= \sum_{x} (\alpha\beta|x) I_2(x) \end{aligned} \tag{4.20}$$

To assemble the short-range exchange contribution we take advantage of the local orbitals and the half transformed representation.

$$\begin{aligned} K_{\alpha\beta} &= \sum_{\gamma,\delta,x,y\mu} (\alpha\gamma|x) M^{-1}_{xy} (y|\beta\delta) L_{\gamma\mu} L_{\mu\delta} \\ &= \sum_{\mu,x,y} (\alpha\mu|x) M^{-1}_{xy} (y|\beta\mu) \\ &= \sum_{\mu,x,y} (\alpha\mu|x) m_{xk} m_{ky} (y|\beta\mu) \\ &= I_K(\alpha\mu,k) I_K(\beta\mu,k) \end{aligned} \tag{4.21}$$

The advantage to this method is that the exchange can be assembled by batching over the local orbitals $|\mu\rangle$ and the local coefficients $L_{\gamma\mu}$ can be used to screen which γ indices to included in the calculation of $(\alpha\gamma|x)$ limiting the range of $\alpha\gamma$ pairs.

The calculation of the direct intermediate $I_2(x)$ can be performed during the assembly of the short-range exchange contribution. Algorithm 1 outlines the construction of the short-range contributions. The algorithm requires the batched primitive integrals $(\alpha_A\beta|y)$, the density matrix, the batched local orbital coefficients $L_{\alpha_A,\mu}$ and both the inverse metric matrix and its symmetric halves $M^{-1}_{xy} = \sum_k m_{x,k} m_{k,y}$.

The main loop for the the exchange contribution is a loop over a batch of local orbitals. The local orbitals are batched into M groups to keep the memory demand low. For a particular orbital $|\mu\rangle$, the magnitude of $L_{\alpha_A,\mu}$ determines which $|\alpha_A\rangle$ to include in the calculation of $(\alpha_A\beta|y)$ for that μ, the threshold to control this screening is $T_{density}$.

Algorithm 1 Short Range Density Fit Fock Contributions

1: $MapA = False\ \forall A$ ▷ Vector to keep track of which A's have been summed into $I_2(x)$
2: **for** M=1 in nM **do** ▷ Batch of local orbitals μ
3: **for** $A \in M$ **do** ▷ Defined by $max(L_{\alpha_A,\mu}) > T_{density}$
4: Obtain the Integral Batch $(\alpha_A\beta|y)$ ▷ $B \in A, Y \in AB$
5: $(\mu\beta|y) += \sum_{\alpha_A} (\alpha_A\beta|y) L_{\alpha_A\mu}$
6: **if** $MapA \neq True$ **then** ▷ Intermediate used for J, only do this once
7: $I_1(\alpha_A,\beta,x) += \sum_y (\alpha_A\beta|y) M^{-1}_{xy}$
8: $I_2(x) += \sum_{\alpha_A\beta} I_1(\alpha_A,\beta,x) D_{\alpha_A\beta}$
9: $MapA$ at $A = True$
10: **end if**
11: **end for**
12: $I_K(\mu\beta,x) += \sum_y (\mu\beta|y) m_{y,x}$
13: $K_{\beta\gamma} += \sum_{\mu k} I_K(\mu\beta,x) I_K(\mu\gamma,x)$ ▷ Exchange Contribution
14: **end for**
15: **for** A **do**
16: Obtain the Integral Batch $(\alpha_A\beta|x)$ ▷ $B \in A, X \in AB$
17: $J_{\alpha_A\beta} = \sum_x (\alpha_A\beta|x) I2(x)$ ▷ Direct Contribution
18: **end for**

The largest matrix kept in memory at one time is the sparse matrix $(\alpha_A\beta|x)$ which will never be too large because the integrals are short range and the size of A is controlled by the user/program. It is important to note that this is independent of the size of the molecule, however the memory usage can become large if very large basis sets are used.

The two main loops in algorithm 1 are a master loop over M and a loop over A, the second loops range is restricted since there is small range of α_A associated with M. There maybe some redundancy in the number of times the integral block $(\alpha_A\beta|x)$ is calculated which happens when different M require the same integrals. It is possible to sort the order of M indices such that $M+1$ is as similar as it can be to M and a rolling set of integral blocks can be kept in memory to minimize the amount of redundant integral calculations. Another option is to label each integral block as a sparse or dense blocks and keep the sparse blocks on disk or even memory. Only the dense block would be recalculated as needed.

To calculate $I_2(x)$ we need to be careful not to double count any integrals, so the first time an integral block is calculated it is added to $I_2(x)$, then if it is recalculated for a different M

it is not added, $MapA$ is used to keep track of this information.

The scaling of this algorithm should be low order, the total amount of data in the primitive integral blocks, $(\alpha_A\beta|y)$ scales linearly with respect to system size. The primitive blocks need to be calculated at least twice to construct J and K. The integrals used in the construction K can be heavily screened with the local orbital coefficients $L_{\alpha_A,\mu}$. Unfortunately, the current code is far from optimized and hard data on the scaling of the algorithm cannot be shown at present.

Loss of numerical precision can occur if one sums many small numbers into a large number. For example if one tried to sum a million numbers that had a value of about 1 but were precise to 15 decimal places, the final summation number would have a value of 1 million but would only be precise to about 9 decimal places. This issue would occur in the construction of the short-range direct term if the $I_2(x)$ intermediate was built by first contracting with the density then contracting with the inverse metric matrix.

$$I_1(y) = \sum_{\alpha\beta} (\alpha\beta|y) D_{\alpha\beta} \tag{4.22}$$

$$I_2(x) = \sum_{y} M^{-1}_{xy} I_1(y) \tag{4.23}$$

Equation 4.23 is referred to as case 1 and this is not implemented in the SCF code because it causes precision problems. The first step in case 1 is summing the $I_1(y)$ intermediate which has larger numerical values than $I_2(x)$ and some numerical precision is lost in this process. Equation 4.25, referred to as case 2, is the order that is implemented in the SCF code and it does not cause issues with numerical precision.

$$I_1(\alpha,\beta,x) = \sum_{y} M^{-1}_{xy} (\alpha\beta|y) \tag{4.24}$$

$$I_2(x) = \sum_{\alpha\beta} I_1(\alpha,\beta,x) D_{\alpha\beta} \tag{4.25}$$

Figure 4.1 depicts the convergence of the SCF method for a simple water dimer, using equations 4.23 and 4.25 to assemble the $I_2(x)$ intermediate. One can see how the SCF oscillates and cannot converge for case 1 because the SCF convergence threshold is tighter than the numerical precision of the intermediate $I_2(x)$.

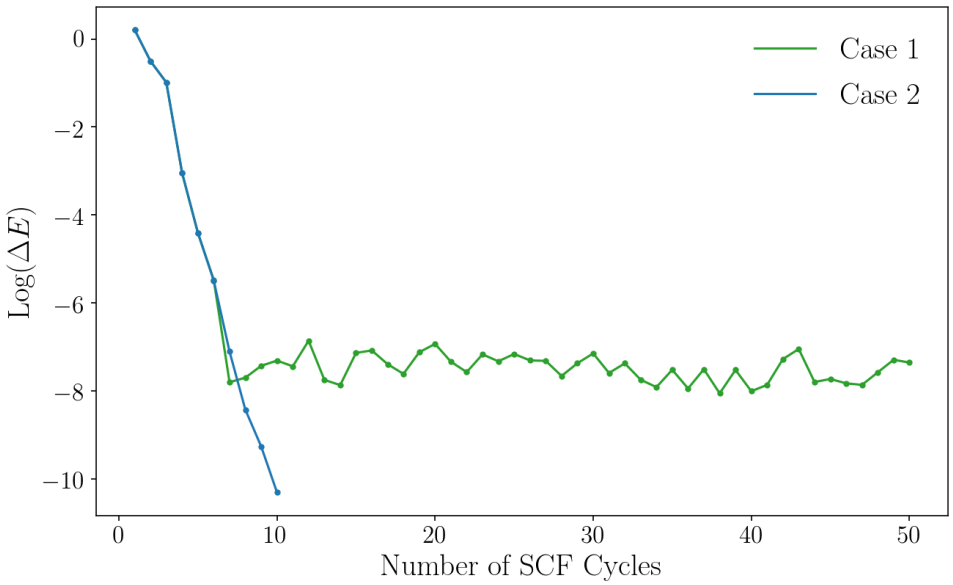

Figure 4.1: The convergence of the SCF energy for a simple water dimer separated by 5Å. Case 1 is where $I_2(x)$ is assembled using equation 4.23 and Case 2 is where $I_2(x)$ is assembled using equation 4.25.

In practice, converging the SCF energy to very tight thresholds is unnecessary for the vast majority of molecules. This very tight convergence is only necessary when the molecule of interest sits on a flat potential energy surface and the user wants to ensure that they are not trapped in a local minima.

4.2.2 Assembly of the Fourier Contributions

The Fourier contribution is evaluated numerically in spherical coordinates using a spherical grid of points and weights. The spherical grids are defined by an angular order (Lebedev[88]) and radial quadrature, while the weights are dependent on the radial potential used. Throughout this work an equidistant radial grid was used, however, Gauss-Legendre[91] or Gauss-Chebyshev[90] quadratures also work. We did not find any significant improvement between these quadratures.

The numerical Fourier contributions to the Fock matrix are presented below in equation 4.27. The direct term is assembled in a mixed AO-LO representation $J_{\alpha\nu}$ which when transformed back to the AO basis contains information about the occupied-virtual and occupied-occupied

blocks of the Fock matrix. To obtain the virtual-virtual block, $J_{\alpha\bar{\nu}}$ must also be calculated.

$$J_{\alpha\nu} = \sum_{\mu,g} (\alpha\mu|g)|w_g|P_g(g|\mu\mu)^* \tag{4.26}$$

$$K_{\alpha\beta} = \sum_{\mu,g} (\alpha\mu|g)|w_g|P_g(g|\beta\mu)^* \tag{4.27}$$

The primitive FTAO integrals are unique from the other primitive integrals because they are complex. The numerical weights are purely real but they cannot be square rooted and attached to the primitive three index integrals because for some potentials the weights are negative. The weights must first be factored into magnitude ($|w_g|$) and phase (P_g), the phase of the weights is either $+1$ or -1, the magnitude of the weights can be square rooted and attached to the primitive three index integrals.

Similar to the short-range integrals, it is assumed that the the primitive integrals can be calculated in a small batch $(\alpha_A\beta_B|g_i)$ where g_i would be a small batch of g points controlled by the user/program. The size of the primitive integrals is limited, independent of system size, because A must always be close to B and the number of g points held in memory at one time is controlled. Local orbitals are essential to the assembly of the Fourier contribution and it is assumed that the local orbital coefficients are in memory and batched into $L_{\alpha_A\mu_M}$.

