Energy Densities: A Systematic Approach To Correlation In Density Functional Theory

Tom J P Irons

Typset In LaTeX

University of Nottingham
UK | CHINA | MALAYSIA

Energy Densities: A Systematic Approach To Correlation In Density Functional Theory

Submitted March 2019 in partial fulfilment of the conditions for the award of the degree PhD Theoretical & Computational Chemistry

Tom James Patrick Irons

Supervised by Dr. Andrew M. Teale

School Of Chemistry
University Of Nottingham

Preface

Abstract

Density functional theory (DFT) has grown to become by far the most widely applied method in the modelling of electronic systems yet, in contrast to wavefunction–based *ab initio* methods, the reliability of a DFT calculation can be uncertain. This is because the essential ingredient required for a DFT calculation to be meaningful – the exchange & correlation energy functionals, are approximations for a type of electronic interaction with an unknown functional form. Whilst there exist types of system for which DFT does not provide a useful model – those with significant dispersion interactions and those with near–degenerate states, seeking improvements to DFT in these areas can be far from straightforward since the exchange & correlation functionals cannot be systematically improved.

In this work, a mathematically rigorous description of DFT is combined with some of the most reliable & accurate *ab initio* electronic structure methods in a bespoke development code to obtain a detailed picture of how the exact correlation energy functional in DFT behaves. Using these insights, new approaches are investigated for modelling the correlation energy functional in local form, allowing a systematic study of how best to approximate the correlation energy functional in different types of system to be pursued.

As an adjunct to this, the present work looks ahead at how to extend & generalise this investigation by considering the behaviour of systems in the presence of a magnetic field. Efficient algorithms are developed and implemented to facilitate this, enabling the advancement of this strand of investigation in subsequent work.

Publications

(1) Irons, T. J. P. & A. M. Teale, The coupling constant averaged exchange–correlation energy density, *Molecular Physics*, **114**(3-4), 484–497, 2016.

(2) Vuckovic, S., T. J. P. Irons, A. Savin, A. M. Teale & P. Gori-Giorgi, Exchange–correlation functionals via local interpolation along the adiabatic connection, *Journal of Chemical Theory and Computation*, **12**(6), 2598–2610, 2016. [*ACS Editors' Choice*]

(3) Vuckovic, S., T. J. P. Irons, L. O. Wagner, A. M. Teale & P. Gori-Giorgi, Interpolated energy densities, correlation indicators and lower bounds from approximations to the strong coupling limit of DFT, *Physical Chemistry Chemical Physics*, **19**(8), 6169–6183, 2017.

(4) Irons, T. J. P., J. Zemen & A. M. Teale, Efficient calculation of molecular integrals over London atomic orbitals, *Journal of Chemical Theory and Computation*, **13**(8), 3636–3649, 2017.

(5) Irons, T. J. P., J. W. Furness, M. S. Ryley, J. Zemen, T. Helgaker & A. M. Teale, Connections between variation principles at the interface of wave-function and density-functional theories, *The Journal of Chemical Physics*, **147**(13), 134107, 2017. [*AIP Editor's Pick*]

Acknowledgements

These pages serve as a record of my study between September 2014 & October 2018. I am grateful to the School of Chemistry for their choosing to gift me this opportunity, given my unusual circumstances of four years ago. I have enjoyed seeing the scientific frontline, where the boundaries of knowledge are unfolded a little further as each day passes.

I am grateful to the people I have met along the way, from colleagues in Nottingham to collaborators in Amsterdam & Oslo. Above all, I am grateful to my supervisor, Andy Teale: for explaining it to me when I didn't understand, for taking a look when it didn't work but most of all for being patient and understanding, when others might not have been.

Tom Irons
March 8, 2019

Contents

1 Introduction To Quantum Chemistry

1.1 The Molecular Schrödinger Equation

In an electronic structure problem, the fundamental motivation is that of solving the *Schrödinger Equation*[1] for the system being considered, and every approach to predicting the electronic structure of a system is simply a method of trying to solve the molecular Schrödinger Equation. In its most general form, the Schrödinger Equation may be written as the following

$$\hat{\mathcal{H}}\Psi = \mathcal{E}\Psi \tag{1.1}$$

where Ψ is the molecular wavefunction, a mathematical function that describes the system in its entirety, $\mathcal{E}$ the energy corresponding to a system with *state* Ψ and $\hat{\mathcal{H}}$ is the *Hamiltonian operator*. In this general case Ψ is an exact solution, or family of solutions Ψ_i, to the Schrödinger Equation as the action of the Hamiltonian on Ψ_i is equal to $\mathcal{E}_i\Psi_i$; in other words, Ψ_i is an *eigenfunction* of $\hat{\mathcal{H}}$ and $\mathcal{E}_i$ is the *eigenvalue* associated with it.

In practical applications of the Schrödinger Equation, the Hamiltonian is usually expressed in terms of the individual operators that it contains, thus the form of the Hamiltonian varies depending on the problem being considered. For the quantum chemical applications covered in this work, the form of the Hamiltonian considered is most often that of the non–relativistic molecular Hamiltonian, in the context of the *time-independent* Schrödinger Equation. For a system consisting of N nuclei, with atomic numbers Z_α, masses M_α ($\alpha = 1, \ldots, N$) and n electrons, this Hamiltonian has the form

$$\begin{aligned}\hat{\mathcal{H}} = &\underbrace{-\sum_{\alpha=1}^{N}\frac{\hbar^2}{2M_\alpha}\nabla_\alpha^2}_{\text{nuclear kinetic energy}} \quad \underbrace{-\sum_{i=1}^{n}\frac{\hbar^2}{2m_e}\nabla_i^2}_{\text{electron kinetic energy}} \quad \underbrace{-\sum_{i=1}^{n}\sum_{\alpha=1}^{N}\left(\frac{e^2}{4\pi\epsilon_0}\right)\frac{Z_\alpha}{|\mathbf{r}_i-\mathbf{R}_\alpha|}}_{\text{electron-nuclear attraction}} \\ &\underbrace{+\sum_{i>j}^{n}\left(\frac{e^2}{4\pi\epsilon_0}\right)\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}}_{\text{electron-electron repulsion}} \quad \underbrace{+\sum_{\alpha>\beta}^{N}\left(\frac{e^2}{4\pi\epsilon_0}\right)\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}}_{\text{nuclear-nuclear repulsion}}\end{aligned} \tag{1.2}$$

Here m_e denotes the mass of the electron, e the elementary charge, $\mathbf{r}_i$ the position of the i^{th} electron, $\mathbf{R}_\alpha$ the position of the α^{th} nucleus, ϵ_0 the permittivity of free space and ∇_i^2 the Laplacian operator $\left(\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i^2}+\frac{\partial^2}{\partial z_i^2}\right)$ with respect to the coordinates of the i^{th} particle.

It is often convenient in quantum chemistry to define quantities in terms of *atomic units*, rather than using conventional *Système International* (SI) units, allowing many constants to be omitted from expressions and simplifying them considerably. Atomic units defined in Table 1.1 are used throughout this work, unless stated otherwise.

Quantity	Atomic Units	Symbol	Standard Units
mass	electron mass	m_e	9.1094×10^{-31} kg
charge	proton charge	e	1.6022×10^{-19} C
length	Bohr	a_0	5.2918×10^{-11} m
energy	Hartree	E_h	4.3597×10^{-18} J

Table 1.1: Atomic units used throughout this work unless stated otherwise.

In atomic units, the molecular Hamiltonian is considerably simplified and reduces to the following form,

$$\hat{\mathcal{H}} = \underbrace{-\sum_{\alpha=1}^{N}\frac{1}{2M_\alpha}\nabla_\alpha^2}_{\hat{T}_N} \quad \underbrace{-\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2}_{\hat{T}_e} \quad \underbrace{-\sum_{i=1}^{n}\sum_{\alpha=1}^{N}\frac{Z_\alpha}{|\mathbf{r}_i-\mathbf{R}_\alpha|}}_{\hat{V}} \quad \underbrace{+\sum_{i>j}^{n}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}}_{\hat{W}} \quad \underbrace{+\sum_{\alpha>\beta}^{N}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}}_{\hat{V}_{NN}} \tag{1.3}$$

1.1.1 The Born-Oppenheimer Approximation

In the time-independent, non-relativistic molecular Schrödinger Equation, the eigenfunctions of the Hamiltonian in (1.3) will describe the molecular system as a set of the individual quantum particles it contains, hence the wavefunctions $\Psi(\{\mathbf{r}\},\{\mathbf{R}\})$ of the system will depend on the sets of electron coordinates $\{\mathbf{r}\}$ and nuclear coordinates $\{\mathbf{R}\}$.

In many electronic structure calculations however, dependence of the wavefunction on both electronic and nuclear coordinates represents an unnecessary complication; as the mass of a nucleus is far greater than the mass of an electron, it can be reasonably assumed that electrons in a molecule respond *adiabatically* to the motion of the nuclei. In other words, it is assumed that nuclear motion is sufficiently gradual that the surrounding electrons do not move between states in response but remain in the same state, which evolves in response to the nuclear motion. This model is formalised as the *Born-Oppenheimer Approximation*[2], in which the Schrödinger Equation is decoupled and the molecular wavefunction resolved into the product of a nuclear and electronic wavefunction,

$$\Psi(\{\mathbf{r}\},\{\mathbf{R}\}) \approx \Xi(\{\mathbf{R}\})\,\Phi(\{\mathbf{r};\mathbf{R}\}) \tag{1.4}$$

where the electronic wavefunction $\Phi(\{\mathbf{r};\mathbf{R}\})$ is only *parametrically* dependent on the nuclear coordinates and the nuclear wavefunction $\Xi(\{\mathbf{R}\})$ is independent of the electronic coordinates.

This approach can be rationalised by considering the individual components of the Hamiltonian as functions of the electronic and nuclear coordinates,

$$\hat{\mathcal{H}}(\mathbf{r},\mathbf{R}) = \underbrace{\hat{T}_N(\mathbf{R}) + \hat{V}_{NN}(\mathbf{R})}_{\hat{\mathcal{H}}_N(\mathbf{R})} + \underbrace{\hat{T}_e(\mathbf{r}) + \hat{V}(\mathbf{r},\mathbf{R}) + \hat{W}(\mathbf{r})}_{\hat{\mathcal{H}}_e(\mathbf{r},\mathbf{R})}, \tag{1.5}$$

which can be written as the sum of an *electronic Hamiltonian* $\hat{\mathcal{H}}_e$ and a second term $\hat{\mathcal{H}}_N$ comprising the elements of (1.5) acting only on the nuclear coordinates. This does not fully separate the electronic and nuclear terms, as the electron–nuclear attraction operator acts on both the electronic and nuclear coordinates. However, at a fixed molecular geometry, the nuclear coordinates become parameters and $\hat{V}$ can be reduced to a one–electron operator, representing the electrostatic interaction between the electrons and a set of static nuclear *point charges*.

Accordingly, an *electronic* Schrödinger Equation can be constructed from the electronic Hamiltonian at a fixed molecular geometry as

$$\hat{\mathcal{H}}_e(\mathbf{r};\mathbf{R})\Phi_i(\{\mathbf{r};\mathbf{R}\}) = \left[\hat{T}_e(\mathbf{r}) + \hat{V}(\mathbf{r};\mathbf{R}) + \hat{W}(\mathbf{r})\right]\Phi_i(\{\mathbf{r};\mathbf{R}\}) = \mathcal{E}_i(\mathbf{R})\Phi_i(\{\mathbf{r};\mathbf{R}\}) \tag{1.6}$$

where the electronic wavefunction Φ_i is an eigenfunction of the electronic Hamiltonian, with the corresponding eigenvalue $\mathcal{E}_i(\mathbf{R})$ being the electronic energy at the given nuclear geometry. The eigenfunctions Φ_i form a complete orthonormal basis, constituting the *electronic energy levels* of the system for a given set of nuclear coordinates, with ground–state electronic wavefunction Φ_0 and energy $\mathcal{E}_0$.

The full Schrödinger Equation for a system at a given molecular geometry can be similarly constructed, generalising (1.6) to include the terms in $\hat{\mathcal{H}}_N$, the nuclear wavefunction and simplifying using the product rule of differentiation,

$$\begin{aligned}\hat{\mathcal{H}}(\mathbf{r},\mathbf{R})\,\Psi_{ij}(\{\mathbf{r}\},\{\mathbf{R}\}) &\approx \left[\hat{\mathcal{H}}_N(\mathbf{R}) + \mathcal{E}_j(\mathbf{R})\right]\Xi_i(\{\mathbf{R}\})\,\Phi_j(\{\mathbf{r};\mathbf{R}\}) \\ &= E_{ij}\,\Xi_i(\{\mathbf{R}\})\,\Phi_j(\{\mathbf{r};\mathbf{R}\}),\end{aligned} \tag{1.7}$$

where E_{ij} is the total energy of the molecular wavefunction in the i^{th} nuclear and j^{th} electronic eigenstate. In the Born–Oppenheimer approximation, it is assumed that the energetic gap between the different electronic states $\mathcal{E}_i$ is sufficiently large that there is no coupling between the electronic and nuclear wavefunctions. In other words as the nuclear positions change, the total wavefunction of the system will evolve through the *same* electronic state. The electrostatic potential experienced by the nuclei due to the electrons is thus represented by the potential energy function of the given electronic state $\mathcal{E}_i(\mathbf{R})$.

Given a set of solutions to (1.6) defining a potential energy function $\mathcal{E}_i(\mathbf{R})$, the *nuclear* Schrödinger Equation can be written as

$$\left[\hat{\mathcal{H}}_N(\mathbf{R}) + \mathcal{E}_i(\mathbf{R})\right]\Xi_j(\{\mathbf{R}\}) = \left[\hat{T}_N(\mathbf{R}) + \mathcal{U}_i(\mathbf{R})\right]\Xi_j(\{\mathbf{R}\}) = \mathcal{E}_j^N\,\Xi_j(\{\mathbf{R}\}) \tag{1.8}$$

where $\mathcal{E}_j^N$ is the j^{th} nuclear energy eigenvalue and $\mathcal{U}_i(\mathbf{R})$ is the *Born–Oppenheimer potential energy surface* for the i^{th} electronic state,

$$\mathcal{U}_i(\mathbf{R}) = \hat{V}_{NN}(\mathbf{R}) + \mathcal{E}_i(\mathbf{R}), \tag{1.9}$$

with which the nuclei may be considered to interact *classically* rather than quantum mechanically. Therefore, although it does constitute a degree of approximation, the Born-Oppenheimer approximation may be successfully applied to the majority of quantum chemical problems - making attempts to obtain accurate and useful solutions to the electronic Schrödinger Equation much more feasible.

1.2 The Self-Consistent Field Approximation

The central problem that electronic structure calculations aim to solve is that of estimating as accurately as possible the ground state of the *electronic* Schrödinger Equation, given nuclei fixed in position $\{\mathbf{R}\}$, for the general n–electron, N–nuclear system, for which the Hamiltonian is

$$\hat{\mathcal{H}} = -\sum_{i=1}^{n} \frac{1}{2}\nabla_{\mathbf{r}_i}^2 - \sum_{i=1}^{n}\sum_{\alpha=1}^{N} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i>j}^{n} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.10}$$

Due to the final term in (1.10), the electron–electron repulsion operator, the non–relativistic electronic Schrödinger Equation cannot be solved analytically for systems of more than one electron (for which this term vanishes). Therefore, all electronic-structure methods that treat many–electron systems are fundamentally methods for numerically approximating the solution to the electronic Schrödinger Equation as accurately and efficiently as possible.

1.2.1 Many–Electron Wavefunctions

Perhaps the most well–established approach to the electronic structure problem is that of *ab initio* theory, in which the known quantum mechanical behaviour of electrons is used to formulate an *ansatz* for the many–electron wavefunction of a system, assumed to be a fair representation of the true wavefunction, which is then optimised such that it is as close as possible to the true wavefunction of the system.

The initial guess of the many–electron wavefunction may be composed in a number of different ways, however the most common and indeed most obvious choice is to use the exact solution for the one–electron system - i.e. the *hydrogenic orbital*, and construct the many–electron wavefunction as a combination of these *one–electron orbitals*.

1.2.1.1 One–Electron Orbitals

An *orbital* is simply the wavefunction of a single electron, thus encompassing a complete mathematical description of its quantum state. The spatial position $\mathbf{r} \in \mathbb{R}^3$ of an electron can be represented by its *spatial orbital* $\varphi_i(\mathbf{r})$; this is defined as an eigenfunction of the Hamiltonian for one electron,

$$\left(-\frac{1}{2}\nabla_{\mathbf{r}}^2 - \sum_{\alpha} \frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|}\right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}), \tag{1.11}$$

and together the spatial orbitals are generally considered to constitute an orthonormal set,

$$\int_{\mathbb{R}^3} \varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r})\,\mathrm{d}\mathbf{r} = \delta_{ij}. \tag{1.12}$$

The spatial orbital however does not fully characterise the quantum state of an electron as it omits any description of the its *spin*. Although the Hamiltonian operator only acts on the spatial variable $\mathbf{r}$ of an electron, it is known from experimental observations such as the Zeeman effect that electrons possess an intrinsic magnetic moment, which can occupy one of two possible states. This is generally referred to as the spin of the electron and can be described by a spin function $\eta(\sigma)$, where σ is the corresponding spin variable.

Typically the magnetic moment of an electron is quantised by its projection along a particular Cartesian axis, thus the two possible spin functions of an electron are eigenfunctions of the spin angular momentum operator $\hat{s}_z$,

$$\hat{s}_z\alpha(\sigma) = \frac{1}{2}\alpha(\sigma) \qquad\qquad \hat{s}_z\beta(\sigma) = -\frac{1}{2}\beta(\sigma), \tag{1.13}$$

where the spin functions α and β have *spin quantum numbers* of $+1/2$ and $-1/2$ respectively. The two possible values of the spin variable σ form the space $\Sigma := \{\sigma_\uparrow, \sigma_\downarrow\}$, in which the spin functions form an orthonormal basis

$$\int_{\Sigma} \eta^*(\sigma)\tilde{\eta}(\sigma)\,\mathrm{d}\sigma = \delta_{\eta\tilde{\eta}}. \tag{1.14}$$

A complete description of an electron can therefore be obtained by taking the product of its spatial orbital and its spin function; the combined function is usually referred to as a *spinorbital*,

$$\phi_i(\mathbf{r}_i, \sigma_i) = \varphi(\mathbf{r}_i) \times \eta(\sigma_i) \tag{1.15}$$

and is simultaneously an eigenfunction of the one–electron Hamiltonian and spin angular momentum operator.