The Fourier method is most accurate for small values of $\mathbf{R}_{\mu\nu}$, for the exchange matrix $\mathbf{R}_{\mu\nu} = \mathbf{R}_{\mu\mu} = 0$ such that the Fourier method is always reasonably accurate. This is not the case for the direct term. The accuracy of the direct term is dependent on the size of the vector $\mathbf{R}_{\mu\nu}$; for small $\mathbf{R}_{\mu\nu}$ the Fourier method should be used, for large $\mathbf{R}_{\mu\nu}$ the multipole method should be used. To identify which $\mu\nu$ pairs are to be calculated with the Fourier method, a map is created called FT_{Map} which controls which $\mu\nu$ pairs are added to the direct Fourier contribution.

The assembly of the Fourier contributions is outlined in algorithm 2. The master loop is over the batches of g which is used to control the size of the primitive integrals. Then similar to the DF assembly we loop over M and all $A \in M$ to create the first intermediate termed $K3a$.

$$K3a(\mu\beta|g) += \sum_{\alpha_A} (\alpha_A\beta|g) L_{\alpha_A\mu} \tag{4.28}$$

Next the appropriate weights are attached $K3b = K3a \cdot |w_g|^{\frac{1}{2}}$, the weights are dependent on the specific potential that is being assembled. The $K3b$ object is sparse and should be pruned at this step to remove any small values. The beauty to this algorithm is that the regularization of the Fourier integrals is very simple, if the integrals are to be regularized, $K3b$ is regularized by subtracting a phase shifted overlap matrix.

$$K3b^R(\mu\beta|g) = K3b(\mu\beta|g) - |w_g|^{\frac{1}{2}} \cdot S_{\mu\beta} e^{i\mathbf{g}\cdot\mathbf{R}_\mu} \tag{4.29}$$

The exchange can be assembled in one loop over the g batches, the direct term requires the primitive integral to be recalculated. The first loop over g creates an intermediate used for the direct term $I_1(\mu,g)$. The second loop over g is where $K3b(\nu\beta|g)$ is created and the FT_{Map} is used to decide which $\mu\nu$ pairs are added to the direct term. A similar loop is needed to create the $J_{\alpha\bar{\nu}}$ term where the virtual orbtials are used instead of the occupied local orbitals.

One of the key advantages to algorithm 2 is that it can be paralelized over the main loop, since each g label is independent and the primitive integrals $(\alpha_A\beta|g)$ are never very large. This is independent of system size. The regularization of the integrals is inexpensive and can be toggled on or off by a simple boolean flag such that no special treatment is required. The regularized integrals are even more sparse than their unregularized variants, which could lead to an increase in efficiency if efficient sparse-sparse matrix multiplies are used. The number of g points is independent of system size and the amount of data required for this algorithm scales linearly with respect to system size.

Algorithm 2 Fourier Fock Contributions

1: **for** G in *gbatch* **do** ▷ Batch of grid points g
2: **for** M=1 in nM **do** ▷ Batch of local orbitals μ
3: **for** $A \in M$ **do** ▷ Defined by $max(L_{\alpha_A \mu}) > Tdensity$
4: Obtain the Integral Batch $(\alpha_A \beta|g)$ ▷ $B \in A$
5: $K3a(\mu\beta|g) += \sum_{\alpha_A} (\alpha_A \beta|g) L_{\alpha_A \mu}$ ▷ Intermediate used for K
6: **end for**
7: $K3b(\mu\beta|g) = K3a(\mu\beta|g) \cdot |w_g|^{\frac{1}{2}}$ ▷ attach the square root weights
8: **if** Regularize **then**
9: $K3b(\mu\beta|g) = K3b(\mu\beta|g) - |w_g|^{\frac{1}{2}} \cdot S_{\mu\beta} e^{i\mathbf{g}\cdot\mathbf{R}_\mu}$
10: **end if**
11: $I_1(\mu,g) = \sum_{\beta} K3b(\mu\beta|g) L_{\beta\mu}$ ▷ Intermediate used for J
12: $K_{\beta\gamma} += \sum_{\mu g} P_g \cdot K3b(\mu\beta|g) \cdot K3b^*(\mu\gamma|g)$ ▷ Exchange Contribution
13: **end for**
14: **end for**
15: Relabel $I_1(\mu,g)$ as $I_1(\nu,g)$
16: **for** G in *gbatch* **do** ▷ Batch of grid points g
17: **for** M=1 in nM **do** ▷ Batch of local orbitals μ
18: **for** $A \in M$ **do** ▷ Defined by $max(L_{\alpha_A \mu}) > Tdensity$
19: Obtain the Integral Batch $(\alpha_A \beta|g)$ ▷ $B \in A$
20: $K3a(\mu\beta|g) += \sum_{\alpha_A} (\alpha_A \beta|x) L_{\alpha_A \mu}$
21: **end for**
22: $K3b(\mu\beta|g) = K3a(\mu\beta|g) \cdot |w_g|^{\frac{1}{2}}$ ▷ attach the square root weights
23: **if** Regularize **then**
24: $K3b^R(\mu\beta|g) = K3b(\mu\beta|g) - |w_g|^{\frac{1}{2}} \cdot S_{\mu\beta} e^{i\mathbf{g}\cdot\mathbf{R}_\mu}$
25: **end if**
26: **for** $\nu \in FT_{Map}$ **do** ▷ Only the $\mu\nu$ pairs that are close
27: $J_{\beta\mu} += \sum_{\nu g} P_g \cdot K3b(\mu\beta|g) \cdot I_1(\nu g)$
28: **end for**
29: **end for**
30: Repeat lines 17 to 29 using local virtual orbitals to obtain J_{VV}
31: **end for**
32: $J = J_{\alpha O} + J_{O\alpha} - J_{OO} + J_{VV}$

4.2.3 Assembly of the Multipole Contributions

The multipole contributions are assembled in Cartesian coordinates for simplicity even though there is redundant information. It is possible to calculate the exchange term through the multipole expansion, however, it is always less accurate than the Fourier method and is therefore

not calculated. The direct contribution in the multipole expansion is defined as:

$$J_{\alpha\nu} = \sum_{k,l,\mu} (\alpha\nu|k) f^{(k,l)}(R_{\mu\nu})(k|\mu\mu) \tag{4.30}$$

The required components to assemble the multipole contribution are the analytic derivatives of the potential, $f^{(k,l)}$, the local orbital coefficients $L_{\alpha\mu}$ and the Cartesian multipole moments $(\alpha\beta|k)$ where k is the multipole level which ranges from $0-4$. Only contributions with large $|\mathbf{R}_{\mu\nu}|$ are assembled through the multipole method, this is controlled by the array MP_{Map} which is the converse of FT_{Map} such that every $\mu\nu$ pair is accounted for between the methods. To obtain the regularized multipole expansion the integrals $(\alpha\beta|0)$ are removed from the sum. For a full derivation of the Cartesian multipole expansion see chapter 3 and the appendix. Only lines 7-14 scale quadratically with respect system size and the term $I_2(\nu,l)$ is of dimension, number of occupied orbitals by multipole level which will always fit in memory.

Algorithm 3 Multipole Contributions

1: **for** μ **do** ▷ Each local orbital μ
2: **for** k **do** ▷ Each multipole level k, skip $k=0$ if regularizing
3: Obtain $(\alpha\beta|k)(\mathbf{R}_\mu)$ ▷ Expansion centre is $\mathbf{R}_\mu$
4: $I_1(\mu,k)=\sum_{\alpha\beta}(\alpha\beta|k)L_{\alpha\mu}L_{\beta\mu}$
5: **end for**
6: **end for**
7: **for** ν **do**
8: **for** μ **do**
9: **if** $\mu\nu\in MP_{map}$ **then**
10: Obtain $f^{l,k}_{\mu\nu}=f^{l,k}(\mathbf{R}_{\mu\nu})$
11: $I_2(\nu,l)+=\sum_k f^{l,k}_{\mu\nu}I_1(\mu,k)$
12: **end if**
13: **end for**
14: **end for**
15: **for** ν **do**
16: **for** l **do** ▷ Skip $l=0$ if regularizing
17: Obtain $(\alpha\beta|l)(R_\nu)$ ▷ Expansion centre is $\mathbf{R}_\nu$
18: Obtain $(\alpha\nu|l)=\sum_\beta(\alpha\beta|l)L_{\beta\nu}$
19: $J_{\alpha\nu}=\sum_l(\alpha\nu|l)\cdot I_2(\nu,l)$
20: **end for**
21: **end for**
22: Repeat lines 15 to 21 using local virtual orbitals to obtain J_{VV}
23: $J=J_{\alpha O}+J_{O\alpha}-J_{OO}+J_{VV}$

4.2.4 Assembly of the One-Body Contributions

The one-body contributions are only assembled if the potential is regularized. Here we assume that the Cholesky coefficients $L_{\alpha\mu}$ and the overlap matrix in mixed representation are available in memory. The mixed direct term and the AO exchange term are:

$$J_{\alpha\nu}=\sum_\mu V_{\alpha\nu}(\mathbf{R}_\mu)S_{\mu\mu}+S_{\alpha\nu}V_{\mu\mu}(\mathbf{R}_\nu)-S_{\alpha\nu}S_{\mu\mu}V(\mathbf{R}_\mu-\mathbf{R}_\nu) \tag{4.31}$$

$$\begin{aligned}K_{\alpha\beta}&=\sum_\mu V_{\alpha\mu}(\mathbf{R}_\mu)S_{\beta\mu}+S_{\alpha\mu}V_{\beta\mu}(\mathbf{R}_\mu)-S_{\alpha\mu}S_{\mu\mu}V(\mathbf{R}_\mu-\mathbf{R}_\mu)\\&=\sum_\mu V_{\alpha\mu}(\mathbf{R}_\mu)S_{\beta\mu}+[\sum_\mu V_{\alpha\mu}(\mathbf{R}_\mu)S_{\beta\mu}]^\intercal-S_{\alpha\mu}S_{\mu\mu}V(0)\end{aligned} \tag{4.32}$$

The two-range potential is defined to have $V(\mathbf{R}=0)=0$ such that the exchange contribution

can be simplified to single term plus its transpose. Depending on the specific potential used, V will change, for the two-range potential V is V_{lr} but for the three-range potential V would be V_{ulr}. The algorithm is consistent independent of the potential used therefore V is used for the potential. The integrals $V_{\alpha\beta}(\mathbf{R}_\nu)$ are nuclear-electron like repulsion integrals with a unit charge located at $\mathbf{R}_\nu$, these integrals are long range and extend throughout the whole molecule.

The assembly of the one-body contributions are outline in algorithm 4. The master loop is a loop over μ which is the total number of electrons. The exchange can fully be calculated in this single loop. This is the big advantage to the regularization as the one-body exchange contribution is simple to calculate and it is the largest contribution to the long-range exchange energy.