It is convenient to simplify the notation for spinorbitals by combining the spin and spatial coordinates into a single variable $\boldsymbol{\tau}_i = (\mathbf{r}_i, \sigma_i)$, where $\boldsymbol{\tau}_i \in \mathbb{R}^3_\Sigma := \mathbb{R}^3 \times \Sigma$, for example

$$\boldsymbol{\tau}_i = \begin{cases} (\mathbf{r}_i, \sigma_\uparrow) & \implies \quad \phi_i(\boldsymbol{\tau}_i) = \varphi(\mathbf{r}_i) \times \alpha(\sigma_i) \\ (\mathbf{r}_i, \sigma_\downarrow) & \implies \quad \phi_i(\boldsymbol{\tau}_i) = \varphi(\mathbf{r}_i) \times \beta(\sigma_i) \end{cases} \tag{1.16}$$

As both the spatial and spin functions belong to orthogonal sets, an orthonormal basis can be made from the resulting spinorbitals,

$$\int_{\mathbb{R}^3_\Sigma} \phi_i^*(\boldsymbol{\tau}) \phi_j(\boldsymbol{\tau}) \, \mathrm{d}\boldsymbol{\tau} = \int_{\mathbb{R}^3} \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}) \, \mathrm{d}\mathbf{r} \int_\Sigma \eta^*(\sigma) \tilde{\eta}(\sigma) \, \mathrm{d}\sigma = \delta_{ij} \delta_{\eta\tilde{\eta}}. \tag{1.17}$$

Assuming the basis is complete, a general wavefunction ψ can be expressed as a linear combination of spinorbitals,

$$\psi(\boldsymbol{\tau}) = \sum_i \phi_i(\boldsymbol{\tau}) c_i \qquad\qquad \sum_i c_i^* c_i = 1, \tag{1.18}$$

where c_i is the expansion coefficient of the i^{th} function.

1.2.1.2 Bra–Ket Notation

It is convenient to describe the expansion a wavefunction in a basis of spinorbitals using *bra–ket notation* (otherwise referred to as *Dirac* notation), in which they form an orthonormal set of *ket* vectors $|\cdot\rangle$,

$$\phi_i(\boldsymbol{\tau}) \mapsto |i\rangle \qquad\qquad \phi_i^*(\boldsymbol{\tau}) \mapsto \langle i| \qquad\qquad \langle i|j\rangle = \int_{\mathbb{R}^3_\Sigma} \phi_i^*(\boldsymbol{\tau}) \phi_j(\boldsymbol{\tau}) \, \mathrm{d}\boldsymbol{\tau} = \delta_{ij}, \tag{1.19}$$

where $\langle\cdot|$ denotes a *bra* vector, the complex conjugate of the ket vector, and $\langle\cdot|\cdot\rangle$ is the inner product of the bra and ket functions. In this notation, the linear expansion of (1.18) becomes

$$|\psi\rangle = \sum_i |i\rangle \, c_i \qquad\qquad \sum_i c_i^* c_i = 1 \tag{1.20}$$

and an individual coefficient c_i of the expansion can be determined by projecting $|\psi\rangle$ onto basis function $|i\rangle$,

$$\langle i|\psi\rangle = \int_{\mathbb{R}^3_\Sigma} \phi_i^*(\boldsymbol{\tau}) \psi(\boldsymbol{\tau}) \, \mathrm{d}\boldsymbol{\tau} = \sum_j \int_{\mathbb{R}^3_\Sigma} \phi_i^*(\boldsymbol{\tau}) \phi_j(\boldsymbol{\tau}) \, \mathrm{d}\boldsymbol{\tau} c_j = \sum_j \langle i|j\rangle \, c_j = \sum_j \delta_{ij} c_j = c_i. \tag{1.21}$$

Substituting this result back into (1.20) yields

$$|\psi\rangle = \sum_i |i\rangle \, c_i = \sum_i |i\rangle \, \langle i|\psi\rangle = \left(\sum_i |i\rangle\langle i| \right) |\psi\rangle = |\psi\rangle \,, \tag{1.22}$$

in which the *resolution of the identity* $\sum_i |i\rangle\langle i| = 1$ is used to simplify the expression, hence the assumption that the basis is complete and normalised. The action of a general operator $\hat{\mathcal{O}}$ on a ket, represented by the expansion in (1.20), yields another ket function,

$$\hat{\mathcal{O}} \, |\psi\rangle = |\chi\rangle \,. \tag{1.23}$$

As $|\psi\rangle$ is a linear combination of basis functions, the transformation may be resolved by considering the action of $\hat{\mathcal{O}}$ on individual basis functions; in a complete space, this yields a linear combination of basis functions,

$$\hat{\mathcal{O}} \, |i\rangle = \sum_j |j\rangle \, \mathcal{O}_{ji}, \tag{1.24}$$

where $\mathcal{O}_{ji}$ is defined as a *matrix element* of the operator $\hat{\mathcal{O}}$, otherwise identified as

$$\mathcal{O}_{ji} = \sum_k \delta_{jk} \mathcal{O}_{ki} = \sum_k \langle j|k\rangle \, \mathcal{O}_{ki} = \langle j|\hat{\mathcal{O}}|i\rangle = \int_{\mathbb{R}^3_\Sigma} \phi_j^*(\boldsymbol{\tau}) \hat{\mathcal{O}} \phi_i(\boldsymbol{\tau}) \, \mathrm{d}\boldsymbol{\tau}. \tag{1.25}$$

Thus the transformation in (1.23) can be re–written in terms of individual basis functions and matrix elements,

$$|\chi\rangle = \hat{\mathcal{O}} \, |\psi\rangle = \hat{\mathcal{O}} \sum_i |i\rangle \, c_i = \sum_j \sum_i |j\rangle \, \mathcal{O}_{ji} c_i. \tag{1.26}$$

If $|\psi_i\rangle$ is an eigenfunction of the linear Hermitian operator $\hat{\mathcal{O}}$, then (1.26) can expressed as an eigenvalue equation,

$$\hat{\mathcal{O}}\,|\psi_i\rangle = \varepsilon_i\,|\psi_i\rangle\,, \tag{1.27}$$

where ε_i is the corresponding eigenvalue. Such an operator is described as an *observable*, here sampling the value of some measurement for the eigenfunction $|\psi_i\rangle$. However, for a system in some general state $|\chi\rangle$, the value of the observable cannot be sampled with in this way. Instead the *expectation value* of the observable is computed by statistically averaging the measurement value for each state, weighted by its probability amplitude,

$$\langle\hat{\mathcal{O}}\rangle = \sum_i c_i^* c_i \varepsilon_i = \sum_i \langle\chi|\varepsilon_i\psi_i\rangle\,\langle\psi_i|\chi\rangle = \sum_i \langle\chi|\hat{\mathcal{O}}|\psi_i\rangle\,\langle\psi_i|\chi\rangle = \langle\chi|\hat{\mathcal{O}}|\chi\rangle = \int_{\mathbb{R}^3_\Sigma} \chi^*(\tau)\hat{\mathcal{O}}\chi(\tau)\,\mathrm{d}\tau, \tag{1.28}$$

where c_i is substituted using the result of (1.21), the resulting $\varepsilon_i\,|\psi_i\rangle$ term substituted according to (1.27) and then the resolution of the identity used to eliminate the $\sum_i |\psi_i\rangle\langle\psi_i|$ term.

1.2.1.3 Systems Of Indistinguishable Particles

In a quantum mechanical system comprising more than one electron, the electrons are considered to be *indistinguishable*; they share identical physical properties and cannot be specifically identified by any measurement. The measurable quantities of such a system are the expectation values of the operators representing observables, hence the indistinguishability of electrons requires that the expectation value of the observable be invariant to the interchange of electron coordinates in the wavefunction.

Considering a general n–particle wavefunction $\Psi(\tau_1,\ldots,\tau_i,\ldots,\tau_j,\ldots,\tau_n)$, where τ_i,τ_j are the coordinates of the i^{th} and j^{th} particles respectively, the indistinguishability of the particles requires that for observable $\hat{\mathcal{O}}$,

$$\begin{aligned}\int_{\mathbb{R}^3_\Sigma}\cdots\int_{\mathbb{R}^3_\Sigma} \Psi^*(\tau_1,\ldots,\tau_i,\ldots,\tau_j,\ldots,\tau_n)\hat{\mathcal{O}}\Psi(\tau_1,\ldots,\tau_i,\ldots,\tau_j,\ldots,\tau_n)\,\mathrm{d}\tau_1\ldots\mathrm{d}\tau_n \\ = \int_{\mathbb{R}^3_\Sigma}\cdots\int_{\mathbb{R}^3_\Sigma} \Psi^*(\tau_1,\ldots,\tau_j,\ldots,\tau_i,\ldots,\tau_n)\hat{\mathcal{O}}\Psi(\tau_1,\ldots,\tau_j,\ldots,\tau_i,\ldots,\tau_n)\,\mathrm{d}\tau_1\ldots\mathrm{d}\tau_n,\end{aligned} \tag{1.29}$$

as an example of the interchange of particle coordinates. Any given arrangement of particle coordinates can be converted into any other by a sequence of two–particle coordinate exchanges; a general *permutation operator* $\hat{\mathcal{P}}$ can be expressed as a product of two–particle permutations $\hat{\mathcal{P}}_{ij}$, where $\hat{\mathcal{P}}_{ij}$ acts on the general wavefunction as

$$\hat{\mathcal{P}}_{ij}\Psi(\tau_1,\ldots,\tau_i,\ldots,\tau_j,\ldots,\tau_n) = \Psi(\tau_1,\ldots,\tau_j,\ldots,\tau_i,\ldots,\tau_n). \tag{1.30}$$