The direct term is more complicated, it needs to be calculated in two parts $J_{\alpha\mu}$ and $J_{\alpha\bar{\mu}}$ such that it can be projected back into the AO space. This algorithm requires a loop over ν and a loop over $\bar{\nu}$. The algorithm scales quadratically with respect to system size which may become cumbersome for large systems. It is still cheaper to calculate the one-body contribution than the DF/Fourier contributions for small to large systems. For very large systems this may become a bottleneck.

Algorithm 4 One-body Fock Contributions

1: Obtain $S_{\gamma\mu}$
2: Obtain $S_{\gamma\bar{\mu}}$
3: **for** μ **do** ▷ Each Cholesky Vector μ
4: Obtain $V_{\alpha\beta}(\mathbf{R}_\mu)$ ▷ V is dependent on the specific potential used
5: Transform to mixed rep. $V_{\alpha\mu}(\mathbf{R}_\mu) = \sum_\beta V_{\alpha\beta}(\mathbf{R}_\mu) L_{\beta\mu}$ ▷ used only for the exchange
6: $I_k(\alpha\gamma) = \sum_\mu V_{\alpha\mu}(\mathbf{R}_\mu) S_{\gamma\mu}$
7: Increment the exchange $K_{\alpha\gamma} += I_k(\alpha\gamma) + I_k^\intercal(\alpha\gamma)$
8: $V_{\alpha\nu}(\mathbf{R}_\mu) = \sum_\beta V_{\alpha\beta}(\mathbf{R}_\mu) L_{\beta\nu}$ ▷ Needed for J
9: $V_{\alpha\bar{\nu}}(\mathbf{R}_\mu) = \sum_\beta V_{\alpha\beta}(\mathbf{R}_\mu) L_{\beta\bar{\nu}}$ ▷ Needed for J_{VV}
10: **end for**
11: **for** ν **do** ▷ Each Cholesky Vector ν
12: Obtain $V_{\alpha\beta}(\mathbf{R}_\nu)$ ▷ V is dependent on the specific potential used
13: $V_{\mu\mu}(\mathbf{R}_\nu) = \sum_{\mu,\alpha,\beta} V_{\alpha\beta}(\mathbf{R}_\nu) L_{\alpha\mu} L_{\beta\mu}$ ▷ This is a constant
14: $J_{\alpha\nu} = \sum_\mu S_{\alpha\nu} V_{\mu\mu}(\mathbf{R}_\nu) + V_{\alpha\nu}(\mathbf{R}_\mu) S_{\mu\mu} - S_{\alpha\nu} S_{\mu\mu} V(\mathbf{R}_\mu - \mathbf{R}_\nu)$
15: **end for**
16: Repeat lines 11 to 15 using local virtual orbitals to obtain J_{VV}
17: $J = J_{\alpha O} + J_{O\alpha} - J_{OO} + J_{VV}$

4.3 SCF Convergence Analysis

The SCF code used in this thesis is PySCF's SCF code[108] with the exception that the Fock matrix is assembled in the custom code using the algorithms described above. Common SCF features such as direct inversion of the iterative subspace (DIIS) are used as they are built into PySCF's SCF engine. The initial guess density from PySCF is not idempotent and this is a poor guess for the range-separated integrals because they rely on local orbitals from the Cholesky decomposition of the density matrix. This guess is not ideal but only adds an extra cycle or two to the SCF method.

The molecules used in this analysis are glycine, toluene and a glycine dimer separated by 10 $\mathring{A}$ to capture the multipole contribution when small α values are used. These molecules are described in more detail in Chapter 4 section "Computational Details". Results from calculations on molecules from the GMTKN55 database[110–112] are presented in the final section to convince the reader that the method works with the recommended parameters on a large variety of molecules.

The two most important parameters for controlling the range-separated SCF method are the range separation parameter α, and the sparsity threshold T_{sparse}. The range separation parameter controls the partitioning of the short and long-range regions. The smaller the value of α the further the short-range regions extends. The Fourier integrand $\eta(\mathbf{g})$ is also smoother for smaller α values and can be integrated accurately with a smaller numerical grid compared to large α values. The sparsity threshold is essential for linear/low order scaling since it controls how many small integrals are discarded. If T_{sparse} is too loose the accuracy of the integrals and hence the SCF method will decrease, if T_{sparse} is too tight then there is a large loss in efficiency which can be detrimental for the calculation of large systems.

4.3.1 Hartree-Fock Exchange Contribution

In this subsection, the HF direct term is calculated without range separation using the J engine from PySCF and is assumed to be error free. The HF exchange is calculated with the range-separated method, the PySCF full-range exchange matrix is considered to be the answer and is used to report on the accuracy of the range-separated method, errors are reported as

log errors.

$$Log(Error) = Log_{10}(|\text{PySCF} - \text{Range-Separated}|) \tag{4.33}$$

The long-range exchange is only calculated with the Fourier method which makes it much simpler than the HF direct contribution. The two parameters that control the exchange contribution are the range-separated parameter and the size of the numerical grid. Figure 4.2 depicts the log of the HF energy error as a function of numerical grid size (N-Grid) for the glycine molecule in the ccpvtz basis. The HF energy error is the converged HF energy error and N-Grid is the Lebedev order and number of equidistant radial points.

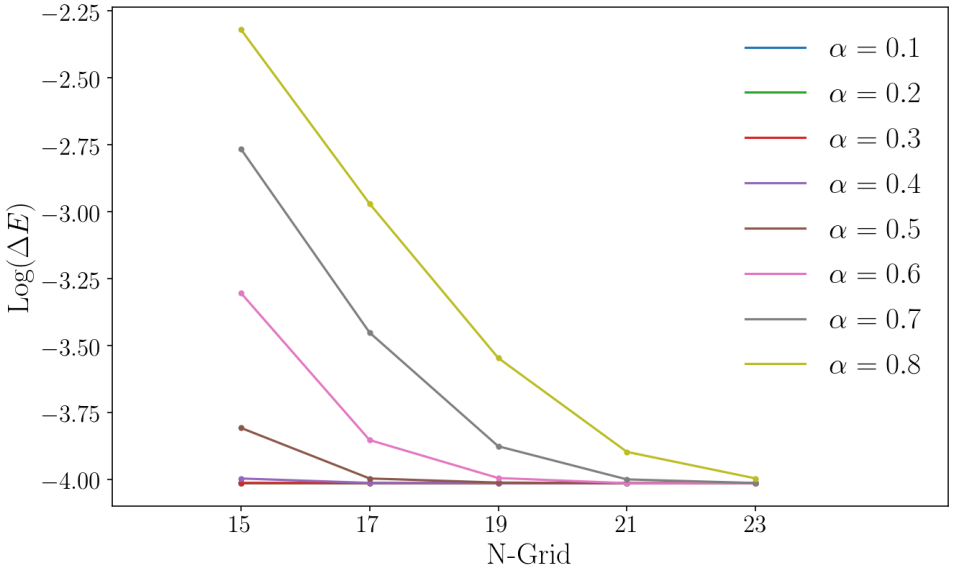

Figure 4.2: The Log of the HF energy error at convergence for the glycine molecule in the ccpvtz basis, the PySCF converged energy is the used as the reference energy. Several values of α are shown.

One can see that for the larger α values a larger numerical grid is needed to obtain accurate results. For the largest α value of 0.8, the accuracy of the method reaches an error of $\approx 10^{-4}$ Hartree at an N-grid of 23, which is still considered a medium size numerical grid containing 4268 grid points. The size of the numerical grid is independent of the system size and the molecule of interest as shown in chapter 3. The errors converge to $\approx 10^{-4}$ because this is the error associated with DF; the accuracy of the Fourier method is better than the accuracy of DF.

Figure 4.2 demonstrates that an aggressive α can be used with a small-medium sized numerical

grid. The advantages of a large α are that the short-range potential is very short-range, greatly increasing the sparsity of the short-range DF integrals. To obtain an accurate calculation, the size of the numerical grid should increase with increasing α however the gird size does not need to be overly large.

The range-separated exchange is relatively simple to construct and can be assembled in linear scaling algorithms. This is quite powerful because the long-range exchange is conventionally the troublesome term to calculate.

4.3.2 Hartree-Fock Direct Contribution

In this subsection the HF direct and exchange contributions are assembled through the range-separated methods.

Range Separation Parameter

Figure 4.3 displays the convergence of the SCF energy for the glycine dimer using the ccpvtz basis set. The glycine monomers are separated by $10\mathring{A}$ such that there is at least some contribution from J_{mp} for all α's. One can easily see that only range separation parameters to converge are α=0.1 and α=0.2, these curves are dashed for clarity. The larger α values did not converge and the range-separated method should not be used for the direct contribution if α is greater than 0.2.

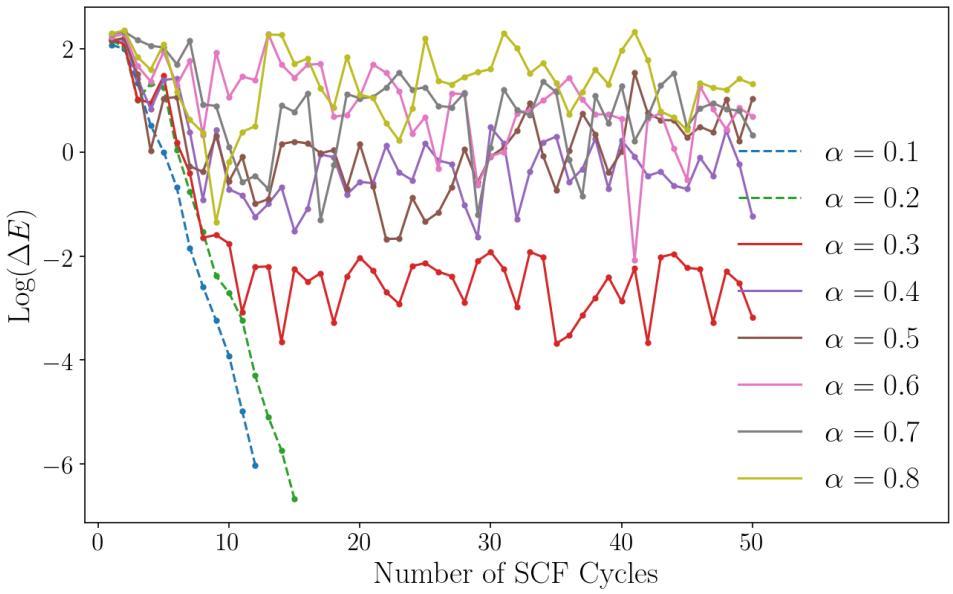

Figure 4.3: The convergence of the HF energy for each SCF cycle. The molecule is a glycine dimer in the ccpvtz basis with a numerical grid of size (23,23).