It can be trivially deduced that applying the permutation operator twice simply restores the original wavefunction,

$$\hat{\mathcal{P}}_{ij}\hat{\mathcal{P}}_{ij} = \hat{\mathcal{I}} \qquad\Longrightarrow\qquad \hat{\mathcal{P}}_{ij}^{-1} = \hat{\mathcal{P}}_{ij} \tag{1.31}$$

where $\hat{\mathcal{I}}$ is the *identity operator* and $\hat{\mathcal{P}}_{ij}^{-1}$ is the *inverse* of $\hat{\mathcal{P}}_{ij}$. Furthermore, this implies the relation

$$\langle\Psi|\hat{\mathcal{P}}_{ij}\Psi\rangle = \langle\hat{\mathcal{P}}_{ij}^{-1}\Psi|\hat{\mathcal{P}}_{ij}^{-1}\hat{\mathcal{P}}_{ij}\Psi\rangle = \langle\hat{\mathcal{P}}_{ij}^{-1}\Psi|\Psi\rangle\,, \tag{1.32}$$

hence the inverse of $\hat{\mathcal{P}}_{ij}$ is also its *adjoint* $\hat{\mathcal{P}}_{ij}^\dagger$.[i] The results of (1.31) and (1.32) can hence be used to show that the permutation operator is both *Hermitian*, as $\hat{\mathcal{P}}_{ij}^\dagger = \hat{\mathcal{P}}_{ij}$, and *unitary*, as $\hat{\mathcal{P}}_{ij}^\dagger\hat{\mathcal{P}}_{ij} = \hat{\mathcal{I}}$. Using these results, (1.29) can be re–arranged as

$$\langle\Psi|\hat{\mathcal{O}}|\Psi\rangle = \langle\hat{\mathcal{P}}_{ij}\Psi|\hat{\mathcal{O}}|\hat{\mathcal{P}}_{ij}\Psi\rangle = \langle\Psi|\hat{\mathcal{P}}_{ij}^\dagger\hat{\mathcal{O}}\hat{\mathcal{P}}_{ij}|\Psi\rangle\,, \tag{1.33}$$

thus it can be shown that the permutation operator and $\hat{\mathcal{O}}$ satisfy the relation

$$\hat{\mathcal{O}} = \hat{\mathcal{P}}_{ij}^\dagger\hat{\mathcal{O}}\hat{\mathcal{P}}_{ij} \quad\Longrightarrow\quad \hat{\mathcal{P}}_{ij}\hat{\mathcal{O}} = \hat{\mathcal{P}}_{ij}\hat{\mathcal{P}}_{ij}^\dagger\hat{\mathcal{O}}\hat{\mathcal{P}}_{ij} \quad\Longrightarrow\quad \hat{\mathcal{P}}_{ij}\hat{\mathcal{O}} = \hat{\mathcal{O}}\hat{\mathcal{P}}_{ij}. \tag{1.34}$$

The operators $\hat{\mathcal{O}}$ and $\hat{\mathcal{P}}_{ij}$ are thus *commutative*; for any two operators, the *commutator* $[\cdot,\cdot]$ is defined as

$$\left[\hat{\mathcal{O}},\hat{\mathcal{P}}\right] = \hat{\mathcal{O}}\hat{\mathcal{P}} - \hat{\mathcal{P}}\hat{\mathcal{O}}, \tag{1.35}$$

hence is equal to zero for operators that commute. In quantum mechanics, if an operator $\hat{\mathcal{O}}$ has eigenfunction Ψ and also is commutative with operator $\hat{\mathcal{P}}$, then Ψ must *simultaneously*[3] be an eigenfunction of $\hat{\mathcal{O}}$ and $\hat{\mathcal{P}}$. Therefore, if Ψ

[i] For an operator $\hat{\mathcal{O}}$, its adjoint $\hat{\mathcal{O}}^\dagger$ is defined by the relation $\langle\Psi|\hat{\mathcal{O}}\Psi\rangle = \langle\hat{\mathcal{O}}^\dagger\Psi|\Psi\rangle$.

is a many–particle wavefunction and an eigenfunction of the Hamiltonian operator $\hat{\mathcal{H}}$, it must also be an eigenfunction of the permutation operator which, given it is Hermitian, has real eigenvalues. This can be written as the eigenvalue equation

$$\hat{\mathcal{P}}_{ij}\Psi = c_{ij}\Psi, \tag{1.36}$$

where c_{ij} is the eigenvalue corresponding to the permutation of particle coordinates i and j. From (1.31) it is follows that

$$\hat{\mathcal{P}}_{ij}\hat{\mathcal{P}}_{ij}\Psi = c_{ij}\hat{\mathcal{P}}_{ij}\Psi = c_{ij}^2\Psi = \hat{\mathcal{I}}\Psi \quad \Longrightarrow \quad c_{ij}^2 = 1 \quad \Longrightarrow \quad c_{ij} = \pm 1, \tag{1.37}$$

as the eigenvalue of the identity operator is simply unity. It is thus possible to deduce that *any eigenfunction of the Hamiltonian must also be an eigenfunction of the permutation operator, with a corresponding eigenvalue of +1 or -1*;

$$\text{If} \quad \hat{\mathcal{H}}\Psi = E\Psi \quad \text{then} \quad \hat{\mathcal{P}}_{ij}\Psi = \pm\Psi. \tag{1.38}$$

The parity of the wavefunction with respect to particle exchange is a feature of the particles themselves;

$$\begin{aligned} \textbf{Symmetric} \text{ under particle exchange} \quad &\Longrightarrow \quad \hat{\mathcal{P}}_{ij}\Psi = +\Psi \quad \Longrightarrow \quad \text{particles are } \textbf{bosons} \\ \textbf{Antisymmetric} \text{ under particle exchange} \quad &\Longrightarrow \quad \hat{\mathcal{P}}_{ij}\Psi = -\Psi \quad \Longrightarrow \quad \text{particles are } \textbf{fermions} \end{aligned} \tag{1.39}$$

The intrinsic difference between bosons and fermions is characterised by the *spin–statistics theorem*[4], in which:

(i) particles with *integer* spin quantum numbers are bosons and satisfy *Bose–Einstein statistics*[5,6],

(ii) particles with *half–integer* spin quantum numbers are fermions and satisfy *Fermi–Dirac statistics*[7,8].

As described in subsubsection 1.2.1.1, electrons have half–integer spin quantum numbers. Therefore, electrons are a member of the fermion class of quantum particles and it follows that *a many–electron wavefunction must be antisymmetric with respect to particle exchange*.

The requirement of the wavefunction to be antisymmetric with respect to electron exchange manifests itself in the *Pauli exclusion principle*, which states that no two electrons may simultaneously occupy the *same quantum state*, else the wavefunction would be symmetric with respect to electron exchange by virtue of their indistinguishability[9,10].

1.2.1.4 The Hartree Product

The simplest *ansatz* for the wavefunction of an n–electron system may be constructed by assuming that the electrons are non–interacting; the Hamiltonian for an independent electron is simply the *one–electron Hamiltonian*,

$$\hat{h}(i) = -\frac{1}{2}\nabla_{\mathbf{r}_i}^2 - \sum_{\alpha}\frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|}. \tag{1.40}$$

It follows therefore that the Hamiltonian for n independent electrons is given by the sum of their one–electron Hamiltonians $\hat{h}(i)$ and the corresponding eigenvalue by the sum of the eigenvalues for each electron ε_i,

$$\hat{\mathcal{H}}_\Pi = \sum_{i=1}^{n}\hat{h}(i) \qquad\qquad \mathcal{E}_\Pi = \sum_{i=1}^{n}\varepsilon_i. \tag{1.41}$$

It can be readily deduced that the wavefunction for n independent electrons consistent with this is the direct product of the wavefunctions of the individual electrons, in other words their spinorbitals; this is the *Hartree product* wavefunction[11],

$$\Theta(\tau_1, \ldots, \tau_n) = \prod_{i=1}^{n}\phi_i(\tau_i). \tag{1.42}$$

The Hartree product wavefunction is however inherently unphysical and as such is an unsuitable model for a many–electron wavefunction. This is because the electrons are not indistinguishable in the Hartree product, but rather are assigned an individual spinorbital, thus violating the Pauli exclusion principle. This failure can be readily demonstrated with a simple two–electron example,

$$\Theta(\tau_1, \tau_2) = \phi_1(\tau_1)\phi_2(\tau_2) \quad \Longrightarrow \quad \hat{\mathcal{P}}_{12}\Theta(\tau_1, \tau_2) = \phi_1(\tau_2)\phi_2(\tau_1) = \Theta(\tau_2, \tau_1) \neq \pm\Theta(\tau_1, \tau_2), \tag{1.43}$$

where the Hartree product clearly does not satisfy the requirement (1.38).

1.2.1.5 The Slater Determinant

Although the Hartree product is an unphysical representation for a many–electron wavefunction, it's construction as a product of one–electron spinorbitals is a logical means of obtaining the wavefunction for the superposition of the one–particle states. This suggests that a suitable *ansatz* for the wavefunction can be derived, using the Hartree product as a starting point.

Given a complete set of spinorbitals $\{\phi_i\}$, an arbitrary function of the electron coordinates ψ can be represented by an expansion in the basis of spinorbitals; for one electron, this takes the form

$$\psi(\tau_1) = \sum_i c_i \phi_i(\tau_1). \tag{1.44}$$

An analogous two–electron function may be expanded in the same space for a fixed value of the second electron coordinate,

$$\psi(\tau_1, \tau_2) = \sum_i c_i^{\tau_2} \phi_i(\tau_1), \tag{1.45}$$

where $c_i^{\tau_2}$ is the value of the expansion coefficient for $\phi_i(\tau_1)$ at the given τ_2. The set of expansion coefficients for each value of τ_2 may be considered a set of functions of this variable, expanded in the basis of spinorbitals as

$$c_i(\tau_2) = \sum_j d_{ij} \phi_j(\tau_2), \tag{1.46}$$

which can be substituted into (1.45) to yield

$$\psi(\tau_1, \tau_2) = \sum_i \sum_j d_{ij} \phi_i(\tau_1) \phi_j(\tau_2), \tag{1.47}$$

where d_{ij} is the expansion coefficient for the i^{th} and j^{th} spinorbitals in electron coordinates τ_1 and τ_2 respectively. This approach can be extended to an arbitrary number of electrons, hence a many–electron wavefunction can be written as a linear combination of the corresponding Hartree product functions.

As shown in (1.43), there are two different Hartree product functions that can be constructed for a simple two–electron system, $\Theta(\tau_1, \tau_2)$ and $\Theta(\tau_2, \tau_1)$. The corresponding general expansion of the form (1.47) is therefore given by

$$\psi(\tau_1, \tau_2) = d_{11}\phi_1(\tau_1)\phi_1(\tau_2) + d_{12}\phi_1(\tau_1)\phi_2(\tau_2) + d_{21}\phi_2(\tau_1)\phi_1(\tau_2) + d_{22}\phi_2(\tau_1)\phi_2(\tau_2). \tag{1.48}$$

In order to satisfy the Pauli exclusion principle, the coefficients d_{11} and d_{22} must necessarily be zero. This values of the other two coefficients can be deduced by applying the permutation operator to the remaining terms in (1.48),

$$\begin{aligned} \hat{\mathcal{P}}_{12}\psi(\tau_1, \tau_2) &= \hat{\mathcal{P}}_{12}\left[d_{12}\phi_1(\tau_1)\phi_2(\tau_2) + d_{21}\phi_2(\tau_1)\phi_1(\tau_2)\right] \\ &= \left[d_{12}\phi_1(\tau_2)\phi_2(\tau_1) + d_{21}\phi_2(\tau_2)\phi_1(\tau_1)\right] = -\psi(\tau_1, \tau_2), \end{aligned} \tag{1.49}$$

then solving the resulting linear equation, given the antisymmetry requirement

$$d_{12}\left[\phi_1(\tau_2)\phi_2(\tau_1) + \phi_1(\tau_1)\phi_2(\tau_2)\right] = -d_{21}\left[\phi_2(\tau_2)\phi_1(\tau_1) + \phi_2(\tau_1)\phi_1(\tau_2)\right] \quad \Longrightarrow \quad d_{12} = -d_{21}. \tag{1.50}$$

A normalised two–electron wavefunction satisfying the Pauli exclusion principle can therefore be constructed as

$$\psi(\tau_1, \tau_2) = \frac{1}{\sqrt{2}}\left[\phi_1(\tau_1)\phi_2(\tau_2) - \phi_1(\tau_2)\phi_2(\tau_1)\right]. \tag{1.51}$$

This may be generalised beyond the two–electron case such that a normalised wavefunction of an arbitrary n–electron system may be constructed from hydrogenic one–electron spinorbitals and will always satisfy the Pauli exclusion principle; this generalised form is known as the *Slater determinant* wavefunction[12],

$$\Phi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\tau_1) & \phi_1(\tau_2) & \dots & \phi_1(\tau_n) \\ \phi_2(\tau_1) & \phi_2(\tau_2) & \dots & \phi_2(\tau_n) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_n(\tau_1) & \phi_n(\tau_2) & \dots & \phi_n(\tau_n) \end{vmatrix} \tag{1.52}$$

The Slater determinant wavefunction may be constructed from a Hartree product by taking the sum of all possible permutations of electron coordinates, each multiplied by a coefficient of $+1$ or -1 depending on whether the given

arrangement is obtained by an applying an even or odd number of individual two–particle permutations to the initial state respectively. This may be written as

$$\Phi(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) = \hat{\mathcal{A}}\Theta(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n), \tag{1.53}$$

where $\hat{\mathcal{A}}$ is the *antisymmetrisation operator*,

$$\hat{\mathcal{A}} = \frac{1}{\sqrt{n!}} \sum_{u=1}^{n!} (-1)^{p_u} \hat{\mathcal{P}}_u, \tag{1.54}$$

with u representing each permutation of the n electrons and p_u being the number of individual two–electron coordinate exchanges required to construct to yield this from the initial state. This operation can be written as a series expansion of terms each containing all possible m–coordinate transpositions, $1 \leqslant m \leqslant n$, as

$$\sum_{u=1}^{n!} (-1)^{p_u} \hat{\mathcal{P}}_u = \hat{\mathcal{I}} - \sum_{ij}^{n} \hat{\mathcal{P}}_{ij} + \sum_{ijk}^{n} \hat{\mathcal{P}}_{ijk} - \ldots, \tag{1.55}$$

where any m–coordinate transposition for $m \geqslant 3$ can be reduced to the product of up to $m-1$ two–index permutations, for example $\hat{\mathcal{P}}_{231} = \hat{\mathcal{P}}_{23}\hat{\mathcal{P}}_{12}$. Evidently there will be $n!$ possible permutations of electron coordinates among the spinorbitals, thus the antisymmetrisation operator is multiplied by a prefactor of $1/\sqrt{n!}$ to ensure the Slater determinant is normalised.

The antisymmetrisation operator is, up to a multiplicative factor, a *projection operator* that extracts from a product of spinorbitals only the component that satisfies the antisymmetry principle. Projection operators are characterised by the properties of *idempotency* and *Hermiticity*; the antisymmetrisation operator can be shown to demonstrate both of these properties:

(i) *Idempotency*[ii] up to a multiplicative factor is deduced by expanding the square of $\hat{\mathcal{A}}$,

$$\begin{aligned} \hat{\mathcal{A}}\hat{\mathcal{A}} &= \left(\frac{1}{\sqrt{n!}}\right)^2 \sum_{u=1}^{n!} (-1)^{p_u} \hat{\mathcal{P}}_u \sum_{v=1}^{n!} (-1)^{p_v} \hat{\mathcal{P}}_v = \frac{1}{n!} \sum_{u=1}^{n!} \sum_{v=1}^{n!} (-1)^{p_u + p_v} \hat{\mathcal{P}}_u \hat{\mathcal{P}}_v \\ &= \frac{1}{n!} \sum_{u=1}^{n!} \sum_{w=1}^{n!} (-1)^{p_w} \hat{\mathcal{P}}_w = \frac{n!}{n!} \sum_{w=1}^{n!} (-1)^{p_w} \hat{\mathcal{P}}_w = \sqrt{n!}\hat{\mathcal{A}} \end{aligned} \tag{1.56}$$

where it is recognised that the product of two permutation operators is another permutation operator of the same set, thus the product the two antisymmetrisation operators is equivalent to a summation of $n!$ identical operators over the same set.

(ii) *Hermiticity*[iii] can be shown using the generic example for an n–electron Hartree product Θ,

$$\left\langle \hat{\mathcal{A}}\Theta \middle| \Theta \right\rangle = \frac{1}{\sqrt{n!}} \sum_{u=1}^{n!} (-1)^{p_u} \left\langle \hat{\mathcal{P}}_u\Theta \middle| \Theta \right\rangle, \tag{1.57}$$

by multiplying both sides of the inner product by the inverse of the permutation operator and simplifying,

$$\frac{1}{\sqrt{n!}} \sum_{u=1}^{n!} (-1)^{p_u} \left\langle \hat{\mathcal{P}}_u^{-1}\hat{\mathcal{P}}_u\Theta \middle| \hat{\mathcal{P}}_u^{-1}\Theta \right\rangle = \frac{1}{\sqrt{n!}} \sum_{u=1}^{n!} (-1)^{p_u} \left\langle \Theta \middle| \hat{\mathcal{P}}_u^{-1}\Theta \right\rangle \tag{1.58}$$

using the identities in (1.29) and (1.31). As the inverse of the permutation operator is equal to itself, it evidently operates on the same set and has the same parity, thus allowing (1.58) to be written as

$$\frac{1}{\sqrt{n!}} \sum_{u=1}^{n!} (-1)^{p_u} \left\langle \Theta \middle| \hat{\mathcal{P}}_u^{-1}\Theta \right\rangle = \frac{1}{\sqrt{n!}} \sum_{v=1}^{n!} (-1)^{p_v} \left\langle \Theta \middle| \hat{\mathcal{P}}_v\Theta \right\rangle = \left\langle \Theta \middle| \hat{\mathcal{A}}\Theta \right\rangle \tag{1.59}$$

for which comparison with (1.57) yields the identity $\langle \hat{\mathcal{A}}\Theta | \Theta \rangle = \langle \Theta | \hat{\mathcal{A}}\Theta \rangle$, hence the antisymmetrisation operator is Hermitian.

[ii] An operator $\hat{\mathcal{O}}$ is idempotent if the square of the operator is equal to the operator itself, i.e. if $\hat{\mathcal{O}}\hat{\mathcal{O}} \equiv \hat{\mathcal{O}}$.
[iii] An operator $\hat{\mathcal{O}}$ is Hermitian if it is equal to its Hermitian adjoint $\hat{\mathcal{O}}^\dagger$, i.e. $\langle \Theta | \hat{\mathcal{O}}\Theta \rangle = \langle \hat{\mathcal{O}}\Theta | \Theta \rangle$ hence $\hat{\mathcal{O}}^\dagger \equiv \hat{\mathcal{O}}$.