The rational we have for the poor convergence of $\alpha \geq 0.3$ is that the accuracy of the direct integrals is poor as was shown in chapter 3. It might be possible to converge the larger α values if very large numerical grids are chosen. The recommended value for α should be between 0.1 and 0.2 for the direct term.

Numerical Grid

The accuracy of the direct term is heavily dependent on the numerical grid size. In the previous chapter it was shown that the direct term is the troublesome term and the long-range HF exchange is evaluated accurately even for very small grids. This is not the case for the direct term, large grids should be used.

Table 4.1 presents data from calculations of the glycine dimer in the ccpvtz basis with increasing grid sizes, the range separation parameter is 0.2 in this case. The angular order and number of radial points were the chosen to be the same value, N-Grid. The smaller grids do not integrate the long-range Fock contribution accurately and the converged SCF energy is very far from the PySCF converged energy. The SCF procedure converges independent of the grid chosen. A grid size of 23 or larger is needed to converge the HF energy to within a milli-Hartree of the PySCF answer.

N-Grid	Number of Grid Points	SCF Cycles	HF Energy Error (Hartree)	Density Matrix Error (Max ΔD)
15	1290	17	3.10	0.15
17	1870	15	0.38	0.03
19	2768	15	6.41×10^{-3}	1.29×10^{-3}
21	3570	15	3.68×10^{-3}	8.84×10^{-4}
23	4462	15	-7.80×10^{-4}	1.32×10^{-3}
27	7176	15	1.51×10^{-4}	4.16×10^{-4}
31	10500	15	-1.57×10^{-5}	6.36×10^{-5}
35	14750	15	5.04×10^{-5}	5.96×10^{-5}
41	23516	15	3.28×10^{-8}	5.59×10^{-5}

Table 4.1: Convergence data for the glycine dimer in the ccpvtz basis, a range separation parameter of $\alpha=0.2$ was used.

The recommended grid size depends on the value of α and for $\alpha=0.1$ an N-Grid value of 23 should be sufficient.

Regularization

The regularized long-range potential is an interesting approach. The regularization procedure explicitly removes the $\frac{1}{\mathbf{g}^2}$ singularity in the Fourier transform and is described in more detail in the previous two data chapters. The regularized Fourier and multipole integrals require local orbitals and local gauge centres but are straight forward to calculate. The regularized direct term is composed of three contributions; short-range, regularized long-range and analytical one-body contributions.

$$\begin{aligned} J &= J_{sr} + J_{lr} \\ &= J_{sr} + J_{lr}^{R} + J_{1Body} \end{aligned} \tag{4.34}$$

Table 4.2 presents the energy breakdown of the regularized direct and exchange energy contributions. One can see that the regularized contributions are very small, less than 0.1 percent for both the direct and exchange contributions. The unfortunate result is that 0.01% of the direct energy contribution is equivalent to about 200 milli-Hartree so it is important that this contribution is calculated accurately.

	Direct Energy Contribution			Exchange Energy Contribution		
Molecule	Short-Range	Regularized Long-Range	one-body	Short-Range	Regularized Long-Range	one-body
Glycine	98.53	0.02	1.45	99.92	0.03	0.05
Toluene	96.98	0.02	3.00	99.81	0.07	0.11
Benzophenone	90.18	0.01	9.80	99.74	0.11	0.15
12-amino Dodecahexanone	85.38	0.01	14.61	99.85	0.06	0.09

Table 4.2: A break down of the energy contribution for the regularized approach. Molecules were calculated in the ccpvtz basis with α=0.2 and a numerical grid of N-Grid = 23.

The accuracy of the regularized integrals is on the same order as the accuracy of the unregularized integrals as shown in the previous chapter. There is no clear significant gain to using regularized integrals in the conventional SCF procedure. However, it is advantageous to converge the SCF method without the regularized long-range contributions, that is, to approximate the long-range direct contribution as the one-body contribution.

$$J \approx J_{sr} + J_{1Body} \tag{4.35}$$

During the final SCF cycle, J_{lr}^{R} should be reintroduced to obtain an accurate SCF calculation. The advantage to this method is that the expensive calculation of the regularized long-range integrals is avoided for all but the final SCF cycles, which leads to much faster SCF iterations. The disadvantage is that this is an approximation so the convergence thresholds need to be loosened slightly, a value of 10^{-6} for the SCF energy convergence was used instead of the default 10^{-10}. In practice, one does not need to converge the SCF to such tight thresholds unless SCF energy is very flat in the orbital rotation space.

Table 4.3 tabulates three schemes for evaluating J and K during the SCF iterations. These schemes neglect different small contributions. The final SCF iteration is always done without neglecting terms. Scheme 1 is the range-separated potential without regularization and is used for reference. Schemes 2 and 3 are the approximate schemes where the regularized Fourier contributions are neglected in all but final SCF iteration.

Scheme 1	$J \approx J_{sr} + J_{FT} + J_{MP}$	$K \approx K_{sr} + K_{FT}$
Scheme 2	$J \approx J_{sr} + J_{MP}^{R} + J_{1Body}$	$K \approx K_{sr} + K_{FT}$
Scheme 3	$J \approx J_{sr} + J_{MP}^{R} + J_{1Body}$	$K \approx K_{sr} + K_{1Body}$

Table 4.3: Regularized SCF schemes

Figure 4.4 depicts the errors of the converged SCF energy and density matrix for the glycine molecule in the ccpvtz basis using schemes 1-3. One can see that scheme 1 is acceptable for α values between 0.1-0.2, an acceptable error is about 0.1 milliHartree as this is roughly the error introduced from the DF approximation. For scheme 2 and 3 one can see that for $\alpha < 0.2$ the accuracy is reasonable, for larger α this approximation should not be used as the error is much larger than the DF error. It is important to note that approximating the long-range exchange as K_{1Body} has very little effect on the accuracy especially for small α.

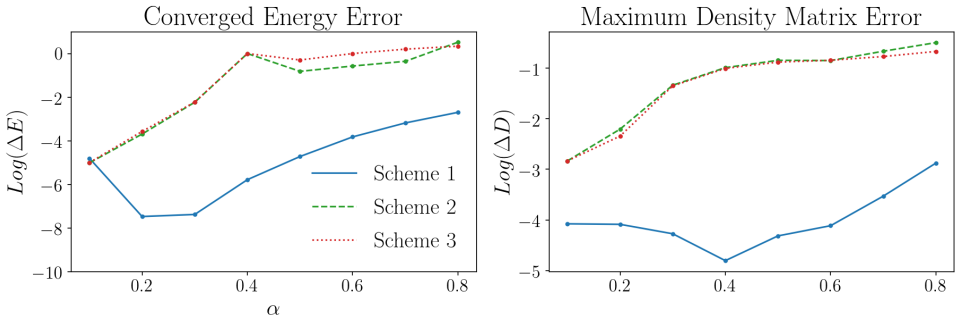

Figure 4.4: The converged SCF energy error (left) and the maximum error in the density matrix (right) for the glycine molecule in the ccpvtz basis.

Figure 4.5 depicts a similar plot as figure 4.4 but for the glycine dimer. In scheme 2 and 3 the regularized multipole contribution is included in each SCF cycle since J_{mp} is inexpensive to calculate. The density matrix obtained from scheme 2 and 3 is a better density matrix than the density matrix obtained from scheme without regularization, which is an interesting result.

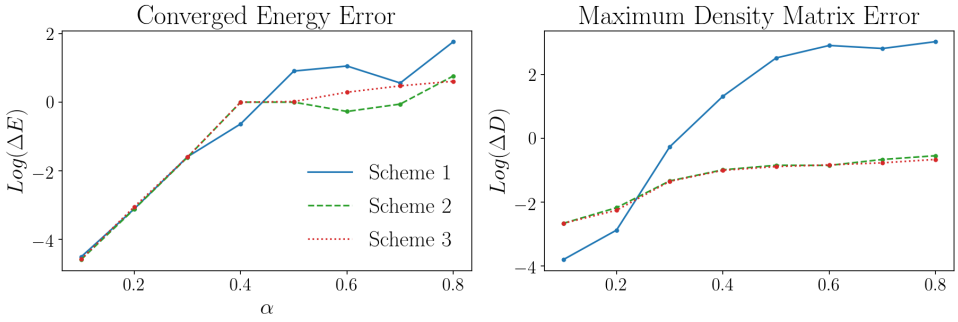

Figure 4.5: The converged SCF energy error (left) and the maximum error in the density matrix (right) for the glycine dimer in the ccpvtz basis.

The regularized potential can be used to converge the SCF method accurately provided that the range separation parameter α is small (< 0.2). By ignoring the expensive regularized

contributions the speed of SCF iterations are greatly increased which is largely advantageous for procedures such as geometry optimizations. The one-body contributions do not scale linearly with respect to system size and this will become a bottleneck for very large systems. A recommended scheme would be to partition long-range direct interactions into near, far and very far contributions. The near contributions can be ignored because the long-range potential is flat near $\mathbf{r}_{12}=0$, the far contributions should be approximated with the one-body contributions, the regularized contributions in this region can be ignored. The very far interactions should be calculated without regularization through the multipole contribution. The final SCF iteration should include all integrals.

4.3.3 Sparsity Thresholds

The sparsity threshold is the threshold that is used to determine if a value is to be pruned from a sparse matrix, this threshold is important for efficiency. Table 4.4 presents the number of SCF cycles required to converge the HF energy to 10^{-10} Hartree, which is the default convergence threshold. For very loose sparsity thresholds of 10^{-6}, many integrals are discarded and the SCF method struggles to converge, which is the expected result. For tighter thresholds the SCF method converges without issue. There are many unnecessarily small integrals included with tight threshold which do not significantly contribute to the calculation. For example the Benzophenone molecule in a minimal basis has approximately 22% more Fourier integrals when a threshold of 10^{-15} is used compared to the looser threshold of 10^{-8}, both calculations converge the SCF energy to the same accuracy. A balance between efficiency and precision is to use a sparsity threshold around 10^{-8}.

Sparsity Thresh	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-11}	10^{-13}	10^{-15}
Glycine	15	15	15	15	15	15	15
Toluene	Did not converge	14	14	14	14	14	14
Benzophenone	Did not converge	21	20	20	20	20	20
12-amino Dodecahexanone	Did not converge	19	20	19	19	19	19

Table 4.4: The number of cycles required to converge the SCF method. The 6-311g basis was used with an α value of 0.2 with a medium size numerical grid of size (23,22).