1.2.1.6 Expectation Value Of The Hamiltonian

The ground–state energy of a system $\mathcal{E}$ described by a Slater determinant wavefunction Φ can be written, using the definition in (1.28), as the expectation value of the Hamiltonian operator on Φ,

$$\mathcal{E}(\Phi) = \langle\Phi|\hat{\mathcal{H}}|\Phi\rangle \tag{1.60}$$

Using the form of the Slater determinant in (1.53) and the properties of the antisymmetrisation operator, this expectation value can be re–arranged in terms of the corresponding n–electron Hartree product wavefunction as

$$\left\langle\Phi\middle|\hat{\mathcal{H}}\middle|\Phi\right\rangle = \left\langle\hat{\mathcal{A}}\Theta\middle|\hat{\mathcal{H}}\middle|\hat{\mathcal{A}}\Theta\right\rangle = \left\langle\Theta\middle|\hat{\mathcal{H}}\middle|\hat{\mathcal{A}}\hat{\mathcal{A}}\Theta\right\rangle = \left\langle\Theta\middle|\hat{\mathcal{H}}\middle|\sqrt{n!}\hat{\mathcal{A}}\Theta\right\rangle = \left\langle\Theta\middle|\hat{\mathcal{H}}\middle|\textstyle\sum_u(-1)^{p_u}\hat{\mathcal{P}}_u\Theta\right\rangle, \tag{1.61}$$

where the properties of $\hat{\mathcal{A}}$ outlined in the preceding discussion are used to simplify the expression.

1.2.2 The Variation Principle

The Slater determinant constitutes an appropriate *ansatz* for the many–electron wavefunction as it is constructed from a basis of one–electron orbitals and satisfies the Pauli exclusion principle. However, it does not represent an exact wavefunction for a many–electron system as it implies that the electrons are independent of each other and only experience an average interaction representing the effects of the other electrons. This is a form of *mean field approximation*, where interactions between particles are considered as interactions between a particle and the average of the others.

Given that the Slater determinant is not an exact representation of a physical wavefunction, it follows that the expectation value of the Hamiltonian with a Slater determinant will not yield the exact ground–state energy of the physical system. The expectation value is related to the exact ground–state energy through the *variation principle*; assuming the eigenfunctions of $\hat{\mathcal{H}}$ constitute a complete set $\{|\Psi_i\rangle\}$, a normalised trial wavefunction $|\Phi\rangle$ satisfying the same boundary conditions as $\{|\Psi_i\rangle\}$ can be expanded in this basis and the expectation value of $\hat{\mathcal{H}}$ on $|\Phi\rangle$ will be an upper bound to the exact ground–state energy $\mathcal{E}_0$.

The variation principle can be demonstrated using methods from subsubsection 1.2.1.2; a trial function satisfying the requisite conditions can be expanded in the eigenfunctions of $\hat{\mathcal{H}}$ as

$$|\Phi\rangle = \sum_i |\Psi_i\rangle\, c_i = \sum_i |\Psi_i\rangle\,\langle\Psi_i|\Phi\rangle \qquad\qquad \langle\Phi| = \sum_i c_i^*\,\langle\Psi_i| = \sum_i \langle\Phi|\Psi_i\rangle\,\langle\Psi_i|, \tag{1.62}$$

hence the normalisation condition for the trial function can be written as

$$\begin{aligned}\langle\Phi|\Phi\rangle &= \sum_{ij}\langle\Phi|\Psi_i\rangle\,\langle\Psi_i|\Psi_j\rangle\,\langle\Psi_j|\Phi\rangle = \sum_{ij}\langle\Phi|\Psi_i\rangle\,\delta_{ij}\,\langle\Psi_j|\Phi\rangle \\ &= \sum_i \langle\Phi|\Psi_i\rangle\,\langle\Psi_i|\Phi\rangle = \sum_i |\langle\Phi|\Psi_i\rangle|^2 = 1.\end{aligned} \tag{1.63}$$

The expectation value of the Hamiltonian on $|\Phi\rangle$ can be expanded in a similar way to yield

$$\begin{aligned}\langle\Phi|\hat{\mathcal{H}}|\Phi\rangle &= \sum_{ij}\langle\Phi|\Psi_i\rangle\,\langle\Psi_i|\hat{\mathcal{H}}|\Psi_j\rangle\,\langle\Psi_j|\Phi\rangle = \sum_{ij}\langle\Phi|\Psi_i\rangle\,\delta_{ij}\mathcal{E}_j\,\langle\Psi_j|\Phi\rangle \\ &= \sum_i \mathcal{E}_i\,\langle\Phi|\Psi_i\rangle\,\langle\Psi_i|\Phi\rangle = \sum_i \mathcal{E}_i\,|\langle\Phi|\Psi_i\rangle|^2.\end{aligned} \tag{1.64}$$

Evidently the variation principle must also require $\mathcal{E}_i \geqslant \mathcal{E}_0$, as $\langle\Phi|\hat{\mathcal{H}}|\Phi\rangle$ is a sum of $\mathcal{E}_i$ multiplied by nonnegative coefficients $|\langle\Phi|\Psi_i\rangle|^2$. It therefore follows from substituting the results of (1.63) into (1.64) that

$$\langle\Phi|\hat{\mathcal{H}}|\Phi\rangle = \sum_i \mathcal{E}_i\,|\langle\Phi|\Psi_i\rangle|^2 \geqslant \sum_i \mathcal{E}_0\,|\langle\Phi|\Psi_i\rangle|^2 = \mathcal{E}_0\sum_i |\langle\Phi|\Psi_i\rangle|^2 = \mathcal{E}_0 \tag{1.65}$$

hence the general expression of the variation principle can be written as

$$\mathcal{E}(\Phi) \geqslant \mathcal{E}_0 \tag{1.66}$$

A feature of the variation principle is that a different trial wavefunction yielding a lower expectation value for the Hamiltonian can be identified as a more accurate representation of the physical system than the original trial wavefunction. Given a model wavefunction expanded in the eigenfunctions of the Hamiltonian according to (1.62), the model may be made more accurate by adjusting the coefficients of the basis such that the expectation value of the Hamiltonian is lower. In other words, the wavefunction may be optimised by minimising the expectation value of the Hamiltonian with respect to changes in the expansion coefficients.

1.2.3 Hartree–Fock Theory

The simplest method of approximating the wavefunction of a many–electron system using the orbital model is the *Hartree–Fock* (HF) method. This arose as a combination of theories developed by Hartree for modelling the many–electron wavefunction[11,13] and those developed by Fock for the antisymmetric description of the wavefunction to correctly account for the Pauli exclusion principle[14]. Although HF theory was not the product of a collaboration, it is so named because it was only after learning of Fock's work that Hartree altered his model to correctly include the antisymmetry effects, thus formalising the Hartree–Fock method.

The main principle underlying the HF approximation is that an accurate representation of the many–electron wavefunction for a system may be obtained by constructing a Slater determinant for the system and optimising it according to the variation principle. By limiting the model to a *single* Slater determinant, the quantum state of each electron is assumed to be fixed, rather than varying according to the instantaneous interaction with each of the other electrons. As such, electrons are treated as being independent, but each experiencing an *effective potential* by virtue of the other electrons in the system, reducing the many–electron problem to a one–electron problem; a mean–field approximation.

The expectation value of the Hamiltonian is given by (1.60) however, to evaluate this expression, it is necessary to resolve the Hamiltonian into its one–electron and two–electron components as

$$\hat{\mathcal{H}} = \sum_i \left(-\frac{1}{2}\nabla^2_{\mathbf{r}_i} - \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \right) + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} = \sum_i \hat{h}(i) + \sum_{i>j} \hat{w}(i,j) \tag{1.67}$$

where $\hat{h}(i)$ is the one–electron Hamiltonian, as given in (1.40), and $\hat{w}(i,j)$ is the two–electron Hamiltonian. The evaluation of the expectation value for each of these terms are described in turn.

1.2.3.1 One–Electron Hamiltonian

The one–electron Hamiltonian contains the kinetic energy and nuclear interaction terms, hence operates on the coordinates of each electron independently. An expression for the expectation value of a one–electron operator with a Slater determinant can be derived by substituting the one–electron operator into the result of (1.61), yielding

$$\left\langle \Phi \middle| \hat{h}(i) \middle| \Phi \right\rangle = \left\langle \phi_1(\tau_1)\phi_2(\tau_2)\dots\phi_n(\tau_n) \middle| \hat{h}(i) \middle| \textstyle\sum_u (-1)^{p_u} \hat{\mathcal{P}}_u \phi_1(\tau_1)\phi_2(\tau_2)\dots\phi_n(\tau_n) \right\rangle . \tag{1.68}$$

Although initially appearing to complicate the expression, (1.68) allows the action of the one–electron Hamiltonian to be considered individually for each permutation of spinorbitals, allowing the following to be deduced:

(i) In the original configuration, there is no permutation hence $\hat{\mathcal{P}}_u = \hat{\mathcal{I}}$ and the corresponding term in (1.68) is resolved as

$$\langle \phi_1(\tau_1)|\phi_1(\tau_1)\rangle \langle \phi_2(\tau_2)|\phi_2(\tau_2)\rangle \dots \langle \phi_i(\tau_i)|\hat{h}(i)|\phi_i(\tau_i)\rangle \dots \langle \phi_n(\tau_n)|\phi_n(\tau_n)\rangle = h_{ii} \tag{1.69}$$

(ii) For a permutation $\hat{\mathcal{P}}_u = \hat{\mathcal{P}}_{ij}$ comprising the exchange of electron coordinates between the i^{th} and j^{th} electrons, the corresponding term in (1.68) is resolved as

$$-\langle \phi_1(\tau_1)|\phi_1(\tau_1)\rangle \langle \phi_2(\tau_2)|\phi_2(\tau_2)\rangle \dots \langle \phi_i(\tau_i)|\hat{h}(i)|\phi_j(\tau_i)\rangle \langle \phi_j(\tau_j)|\phi_i(\tau_j)\rangle \dots \langle \phi_n(\tau_n)|\phi_n(\tau_n)\rangle = 0 \tag{1.70}$$

This outcome is a direct consequence of the orthonormality of the spinorbitals as shown in subsubsection 1.2.1.1, because of which all terms of the type $\langle \phi_j|\phi_i\rangle$ are necessarily zero. The same result is obtained for all other permutations, with the exception of the identity permutation, hence the resulting product expansion of matrix elements always contains zero–valued terms.