4.4 Future Improvements

The current code utilizes sparse-sparse matrix operations for efficiency and the sparsity in the AO integrals is key to achieving low order scaling the in HF algorithm. The most expensive step in a SCF calculation is the recalulation of the two-electron integrals, in our case this is the recalculation of the range-separated integrals, many of which are sparse. In an optimal code this sparsity should be taken advantage of by first partitioning the three centre integrals $(\alpha\beta|k)$ into blocks around their tile centres $(\alpha_A\beta_B|k_K)$. The label k is used for general three centre index, for short-range integrals this will be x the auxiliary index, g for Fourier integrals and l for multipole. The idea would be to label the $(\alpha_A\beta_B|k_K)$ integral blocks as either sparse or dense blocks and to keep the sparse blocks in memory or on disk. Only the dense integrals would be recalculated at every step, this would be a significant time save.

A more complicated approach would be to partition the direct term into close, medium, far and very far interactions. The close interactions would be assembled with the Fourier method, the medium interactions would use the regularized Fourier method and the far interactions would use the regularized multipole method. This method would ensure that the switching occurs between regularized Fourier and regularized multipole methods which is the best methods to switch between because they are the smallest. The very far interactions should be calculated with the pure multipole method such that the one-body terms do not become a bottleneck in very large systems.

4.5 Conclusion

The range-separated SCF procedure behaves as expected. If there are large errors in the AO integrals then the SCF procedure will struggle to converge. If accurate integrals are used the SCF procedure will converge, provided the convergence criterion is looser than the numerical precision. The exchange contribution can be assembled with an aggressively large α and a medium numerical grid. The is very powerful because the long-range exchange is conventionally the expensive term to calculate.

The long-range direct contribution is more difficult to calculate accurately, only small α values with large numerical grids should be used, we recommended an α value of 0.1-0.3 with a

medium to large numerical grid. For small α values the long-range direct contribution can be approximated as J_{1body} which is much cheaper to calculate. The SCF method can be converged with this approximation. Post HF methods in practice do not require a highly converged HF calculation, it is usually acceptable to preform a post-HF calculations starting from a reasonably converged SCF calculation.

To convince the reader that the range-separated method works with a large variety of molecules the log energy errors (PySCF is used as the reference energy) of GMNTK database molecules are presented in figure 4.6. This data used the converged PySCF density matrix such that a larger basis set (ccpvtz) could be used. The calculations were run with an α value of 0.1 and numerical grid size of Lebedev order 23 and 22 equidistant radial points.

As one can see all errors are below -4.5 which is accurate to 0.1 milli-Hartree, most calculations are more accurate than this.

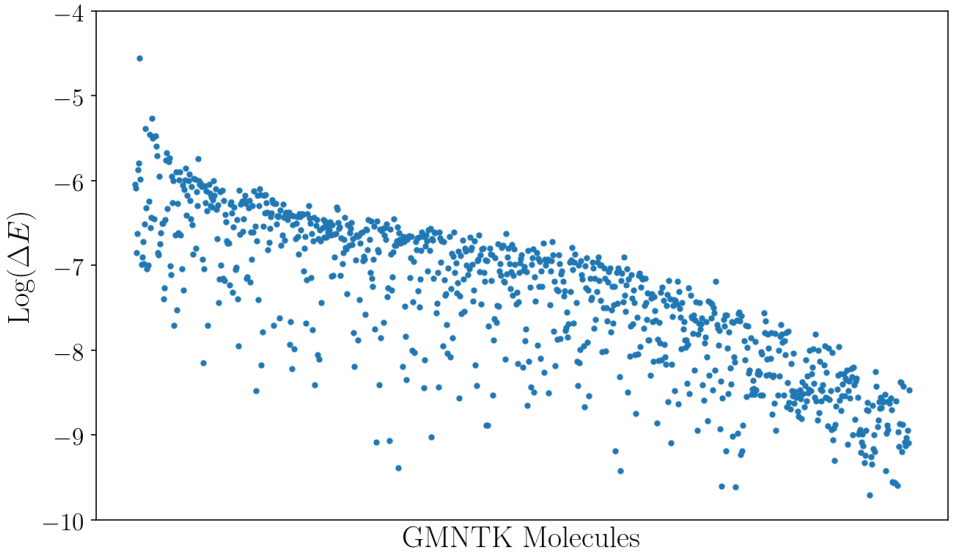

Figure 4.6: The log error of molecules from the GMNTK database in the ccpvtz basis. The range-separated Fock matrix was calculated with the converged PySCF density matrix as input. Errors are relative to the PySCF energy. the value of α is 0.1 and the numerical grid is (23,22), the Fourier method is used for direct contributions within 24Å and the multipole method was used for contributions larger than 24Å. The molecules are sorted by their largest orbital extents, from largest to smallest.

Most of the accuracy issues are associated with the calculation of the direct term, conventionally the direct term is the simpler term, it is possible to use modern J engines to calculate the direct term and use range-separated methods to calculate the exchange term. The short-range exchange can be calculated in modern K engines. The long-range exchange can be calculated

using the Fourier approach which is linear scaling and parallelizable. This is advantageous because the exchange can be calculated with a large α and a small numerical grid.

Chapter 5

Conclusions and Outlook

5.1 Conclusions

This thesis establishes groundwork for a range-separated Hamiltonian that one day might be useful for range-separated correlated calculations. The three data chapters in this thesis build on each other starting with the Coulomb potential in chapter 2, then molecular integrals in chapter 3, and concluding with range-separated integral direct Hartree-Fock in chapter 4.

The range-separated potential is introduced by partitioning the Coulomb potential into short and long-range regions using Ewald partitioning. The partitioning is controlled by a single parameter α called the range separation parameter. Gaussians are added and subtracted from the short and long-range potentials using optimal parameters related to α to ensure that the long-range potential is zero at $\mathbf{r}_{12}=0$ and that it is as flat as possible for small $\mathbf{r}_{12}$. The typical range of α is between 0.1 and 0.8, the smaller the value of α, the further the short-range potential extends.

$$V_{sr}(\mathbf{r}_{12}) = \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{12}|)}{|\mathbf{r}_{12}|} + X_0 e^{-\gamma|\mathbf{r}_{12}|^2} \tag{5.1}$$

$$V_{lr}(\mathbf{r}_{12}) = \frac{1}{|\mathbf{r}_{12}|} - V_{sr}(|\mathbf{r}_{12}|) \tag{5.2}$$

The short-range potential is well behaved and the treatment of the long-range potential is the focus of this chapter. The long-range potential can be evaluated numerically either through a Fourier transform or through a multipole expansion. The Fourier transform of the long-range potential introduces a $\mathbf{g}^{-2}$ singularity which is cancelled out by choosing to evaluate the integral in spherical coordinates. The Fourier transform takes on the general form

$$V_{lr}(\mathbf{r}_{12}) = \int dg^3 \eta(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}_{12}} \tag{5.3}$$

The function $\eta(\mathbf{g})$ depends on the specific radial potential used. The transform is integrated numerically using a numerical grid defined by radial and angular grid points. Lebedev quadrature is used to define the angular grid and equidistant radial points are used for the radial quadrature. Other radial quadratures such as Gauss-Hermite, Gauss-Legendre and Gauss-Chebyshev can also be used but no significant gain over the equidistant grid was found. The numerical weights are determined by the function $\eta(\mathbf{g})$, which for the long-range potential quickly decays limiting the range of $\mathbf{g}$, the extent of the radial grid does not need to increase with increasing $\mathbf{r}_{12}$. The term $e^{i\mathbf{g}\cdot\mathbf{r}_{12}}$ is troublesome for large $\mathbf{r}_{12}$ because it oscillates rapidly. The Fourier method will always breakdown at large enough $\mathbf{r}_{12}$ which is why the multipole method is essential for very long-range interactions. The multipole expansion is evaluated in Cartesian coordinates for simplicity even though there is some redundant information. The long-range potential does not satisfy the Laplace equation and therefore conventional multipole expansions are not immediately applicable.

Chapter 2 introduces the regularization method to further partition the long-range potential and it is used throughout the thesis. The regularized long-range potential utilizes fixed points called gauge centres ($\mathbf{R}_i$) defined as points near the coordinates $\mathbf{r}_i$ to explicitly remove the singularity in the Fourier transform. The resulting long-range potential consists of a regularized long-range potential V_{lr}^{R}, and an analytical one-body potential V_{1body}. The one-body potential makes up approximately 99% of the long-range potential while the regularized long-range potential contributes only about 1% of the long-range potential. The regularized long-range potential is treated with the same methods as the long-range potential (Fourier and multipole) but it is significantly smaller.

The potentials and methods defined in chapter 2 establish the groundwork needed to introduce the range-separated molecular integrals which is the theme of chapter 3. The two-electron repulsion integrals are the main focus of this chapter, they are partitioned into three parts.

$$\begin{aligned}(\alpha\beta|\gamma\delta)_{2range} = &\sum_{xy}(\alpha\beta|x)_{sr}M_{xy}^{-1}(y|\gamma\delta)_{sr} \\ &+\sum_{\mathbf{g}}(\alpha\beta|\mathbf{g})\Theta(\mathbf{R}_{12})\eta(\mathbf{g})(\gamma\delta|\mathbf{g})^{*} \\ &+\sum_{mn}(\alpha\beta|m)f^{mn}(\mathbf{R}_{12})\bar{\Theta}(\mathbf{R}_{12})(\gamma\delta|n)\end{aligned} \tag{5.4}$$

The short-range integrals are represented through density fitting, which is commonly used in large scale quantum chemistry packages. The long-range integrals are assembled through Fourier and multipole methods similar to how the bare potential is assemble in chapter 2. The term $\Theta(\mathbf{R}_{12})$ is a switching function used to determine which method is used for a specific long-range integral. The switch is dependent on $\mathbf{R}_{12}$ which is the distance between the gauge centres $\mathbf{R}_1$ and $\mathbf{R}_2$. In equation 5.4 these gauge centres would be the atomic orbital (AO) gauge centres $\mathbf{R}_\alpha$ and $\mathbf{R}_\gamma$ which are poorly defined and ambiguous. Local orbitals $|\mu\rangle$ are introduced in this chapter as a means solve this ambiguity in the gauge centre definitions. The long-range integrals can also be regularized using a similar method described in chapter 2, local orbitals are essential for the regularization of the long-range integrals. Regularized Fourier and regularized multipole representations are derived in this chapter as part of the regularization method. The primitive three centre integrals are the same if the potential is regularized or unregularized, this is convenient for development because these integrals are readily available in most integral packages. These methods are general for any radial potentials, we are not restricted to the two-range potential.

The amount of data in the short-range and Fourier integrals scale linearly with respect to system size. The linear scaling in the short-range integrals arises from the added sparsity in the auxiliary index, which is not present in full-range DF integrals. The Fourier integrals scale linearly because the size on the numerical grid does not need to increase with respect to system size, this is because $\eta(\mathbf{g})$ quickly decays as a function of $\mathbf{g}$. The multipole and one-body integrals do not scale linearly with respect to system size but they are much cheaper to evaluate than the DF and Fourier methods. Integral pre-screening and sparse matrices are essential to obtain linear scaling in the three centre integrals, both are discussed in chapter 3.

The accuracy of the methods is reported using the accuracy of the Hartree-Fock direct and exchange matrices because the four centre integrals do not fit in memory and are never fully assembled in practice. The long-range exchange is assembled though only the Fourier transform, it was found that a large α and a medium numerical grid were sufficient to accurately represent the exchange. The long-range direct term is more complicated and requires the Fourier method for short/medium range contributions and the the multipole method for long-range interactions. To evaluate the direct term accurately one needs to use a small α and a sizeable numerical grid.

A range-separated Hartree-Fock method is presented in the final data chapter. The range-separated integrals are used to assemble the HF direct and exchange contributions. In principle it is possible to assemble the V^{Ne} integrals contributing to the Fock matrix but these integrals are typically orders of magnitude cheaper to calculate. The direct and exchange contributions are represented by three terms if regularization is not used.

$$J = J_{DF} + J_{FT} + J_{mp} \tag{5.5}$$

$$K = K_{DF} + K_{FT} \tag{5.6}$$

or four terms if regularization is used.

$$J = J_{DF} + J^R_{FT} + J^R_{mp} + J_{1body} \tag{5.7}$$

$$K = K_{DF} + K^R_{FT} + K_{1body} \tag{5.8}$$

Low order scaling algorithms to assemble all contributions are presented in this chapter. The algorithms rely on sparse matrix multiplies and integral screening to achieve this scaling. The code used to assemble these contributions is unoptimized and hard data supporting the low order scaling is not available at this time. The regularization of the methods is controlled through a single boolean switch and it is possible to pick and choose which interactions are regularized.

In principle, the short-range DF contributions can be assembled in modern J/K engines by swapping integral codes, the added sparsity from the short-range integrals can only improve the efficiency of these codes. The long-range exchange is well behaved, it can be assembled using an aggressive α and a medium sized numerical grid. This is very powerful because typically the long-range exchange is expensive to calculate and it does not scale linearly with respect to system size. In our representation the long-range exchange is straightforward to calculate and it can be assembled in a highly parallelized algorithm with a medium sized numerical grid that does not need to increase with respect to system size. The large α value will ensure that the short-range integrals are very sparse increasing the efficiency of the DF code as well.

The direct term is difficult in this representation as it requires both the Fourier and multipole methods, local occupied and local virtual orbitals and proper projections to obtain the correct

AO J matrix. On top of this, the accuracy is not very good unless small α values are used with moderate numerical grids. There are many excellent linear scaling algorithms in the literature to construct J such as the continuous fast multipole method (CFMM)[70,71] and J-engine method[73] and the Fourier-transform Coulomb (FTC)[75–78] method. Advantages of using the the range-separated direct term over these conventional methods is not clear at this time.

The Fourier contribution is the most expensive long-range term to calculate, the same is true if regularization is used. In the regularization scheme J_{lr}^{R} and K_{lr}^{R} are the smallest contributions and it is possible to converge the SCF calculation without these integrals, called the regularized SCF approximation.

$$J = J_{DF} + J_{mp}^{R} + J_{1body} \tag{5.9}$$

$$K = K_{DF} + K_{1body} \tag{5.10}$$

Using this approximation the converged density matrix is similar enough to the correct density matrix that accurate HF energies can be obtained from this approximate density matrix by including J_{lr}^{R} and K_{lr}^{R} in the final SCF iteration. This approximation is powerful since the expense of the long-range calculation is greatly reduced. This would be advantageous in expensive calculations such as geometry optimizations where one doesn't really need a highly converged SCF method at early stages. It is also possible to proceed with coupled cluster calculations if the HF calculation is qualitatively correct, which appears to be the case with this approximation.

To summarize, using the range-separated method, all the major issues lie in the calculation of the long-range direct contributions. This is because both the Fourier and multipole methods must be used to represent the whole long-range potential. Using both methods requires a switch which complicates the assembly and if not done correctly can ruin the overall accuracy of the method. The Fourier method is excellent for small $\mathbf{r}_{12}$, however at some point $\mathbf{r}_{12}$ will become large enough that the Fourier method will always breakdown. The multipole method has the opposite behaviour it is an excellent approximation for very long-range interactions but it does not approximate the smaller $\mathbf{r}_{12}$ region well. The region where the Fourier method breaks down and the the region where the multipole method starts to accurately represent the potential do not overlap well. If a region existed where both the Fourier and multipole methods

accurately represented the potential then there would not be major issues with accuracy. This was the idea behind the three range potential introduced in chapter 2.

$$V_{Coul}(\mathbf{r}_{12}) = V_{sr}(\mathbf{r}_{12}) + V_{inter}(\mathbf{r}_{12}) + V_{ulr}(\mathbf{r}_{12}) \tag{5.11}$$

Here the ultra long-range potential $V_{ulr}(\mathbf{r}_{12})$ could be accurately represented by the multipole expansion and the intermediate-range $V_{inter}(\mathbf{r}_{12})$ potential could be accurately represented by the Fourier transform, without the need for switching. In practice, through anecdotal evidence not reported in the thesis, this did not work and a switch was still required for the ultra long-range potential.

The method works excellently for the long-range exchange, and the regularized SCF approximation is quite powerful when tight SCF convergence is not needed.

5.2 Future Outlook

Some very useful lessons were learned while building a Hartree-Fock code using the range-separated integral representation. Work has started in the Nooijen research group on applying these new techniques to cluster in molecules (CIM) calculations, which we mentioned as a major long term goal in the introduction to this thesis. The first steps in a CIM calculation are a HF and MP2 calculation on the complete system. These steps are necessary to take into account long-range Coulomb effects both at the mean field and correlated level. For these steps we need all the machinery developed here and all three layers: Density fitting of short-range (3-center) integrals, Fourier representation at intermediate range and multipole expansions for the truly long range, including a switching mechanism between Fourier and Multipole. We have recently submitted a paper in which we describe an efficient algorithm to perform so-called Laplace-MP2 calculations[40], in which the cumbersome denominator in MP2 is replaced by a factorizable time integration (Laplace transform) that can be done using a 1d numerical integration with few points (8 or so). We use the same sparse matrix representations of all (local) quantities as pursued in the thesis and the amount of data processed scales linearly with the size of the system. This work on Laplace MP2 did not yet include the multipole layer and switching and this is currently in progress. We have established that it is needed.

The next steps in CIM are in principle easier, although a lot of work will be involved in the efficient implementation. The key step is to determine a domain of occupied and virtual orbitals associated with a particular localized orbital i, or a small subset of these, defining a central cell. In our group we have established that the exchange matrix associated with the central orbitals provides an excellent path to selecting the remaining occupied and virtual orbitals in the domain. The algorithm is a bit involved and requires additional exchange matrices to be calculated[40] (e.g. corresponding to the full domain of occupied orbitals, and not only the central orbitals). This allows for highly efficient calculations in the context of this thesis. We have amply demonstrated that accurate exchange matrices can be obtained using short-range integrals with a large α, implying they decay very fast, and a small grid for the Fourier transform. No multipole layer or cumbersome switching is required. Since the orbitals selected in a particular orbital domain are all spanning the same vicinity in space we can anticipate that accurate integrals can be calculated using the same two-layer strategy, fast decaying short-range and relatively low-level Fourier grid. Given the extents of the orbitals in the domain all primitive AO integrals and integral transformations can be assembled in a direct fashion using the screening techniques developed here. Most importantly, we do not anticipate complicated algorithms. The hard developmental work is done in the context of Hartree-Fock, and one should be able to focus on technical aspects of the implementation and focus on improved efficiency, coding up certain critical steps in Fortran or C++, rather than Python. This work will be left to a new generation of students and long-standing collaborator Ondrej Demel, who was the principal developer of the Laplace MP2 module.

Another interesting avenue concerns accurate correlated calculations for solids. The calculation of short-range interactions like Hartree-Fock exchange and the exchange terms in MP2 as well as local correlation effects using Coupled Cluster can most likely be done using the two-layer Short-range / Fourier representation of integrals. It is straightforward to include translational symmetry using k-vectors in the first Brillouin zone. The long-range direct Coulomb interaction in Hartree-Fock and MP2 is more cumbersome. It is likely that all four layers discussed in this thesis: Short-range, Fourier, Multipole and one-body should be deployed to create an accurate and efficient implementation for solids. This may require the blood, sweat and tears of another graduate student, dedicating a complete PhD trajectory to the project. It will be a challenge.

Once these building blocks are in place the generalizations from ground state calculations to

excited states, multireference situations can commence. The work never ends.

References

1. Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.
2. Shavitt, I.; Bartlett, R. J. *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*; Cambridge University Press, 2009.
3. Bartlett, R. J. Coupled-cluster theory and its equation-of-motion extensions. *WIREs Computational Molecular Science* **2012**, *2*, 126–138.
4. Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley, 2000.
5. Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
6. Cremer, D. Møller-Plesset perturbation theory: from small molecule methods to methods for thousands of atoms: Møller-Plesset perturbation theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 509–530.
7. Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *The Journal of Chemical Physics* **2009**, *130*, 114108.
8. Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of Chemical Physics* **2016**, *144*, 024109.
9. Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *Journal of Chemical Theory and Computation* **2015**, *11*, 1525–1539, Publisher: American Chemical Society.
10. Liakos, D. G.; Guo, Y.; Neese, F. Comprehensive Benchmark Results for the Domain Based Local Pair Natural Orbital Coupled Cluster Method (DLPNO-CCSD(T)) for Closed- and Open-Shell Systems. *The Journal of Physical Chemistry A* **2020**, *124*, 90–100, Publisher: American Chemical Society.
11. Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. 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.
12. Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **2013**, *138*, 034106.
13. Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *The Journal of Chemical Physics* **2015**, *143*, 034108.

14. Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable Electron Correlation Methods I.: PNO-LMP2 with Linear Scaling in the Molecular Size and Near-Inverse-Linear Scaling in the Number of Processors. *Journal of Chemical Theory and Computation* **2015**, *11*, 484–507, Publisher: American Chemical Society.

15. Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. *Journal of Chemical Theory and Computation* **2015**, *11*, 5291–5304, Publisher: American Chemical Society.

16. Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *Journal of Chemical Theory and Computation* **2017**, *13*, 3650–3675, Publisher: American Chemical Society.

17. Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals. *The Journal of Chemical Physics* **2017**, *146*, 174108.

18. Pavošević, F.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals. *The Journal of Chemical Physics* **2016**, *144*, 144109.

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

20. Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *The Journal of Chemical Physics* **1993**, *98*, 7029–7039.

21. Nooijen, M.; Bartlett, R. J. Similarity transformed equation-of-motion coupled-cluster theory: Details, examples, and comparisons. *The Journal of Chemical Physics* **1997**, *107*, 6812–6830.