Therefore, the expectation value of a one–electron operator is simply h_{ii}, which can be summed over all electrons to give the expectation value of the one–electron Hamiltonian as,

$$\langle \Phi | \sum_i \hat{h}(i) | \Phi \rangle = \sum_i \langle \phi_i(\tau_i)|\hat{h}(i)|\phi_i(\tau_i)\rangle = \sum_i \langle i|\hat{h}|i\rangle = \sum_i h_{ii} \tag{1.71}$$

1.2.3.2 Two–Electron Hamiltonian

The two–electron term in the Hamiltonian represents the *electron–electron repulsion* interaction. In a similar way as for the one–electron operator, its expectation value for a Slater determinant can be simplified using the result of of (1.61),

$$\langle\Phi|\hat{w}(i,j)|\Phi\rangle = \left\langle \phi_1(\tau_1)\phi_2(\tau_2)\dots\phi_n(\tau_n)\middle|\hat{w}(i,j)\middle|\sum_u(-1)^{p_u}\hat{\mathcal{P}}_u\phi_1(\tau_1)\phi_2(\tau_2)\dots\phi_n(\tau_n)\right\rangle . \tag{1.72}$$

However, unlike the one–electron operator for which only matrix elements of two spinorbitals with the same index as the operator are significant, the two–electron operator acts on two electronic coordinates simultaneously thus will have non–vanishing matrix elements between pairs of spinorbitals with the relevant indices. As for the one–electron operator, this expectation value can be simplified by considering individual permutations of the spinorbitals in turn:

(i) In the original configuration, there is no permutation hence $\hat{\mathcal{P}}_u$ is the identity operator and the corresponding term in (1.72) is given by

$$\langle\phi_1(\tau_1)|\phi_1(\tau_1)\rangle\langle\phi_2(\tau_2)|\phi_2(\tau_2)\rangle\dots\langle\phi_i(\tau_i)\phi_j(\tau_j)|\hat{w}(i,j)|\phi_i(\tau_i)\phi_j(\tau_j)\rangle\dots\langle\phi_n(\tau_n)|\phi_n(\tau_n)\rangle = J_{ij} \tag{1.73}$$

(ii) For a single permutation of electron coordinates $\hat{\mathcal{P}}_u = \hat{\mathcal{P}}_{ij}$, unlike for the one–electron operator, the corresponding term in (1.72) does not vanish but takes the form

$$-\langle\phi_1(\tau_1)|\phi_1(\tau_1)\rangle\langle\phi_2(\tau_2)|\phi_2(\tau_2)\rangle\dots\langle\phi_i(\tau_i)\phi_j(\tau_j)|\hat{w}(i,j)|\phi_j(\tau_i)\phi_i(\tau_j)\rangle\dots\langle\phi_n(\tau_n)|\phi_n(\tau_n)\rangle = -K_{ij} \tag{1.74}$$

(iii) For other permutations, the resulting product of terms would contain those of the type $\langle\phi_j|\phi_i\rangle$; necessarily zero due to the orthogonality of the spinorbitals.

The terms J_{ij} and K_{ij} are the *Coulomb* and *exchange* matrix elements respectively. The Coulomb term can be interpreted physically as the classical electrostatic interaction between the charge distributions of two electrons. The exchange term is less easily characterised, but may be described as the energetic contribution arising from electrons of parallel spin states having different spatial wavefunctions.

For simplicity of notation, the matrix elements of the two–electron operator are often written as

$$\begin{aligned} \textbf{Coulomb:} \quad & \langle\phi_i(\tau_i)\phi_j(\tau_j)|\hat{w}(i,j)|\phi_i(\tau_i)\phi_j(\tau_j)\rangle = \langle ij|ij\rangle = (ii|jj) = J_{ij} \\ \textbf{Exchange:} \quad & \langle\phi_i(\tau_i)\phi_j(\tau_j)|\hat{w}(i,j)|\phi_j(\tau_i)\phi_i(\tau_j)\rangle = \langle ij|ji\rangle = (ij|ji) = K_{ij} \end{aligned} \tag{1.75}$$

where the notation $\langle ij|ij\rangle$ is generally referred to as *Physicist's notation*, where there is a coordinate from each of the two electrons on both sides of the operator, whilst $(ii|jj)$ is *Mulliken notation*, in which both coordinates from one electron appear on one side of the operator whilst both coordinates from the other electron are collected on the other side of the operator. The expectation of the two–electron Hamiltonian on a Slater determinant is given concisely as

$$\langle\Phi|\sum_{i>j}\hat{w}(i,j)|\Phi\rangle = \frac{1}{2}\sum_{ij}\left(\langle ij|ij\rangle - \langle ij|ji\rangle\right) = \frac{1}{2}\sum_{ij}\left(J_{ij} - K_{ij}\right) = \frac{1}{2}\sum_{ij}\langle ij||ij\rangle \tag{1.76}$$

where the Coulomb and exchange terms are combined into an single antisymmetrised matrix element,

$$\langle ij||ij\rangle = \langle ij|ij\rangle - \langle ij|ji\rangle . \tag{1.77}$$

Combining the one and two–electron terms, the expectation value of the Hamiltonian (1.10) for a Slater determinant wavefunction can now be presented in the simplified form

$$\langle\Phi|\hat{\mathcal{H}}|\Phi\rangle = \sum_i\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{ij}\langle ij||ij\rangle . \tag{1.78}$$

It is convenient to define each of these resulting components in terms of their own operators acting on a single electron coordinate. The one–electron operator is clearly defined in these terms, shown in (1.71); similarly the Coulomb and exchange operators can be defined as

$$\begin{aligned} \textbf{Coulomb:} \quad & \hat{\mathcal{J}}_q(i)\phi_p(\tau_i) = \left[\int\phi_q^*(\tau_j)\hat{w}(i,j)\phi_q(\tau_j)\mathrm{d}\tau_j\right]\phi_p(\tau_i) \\ \textbf{Exchange:} \quad & \hat{\mathcal{K}}_q(i)\phi_p(\tau_i) = \left[\int\phi_q^*(\tau_j)\hat{w}(i,j)\phi_p(\tau_j)\mathrm{d}\tau_j\right]\phi_q(\tau_i) \end{aligned} \tag{1.79}$$

with the result that the Hamiltonian can be written as a series of one–electron operators, each being a *Fock operator* and acting on a single electron coordinate

$$\hat{f}(i) = \hat{h}(i) + \sum_q\left(\hat{\mathcal{J}}_q(i) - \hat{\mathcal{K}}_q(i)\right) \tag{1.80}$$

1.2.3.3 The Hartree–Fock Equations

The HF equations can be derived in spinorbital form as the minimisation of the expectation value of the Hamiltonian on a single Slater determinant, yielding a set of coupled integro–differential equations; the Hartree–Fock eigenvalue equations.

Given a trial function Φ, the expectation value can be evaluated as a functional of Φ according to (1.60). If a small perturbation is applied to the function, i.e. $\Phi \rightarrow \Phi + \delta\Phi$, the energy is perturbed as

$$\mathcal{E}(\Phi + \delta\Phi) = \langle \Phi + \delta\Phi | \hat{\mathcal{H}} | \Phi + \delta\Phi \rangle = \mathcal{E}(\Phi) + \left\{ \langle \delta\Phi | \hat{\mathcal{H}} | \Phi \rangle + \langle \Phi | \hat{\mathcal{H}} | \delta\Phi \rangle \right\} + \ldots = \mathcal{E}(\Phi) + \delta\mathcal{E} + \ldots, \tag{1.81}$$

where $\delta\mathcal{E}$ is the first–order variation in the energy, including terms linear in $\delta\Phi$. According to the variation principle, the optimised wavefunction Φ is that which minimises $\mathcal{E}(\Phi)$, i.e. for which $\delta\mathcal{E} = 0$, which ensures that $\mathcal{E}(\Phi)$ is stationary with respect to variation in Φ.