22. Förner, W. Coupled cluster studies. IV. Analysis of the correlated wavefunction in canonical and localized orbital basis for ethylene, carbon monoxide, and carbon dioxide. *Chemical Physics* **1987**, *114*, 21–35.

23. Li, S.; Ma, J.; Jiang, Y. Linear scaling local correlation approach for solving the coupled cluster equations of large systems. *Journal of Computational Chemistry* **2002**, *23*, 237–244.

24. Li, S.; Li, W.; Fang, T. An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules. *Journal of the American Chemical Society* **2005**, *127*, 7215–7226.

25. Li, S.; Shen, J.; Li, W.; Jiang, Y. An efficient implementation of the "cluster-in-molecule" approach for local electron correlation calculations. *The Journal of Chemical Physics* **2006**, *125*, 074109.

26. Li, W.; Li, S.; Jiang, Y. Generalized Energy-Based Fragmentation Approach for

Computing the Ground-State Energies and Properties of Large Molecules. *The Journal of Physical Chemistry A* **2007**, *111*, 2193–2199.

27. Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled-Cluster and Møller+Plesset Perturbation Theories. *The Journal of Physical Chemistry A* **2010**, *114*, 6721–6727.

28. Li, W.; Guo, Y.; Li, S. A refined cluster-in-molecule local correlation approach for predicting the relative energies of large systems. *Physical Chemistry Chemical Physics* **2012**, *14*, 7854.

29. Li, W.; Li, S. Cluster-in-molecule local correlation method for large systems. *Science China Chemistry* **2014**, *57*, 78–86.

30. Guo, Y.; Li, W.; Li, S. Improved Cluster-in-Molecule Local Correlation Approach for Electron Correlation Calculation of Large Systems. *The Journal of Physical Chemistry A* **2014**, *118*, 8996–9004.

31. Li, W.; Ni, Z.; Li, S. Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems. *Molecular Physics* **2016**, *114*, 1447–1460.

32. Ni, Z.; Li, W.; Li, S. Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems: Fully Optimized Implementation of the Cluster-in-Molecule Local Correlation Approach for Electron Correlation Calculations of Large Systems. *Journal of Computational Chemistry* **2019**, *40*, 1130–1140.

33. Wang, Y.; Ni, Z.; Li, W.; Li, S. Cluster-in-Molecule Local Correlation Approach for Periodic Systems. *Journal of Chemical Theory and Computation* **2019**, *15*, 2933–2943.

34. Rolik, Z.; Kállay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *The Journal of Chemical Physics* **2011**, *135*, 104111.

35. Flocke, N.; Bartlett, R. J. A natural linear scaling coupled-cluster method. *The Journal of Chemical Physics* **2004**, *121*, 10935.

36. Kristensen, K.; Kjærgaard, T.; Høyvik, I.-M.; Ettenhuber, P.; Jørgensen, P.; Jansik, B.; Reine, S.; Jakowski, J. The divide–expand–consolidate MP2 scheme goes massively parallel. *Molecular Physics* **2013**, *111*, 1196–1210.

37. Kristensen, K.; Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. A Locality Analysis of the Divide–Expand–Consolidate Coupled Cluster Amplitude Equations. *Journal of Chemical Theory and Computation* **2011**, *7*, 1677–1694, Publisher: American Chemical Society.

38. Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. 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, Publisher: American Chemical Society.

39. Kjærgaard, T.; Baudin, P.; Bykov, D.; Eriksen, J. J.; Ettenhuber, P.; Kristensen, K.; Larkin, J.; Liakh, D.; Pawlowski, F.; Vose, A.; Wang, Y. M.; Jørgensen, P. Massively parallel and linear-scaling algorithm for second-order Møller–Plesset perturbation theory

applied to the study of supramolecular wires. *Computer Physics Communications* **2017**, *212*, 152–160.

40. Demel, O.; Lecours, M.; Habrovsky, R.; Noojen, M. Towards Laplace MP2 method using range separated Coulomb potential and Orbital Selective Virtuals [submitted]. **2021**,

41. Rebolini, E.; Baardsen, G.; Hansen, A. S.; Leikanger, K. R.; Pedersen, T. B. Divide–Expand–Consolidate Second-Order Møller–Plesset Theory with Periodic Boundary Conditions. *Journal of Chemical Theory and Computation* **2018**, *14*, 2427–2438, Publisher: American Chemical Society.

42. Ye, H.-Z.; Berkelbach, T. C. Fast periodic Gaussian density fitting by range separation. *The Journal of Chemical Physics* **2021**, *154*, 131104.

43. Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. Vanishing-Overhead Linear-Scaling Random Phase Approximation by Cholesky Decomposition and an Attenuated Coulomb-Metric. *Journal of Chemical Theory and Computation* **2017**, *13*, 1647–1655, Publisher: American Chemical Society.

44. McClain, J.; Sun, Q.; Chan, G. K.-L.; Berkelbach, T. C. Gaussian-Based Coupled-Cluster Theory for the Ground-State and Band Structure of Solids. *Journal of Chemical Theory and Computation* **2017**, *13*, 1209–1218, Publisher: American Chemical Society.

45. Patterson, C. H. Density fitting in periodic systems: Application to TDHF in diamond and oxides. *The Journal of Chemical Physics* **2020**, *153*, 064107.

46. Wang, X.; Lewis, C. A.; Valeev, E. F. Efficient evaluation of exact exchange for periodic systems via concentric atomic density fitting. *The Journal of Chemical Physics* **2020**, *153*, 124116.

47. McWeeny, R. Hartree-Fock Theory with Nonorthogonal Basis Functions. *Phys. Rev.* **1959**, *114*, 1528–1529.

48. Hernández, E.; Gillan, M. J.; Goringe, C. M. Linear-scaling density-functional-theory technique: The density-matrix approach. *Phys. Rev. B* **1996**, *53*, 7147–7157.

49. McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

50. Rudberg, E.; Rubensson, E. H. Assessment of density matrix methods for linear scaling electronic structure calculations. *Journal of Physics: Condensed Matter* **2011**, *23*, 075502.

51. Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Trace resetting density matrix purification in O(N) self-consistent-field theory. *The Journal of Chemical Physics* **2003**, *118*, 8611–8620.

52. Li, X.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J.; Schlegel, H. B. Density matrix search using direct inversion in the iterative subspace as a linear scaling alternative to diagonalization in electronic structure calculations. *The Journal of Chemical Physics* **2003**, *119*, 7651–7658.

53. Shao, Y.; Saravanan, C.; Head-Gordon, M.; White, C. A. Curvy steps for density

matrix-based energy minimization: Application to large-scale self-consistent-field calculations. *The Journal of Chemical Physics* **2003**, *118*, 6144–6151.

54. Helgaker, T.; Larsen, H.; Olsen, J.; Jørgensen, P. Direct optimization of the AO density matrix in Hartree–Fock and Kohn–Sham theories. *Chemical Physics Letters* **2000**, *327*, 397–403.

55. Ochsenfeld, C.; Head-Gordon, M. A reformulation of the coupled perturbed self-consistent field equations entirely within a local atomic orbital density matrix-based scheme. *Chemical Physics Letters* **1997**, *270*, 399–405.

56. Millam, J. M.; Scuseria, G. E. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *The Journal of Chemical Physics* **1997**, *106*, 5569–5577.

57. Jordan, D. K.; Mazziotti, D. A. Comparison of two genres for linear scaling in density functional theory: Purification and density matrix minimization methods. *The Journal of Chemical Physics* **2005**, *122*, 084114.

58. Challacombe, M. A simplified density matrix minimization for linear scaling self-consistent field theory. *The Journal of Chemical Physics* **1999**, *110*, 2332–2342.

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

60. Høst, S.; Olsen, J.; Jansík, B.; Thøgersen, L.; Jørgensen, P.; Helgaker, T. The augmented Roothaan–Hall method for optimizing Hartree–Fock and Kohn–Sham density matrices. *The Journal of Chemical Physics* **2008**, *129*, 124106.

61. Garza, A. J.; Scuseria, G. E. Comparison of self-consistent field convergence acceleration techniques. *The Journal of Chemical Physics* **2012**, *137*, 054110.

62. Kudin, K. N.; Scuseria, G. E.; Cancès, E. A black-box self-consistent field convergence algorithm: One step closer. *The Journal of Chemical Physics* **2002**, *116*, 8255–8261.

63. Almlöf, J.; Faegri Jr., K.; Korsell, K. Principles for a direct SCF approach to LICAO–MOab-initio calculations. *Journal of Computational Chemistry* **1982**, *3*, 385–399.

64. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters* **1989**, *162*, 165–169.

65. Hollman, D. S.; Schaefer, H. F.; Valeev, E. F. A tight distance-dependent estimator for screening three-center Coulomb integrals over Gaussian basis functions. *The Journal of Chemical Physics* **2015**, *142*, 154106.

66. Lambrecht, D. S.; Ochsenfeld, C. Multipole-based integral estimates for the rigorous description of distance dependence in two-electron integrals. *The Journal of Chemical Physics* **2005**, *123*, 184101.

67. Almlöf, J.; Truhlar, D. G.; Lybrand, T. P. Supercomputer Chemistry Structure, Dynamics, and Biochemical Applications. *Interdisciplinary Science Reviews* **1990**, *15*, 252–263.

68. Reza Ahmadi, G.; Almlöf, J. The Coulomb operator in a Gaussian product basis. *Chemical Physics Letters* **1995**, *246*, 364–370.

69. Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *Journal of Computational Physics* **1987**, *73*, 325–348.

70. White, C. A.; Johnson, B. G.; Gill, P. M.; Head-Gordon, M. The continuous fast multipole method. *Chemical Physics Letters* **1994**, *230*, 8–16.

71. White, C. A.; Head-Gordon, M. Derivation and efficient implementation of the fast multipole method. *The Journal of Chemical Physics* **1994**, *101*, 6593–6605.

72. Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *Science* **1996**, *271*, 51–53.

73. White, C. A.; Head-Gordon, M. A J matrix engine for density functional theory calculations. *The Journal of Chemical Physics* **1996**, *104*, 2620–2629.

74. Neese, F. An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix. *Journal of Computational Chemistry* **2003**, *24*, 1740–1747.

75. Füsti-Molnár, L.; Pulay, P. The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. *The Journal of Chemical Physics* **2002**, *117*, 7827–7835.

76. Füsti-Molnar, L.; Pulay, P. Accurate molecular integrals and energies using combined plane wave and Gaussian basis sets in molecular electronic structure theory. *The Journal of Chemical Physics* **2002**, *116*, 7795–7805.

77. Füsti-Molnár, L.; Pulay, P. Gaussian-based first-principles calculations on large systems using the Fourier Transform Coulomb method. *Journal of Molecular Structure-theochem* **2003**, *666*, 25–30.

78. Füsti-Molnár, L.; Kong, J. Fast and accurate Coulomb calculation with Gaussian functions. *The Journal of Chemical Physics* **2005**, *122*, 074108.

79. Schwegler, E.; Challacombe, M.; Head-Gordon, M. 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.

80. Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree–Fock-type exchange matrices. *The Journal of Chemical Physics* **1998**, *109*, 1663–1669.

81. Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange. *Chemical Physics* **2009**, *356*, 98–109, Moving Frontiers in Quantum Chemistry:.

82. Laqua, H.; Thompson, T. H.; Kussmann, J.; Ochsenfeld, C. Highly Efficient, Linear Scaling Seminumerical Exact Exchange Method for Graphic Processing Units. *Journal of Chemical Theory and Computation* **2020**, *16*, 1456–1468, PMID: 32053375.

83. Kussmann, J.; Laqua, H.; Ochsenfeld, C. Highly Efficient Resolution-of-Identity Density Functional Theory Calculations on Central and Graphics Processing Units. *Journal of Chemical Theory and Computation* **2021**, *17*, 1512–1521, Publisher: American Chemical Society.

84. Csóka, J.; Kállay, M. Speeding up density fitting Hartree–Fock calculations with multipole approximations. *Molecular Physics* **2020**, *118*, e1769213.

85. Laqua, H.; Kussmann, J.; Ochsenfeld, C. Accelerating seminumerical Fock-exchange calculations using mixed single- and double-precision arithmethic. *The Journal of Chemical Physics* **2021**, *154*, 214116.

86. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369*, 253–287.

87. Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Current Opinion in Structural Biology* **2002**, *12*, 190–196.

88. Lebedev, V. Quadratures on a sphere. *USSR Computational Mathematics and Mathematical Physics* **1976**, *16*, 10–24.

89. Shizgal, B. A Gaussian quadrature procedure for use in the solution of the Boltzmann equation and related problems. *Journal of Computational Physics* **1981**, *41*, 309–328.

90. Riess, R. D.; Johnson, L. W. Estimating Gauss-Chebyshev Quadrature Errors. *SIAM Journal on Numerical Analysis* **1969**, *6*, 557–559.

91. Golub, G. H.; Welsch, J. H. Calculation of Gauss Quadrature Rules. *Mathematics of Computation* **1969**, *23*, 221–s10.

92. Schlömer Nico. Papior, D., Nick. Arnold; Zetter, R. Numerical integration (quadrature, cubature) in Python, QuadPy v0.16.6. **2021**,

93. Toulouse, J. Extension multid´eterminantale de la méthode de Kohn-Sham en théorie de la fonctionnelle de la densité par décomposition de l'interaction électronique en contributions de longue portée et de courte portée. Ph.D. thesis, 2005.

94. Whitten, J. L. Coulombic potential energy integrals and approximations. *The Journal of Chemical Physics* **1973**, *58*, 4496–4501.

95. Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *The Journal of Chemical Physics* **1979**, *71*, 4993.

96. Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of Xα theory. **1979**, *71*, 8.

97. Dunlap, B. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *Journal of Molecular Structure: THEOCHEM* **2000**, *529*, 37–40.

98. Manby, F. R. Density fitting in second-order linear-r12 Møller–Plesset perturbation theory. *The Journal of Chemical Physics* **2003**, *119*, 4607–4613.

99. Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *The Journal of Chemical Physics* **2003**, *118*, 8149–8160.

100. Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear scaling density fitting. *The Journal of Chemical Physics* **2006**, *125*, 194109.

101. Aquilante, F.; Gagliardi, L.; Pedersen, T. B.; Lindh, R. Atomic Cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency. *The Journal of Chemical Physics* **2009**, *130*, 154107.

102. Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear scaling density fitting. *The Journal of Chemical Physics* **2006**, *125*, 194109.

103. Sodt, A.; Head-Gordon, M. Hartree-Fock exchange computed using the atomic resolution of the identity approximation. *The Journal of Chemical Physics* **2008**, *128*, 104106.

104. Polly, R.; *, H.-J. W.; Manby, F. R.; Knowles, P. J. Fast Hartree–Fock theory using local density fitting approximations. *Molecular Physics* **2004**, *102*, 2311–2321.

105. Jensen, F. *Introduction to Computational Chemistry*; John Wiley and Sons Ltd, 2006.

106. A., S.; NS, O. *Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory*; 1996.

107. Kussmann, J.; Beer, M.; Ochsenfeld, C. Linear-scaling self-consistent field methods for large molecules. *WIREs Computational Molecular Science* **2013**, *3*, 614–636.

108. Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K. PySCF: the Python-based simulations of chemistry framework. 2017; `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340`.

109. Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. *Journal of Computational Chemistry* **2015**, *36*, 1664–1671.

110. Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions + Assessment of Common and Reparameterized (*meta* -)GGA Density Functionals. *Journal of Chemical Theory and Computation* **2010**, *6*, 107–126.

111. Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2011**, *7*, 291–309.

112. Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215.

113. Aquilante, F.; Pedersen, T. B.; Lindh, R. 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.

114. Saravanan, C.; Shao, Y.; Baer, R.; Ross, P. N.; Head-Gordon, M. Sparse matrix multiplications for linear scaling electronic structure calculations in an atom-centered basis set using multiatom blocks. *Journal of Computational Chemistry* **2003**, *24*, 618–622.

115. Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

116. Ni, Z.; Wang, Y.; Li, W.; Pulay, P.; Li, S. Analytical Energy Gradients for the Cluster-in-Molecule MP2 Method and Its Application to Geometry Optimizations of Large Systems. *Journal of Chemical Theory and Computation* **2019**, *15*, 3623–3634.

117. Kleier, D. A.; Halgren, T. A.; Hall, J. H.; Lipscomb, W. N. Localized molecular orbitals for polyatomic molecules. I. A comparison of the Edmiston-Ruedenberg and Boys localization methods. *The Journal of Chemical Physics* **1974**, *61*, 3905–3919.

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

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

120. Higham, N. J. Cholesky factorization. *WIREs Computational Statistics* **2009**, *1*, 251–254.

121. Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *Journal of Computational Chemistry* **1989**, *10*, 104–111.

122. Maurer, S. A.; Lambrecht, D. S.; Flaig, D.; Ochsenfeld, C. Distance-dependent Schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. rigorous upper bounds. *The Journal of Chemical Physics* **2012**, *136*, 144107.

Appendices

A Derivatives of V_{lr}

Using the substitution that $s=\frac{1}{2}\mathbf{r}_{12}\cdot\mathbf{r}_{12}$ the long-range potential becomes:

$$V_{lr}=\frac{\mathrm{erf}(\alpha\sqrt{2s})}{\sqrt{2s}}-X_0e^{-2\gamma s} \tag{1}$$

$$\frac{\partial V_{lr}}{\partial s}=\frac{X_0\sqrt{2s}e^{-2\alpha^2 s}-\mathrm{erf}(\alpha\sqrt{2s})}{2s\sqrt{2s}}-(-2\gamma)X_0e^{-2\gamma s} \tag{2}$$

$$\frac{\partial^2 V_{lr}}{\partial s^2}=-\frac{X_0\sqrt{2s}e^{-2\alpha^2 s}(3+4\alpha^2 s)-3\mathrm{erf}(\alpha\sqrt{2s})}{4s^2\sqrt{2s}}-(-2\gamma)^2X_0e^{-2\gamma s} \tag{3}$$

$$\frac{\partial^3 V_{lr}}{\partial s^3}=\frac{X_0\sqrt{2s}e^{-2\alpha^2 s}(15+20\alpha^2 s+16\alpha^4 s^2)-15\mathrm{erf}(\alpha\sqrt{2s})}{8s^3\sqrt{2s}}-(-2\gamma)^3X_0e^{-2\gamma s} \tag{4}$$

The n^{th} derivative of $V_{lr}(s)$ takes on the form:

$$\frac{\partial^n V_{lr}}{\partial s^n}=(-1)^{n+1}\frac{X_0\sqrt{2s}e^{-2\alpha^2 s}\left[\sum_{i=1}^{n}(4\alpha^2 s)^{i-1}\prod_{k=i}^{n-1}(2k+1)\right]-\prod_{k=0}^{n-1}(2k+1)\mathrm{erf}(\alpha\sqrt{2s})}{\sqrt{2s}(2s)^n}$$
$$-(-2\gamma)^nX_0e^{-2\gamma s} \tag{5}$$

Using L'Hospital's rule the limits of $V_{lr}(0)$ can be obtained.

$$\lim_{s\to 0}\frac{\partial^n V}{\partial s^n}=(-1)^nX_0(\frac{(2\alpha^2)^n}{2n+1})-Xo(-1)^{n+1}(2\gamma)^n \tag{6}$$

B The $f^{k,l}$ Terms for the Two-Range Potential

The terms $f^{k,l}$ can be related to the terms $f^{k,0}$. These relationships are shown below.

	l				
k	0	1	2	3	4
0	f^{00}	$-f^{10}$	f^{20}	$-f^{30}$	f^{40}
1	f^{10}	$-2f^{20}$	$3f^{30}$	$-4f^{40}$	
2	f^{20}	$-3f^{30}$	$6f^{40}$		
3	f^{30}	$-4f^{40}$			
4	f^{40}				

Where the $f^{k,0}(\mathbf{R})$ terms are.

$$f^{0,0} = V(\mathbf{R}) \tag{7}$$

$$f^{1,0} = -\frac{\partial V}{\partial s} R^i \tag{8}$$

$$f^{2,0} = \frac{1}{2}[\frac{\partial^2 V}{\partial s^2}(\mathbf{R}) R^i R^j + \delta_{ij} \frac{\partial V}{\partial s}] \tag{9}$$

$$f^{3,0} = -\frac{1}{3!}[\frac{\partial^3 V}{\partial s^3}(\mathbf{R}) R^i R^j R^k + \frac{\partial^2 V}{\partial s^2}(\mathbf{R})(\delta_{ij} R^k + \delta_{ik} R^j + \delta_{jk} R^i)] \tag{10}$$

$$\begin{aligned} f^{4,0} = \frac{1}{3!}[&\frac{\partial^4 V}{\partial s^4}(\mathbf{R}) R^i R^j R^k R^l \\ &+ \frac{\partial^3 V}{\partial s^3}(\mathbf{R})(\delta_{ij} R^k R^l + \delta_{ik} R^j R^l + \delta_{il} R^k R^j + \delta_{kj} R^i R^l + \delta_{kl} R^i R^j + \delta_{lj} R^i R^k) \\ &+ \frac{\partial^2 V}{\partial s^2}(\mathbf{R})(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})] \end{aligned} \tag{11}$$

$$\tag{12}$$