Given a Slater determinant is a linear expansion in a basis of spinorbitals $\{\phi_p\}$, the problem can be cast as the minimisation of the corresponding expectation value with respect to the spinorbitals, subject to the constraint that they remain orthonormal, i.e. that $\langle p|q\rangle - \delta_{pq} = 0$. This constraint can be included in the variational problem using *Lagrange's method of undetermined multipliers*, in which the Lagrangian functional to be minimised is defined as

$$\begin{aligned} \mathcal{L}(\{\phi_p\}) &= \mathcal{E}(\{\phi_p\}) - \sum_{pq} \varepsilon_{qp} \left(\langle p|q\rangle - \delta_{pq} \right) \\ &= \sum_{p} \langle p|\hat{h}|p\rangle + \frac{1}{2} \sum_{pq} \langle pq||pq\rangle - \sum_{pq} \varepsilon_{qp} \left(\langle p|q\rangle - \delta_{pq} \right) \end{aligned} \tag{1.82}$$

where the $\{\varepsilon_{qp}\}$ are a set of undetermined Lagrange multipliers, which must be Hermitian given that $\mathcal{L}$ is real and $\langle p|q\rangle = \langle q|p\rangle^*$. Applying a small perturbation to the spinorbitals, i.e. $\phi_p \rightarrow \phi_p + \delta\phi_p$, and setting the first order variation in the Lagrangian equal to zero yields the equation

$$\delta\mathcal{L} = \delta\mathcal{E} - \sum_{pq} \varepsilon_{qp} \delta \langle p|q\rangle = 0 \tag{1.83}$$

to be solved. Using the chain rule of differentiation, this is expanded in terms of spinorbitals as

$$\begin{aligned} \delta\mathcal{L} = &\sum_{p} \Big(\langle \delta\phi_p | \hat{h} | \phi_p \rangle + \langle \phi_p | \hat{h} | \delta\phi_p \rangle \Big) \\ &+ \frac{1}{2} \sum_{pq} \left(\langle \delta\phi_p \phi_q | \phi_p \phi_q \rangle + \langle \phi_p \delta\phi_q | \phi_p \phi_q \rangle + \langle \phi_p \phi_q | \delta\phi_p \phi_q \rangle + \langle \phi_p \phi_q | \phi_p \delta\phi_q \rangle \right) \\ &- \frac{1}{2} \sum_{pq} \left(\langle \delta\phi_p \phi_q | \phi_q \phi_p \rangle + \langle \phi_p \delta\phi_q | \phi_q \phi_p \rangle + \langle \phi_p \phi_q | \delta\phi_q \phi_p \rangle + \langle \phi_p \phi_q | \phi_q \delta\phi_p \rangle \right) \\ &- \sum_{pq} \varepsilon_{qp} \left(\langle \delta\phi_p | \phi_q \rangle + \langle \phi_p | \delta\phi_q \rangle \right) = 0 \end{aligned} \tag{1.84}$$

which can be simplified using the permutational symmetries of the matrix elements to give

$$\delta\mathcal{L} = \sum_{p} \langle \delta\phi_p | \hat{h} | \phi_q \rangle + \sum_{pq} \left(\langle \delta\phi_p \phi_q | \phi_p \phi_q \rangle - \langle \delta\phi_p \phi_q | \phi_q \phi_p \rangle \right) - \sum_{pq} \varepsilon_{qp} \langle \delta\phi_p | \phi_q \rangle + \text{c.c.} = 0 \tag{1.85}$$

where c.c. denotes the corresponding complex conjugate expression. This may be re–written as an integral over the one–electron, Coulomb and exchange operators as

$$\delta\mathcal{L} = \sum_{p} \int \delta\phi_p^*(\boldsymbol{\tau}_i) \left[\hat{h}(i)\phi_p(\boldsymbol{\tau}_i) + \sum_{q} \left(\hat{\mathcal{J}}_q(i) - \hat{\mathcal{K}}_q(i) \right) \phi_p(\boldsymbol{\tau}_i) - \sum_{q} \varepsilon_{qp} \phi_q(\boldsymbol{\tau}_i) \right] \mathrm{d}\boldsymbol{\tau}_i + \text{c.c.} = 0, \tag{1.86}$$

for a given electron i. Since the perturbation $\delta\phi_p^*$ is arbitrary in value, for this expression to be always true the bracketed term must necessarily be zero for all spinorbitals, from which it follows that

$$\left[\hat{h}(i) + \sum_{q} \left(\hat{\mathcal{J}}_q(i) - \hat{\mathcal{K}}_q(i) \right) \right] \phi_p(\boldsymbol{\tau}_i) = \sum_{q} \varepsilon_{qp} \phi_q(\boldsymbol{\tau}_i) \tag{1.87}$$

which is a one–electron eigenvalue problem, more succinctly written in terms of the Fock operator as

$$\hat{f}\,|\phi_p\rangle = \sum_q \varepsilon_{qp}\,|\phi_q\rangle\,. \tag{1.88}$$

This form of the HF eigenvalue equations however can be further simplified by application of a unitary transform to the spinorbitals; the transformation of a set of spinorbitals $\{\phi_p\}$ by a unitary matrix $\mathbf{U}$, for which $\mathbf{U}^\dagger = \mathbf{U}^{-1}$, into another set $\{\tilde{\phi}_p\}$,

$$\tilde{\phi}_p = \sum_q \phi_q U_{qp}, \tag{1.89}$$

whilst preserving orthonormality. For a matrix of spinorbitals $\mathbf{O}$, the normalised Slater determinant can be written as $\Phi = (n!)^{-1/2}\det\{\mathbf{O}\}$, hence the Slater determinant of transformed spinorbitals can be expressed as

$$\tilde{\Phi} = \frac{1}{\sqrt{n!}}\det\{\mathbf{UO}\} = \frac{1}{\sqrt{n!}}\det\{\mathbf{U}\}\det\{\mathbf{O}\} = \det\{\mathbf{U}\}\Phi. \tag{1.90}$$

Given that $\mathbf{U}$ is unitary, it can be shown that its determinant is simply a phase factor $\exp(\imath\omega)$.[iv] It follows that the expectation values of observables are unaffected by a unitary transformation of the spinorbitals in a Slater determinant.

The unitary transformation of the spinorbitals can be used to derive a canonical form of the HF eigenvalue equations. Left–multiplying (1.88) by $\langle\phi_t|$ reveals that the Lagrange multipliers are in fact matrix elements of the Fock operator,

$$\langle\phi_t|\hat{f}|\phi_p\rangle = \sum_q \varepsilon_{qp}\,\langle\phi_t|\phi_q\rangle = \sum_q \varepsilon_{qp}\delta_{tq} = \varepsilon_{tp}, \tag{1.91}$$

hence the effect of a unitary transformation of the spinorbitals of the Lagrange multipliers is given by

$$\tilde{\varepsilon}_{pq} = \int \tilde{\phi}_p^*(\tau_i)\hat{f}(i)\tilde{\phi}_q(\tau_i)\mathrm{d}\tau_i = \sum_{tu} U_{tp}^* U_{uq}\int \phi_t^*(\tau_i)\hat{f}(i)\phi_u(\tau_i)\mathrm{d}\tau_i = \sum_{tu} U_{tp}^*\varepsilon_{tu}U_{uq} \quad\Longrightarrow\quad \tilde{\varepsilon} = \mathbf{U}^\dagger\varepsilon\mathbf{U}. \tag{1.92}$$

As the matrix of Lagrange multipliers ε is Hermitian, it is always possible to find a unitary matrix that diagonalises it. Therefore, there must exist a set of spinorbitals for which the matrix of Lagrange multipliers is diagonal. These spinorbitals and Lagrange multipliers are defined as the solution to the *canonical Hartree–Fock equations*,

$$\hat{f}\,|\phi_i\rangle = \varepsilon_i\,|\phi_i\rangle \tag{1.93}$$

1.2.3.4 Spin Resolution In Hartree–Fock

In the preceding sections, the HF equations have been presented in spinorbital form. However, the orthogonality of the spin functions allows two general forms of the HF equations to be considered; a *restricted Hartree–Fock* (RHF) model for systems with no unpaired electrons and an *unrestricted Hartree–Fock* (UHF) model for systems containing unpaired electrons.

Matrix elements of the one–electron Hamiltonian can be resolved into integrals over spin and spatial functions,

$$\langle p|\hat{h}|q\rangle = \int \varphi_p^*(\mathbf{r})\hat{h}\varphi_q(\mathbf{r})\mathrm{d}\mathbf{r}\int \eta_p^*(\sigma)\eta_q(\sigma)\mathrm{d}\sigma = (p|\hat{h}|q)\delta_{\eta_p\eta_q}, \tag{1.94}$$

where the term $(p|\hat{h}|q)$ is an integral over spatial orbitals, whilst the spin–integration reveals that $\langle p|\hat{h}|q\rangle$ is only non–zero if both spinorbitals have the same spin component. A similar analysis can be applied to the two–electron matrix elements as follows,

$$\langle pq|rs\rangle = \iint \frac{\varphi_p^*(\mathbf{r}_1)\varphi_q(\mathbf{r}_2)\varphi_r^*(\mathbf{r}_1)\varphi_s(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\int\eta_p^*(\sigma_1)\eta_r(\sigma_1)\mathrm{d}\sigma_1\int\eta_q^*(\sigma_2)\eta_s(\sigma_2)\mathrm{d}\sigma_2 = (pr|qs)\,\delta_{\eta_p\eta_r}\delta_{\eta_q\eta_s} \tag{1.95}$$

where the term $(pr|qs)$ is a two–electron integral over spatial orbitals in Mulliken notation. As Coulomb integrals have the form $(ii|jj)$, all elements are non–zero following spin integration. However, exchange integrals of the form $(ij|ji)$ are necessarily zero if the two spinorbitals i and j do not have the same spin function.

[iv] As $\det\{\mathbf{U}^\dagger\mathbf{U}\} = \det\{\mathbf{I}\} = 1$, it follows that $\det\{\mathbf{U}^\dagger\mathbf{U}\} = \det\{\mathbf{U}^\dagger\}\det\{\mathbf{U}\} = \det\{\mathbf{U}\}^*\det\{\mathbf{U}\} = 1$, hence $\det\{\mathbf{U}\} = \exp(\imath\omega)$.

Denoting a spinorbital index as a spatial orbital index with a given spin function, i.e. i^α and i^β, the only non–zero matrix elements of the one and two–electron Hamiltonian are those of the form

$$\begin{aligned}
\langle i|\hat{h}|i\rangle: &\quad (i^\alpha|\hat{h}|i^\alpha) \qquad (i^\beta|\hat{h}|i^\beta) \\
\langle ij|ij\rangle: &\quad (i^\alpha i^\alpha|j^\alpha j^\alpha) \quad (i^\alpha i^\alpha|j^\beta j^\beta) \quad (i^\beta i^\beta|j^\alpha j^\alpha) \quad (i^\beta i^\beta|j^\beta j^\beta) \\
\langle ij|ji\rangle: &\quad (i^\alpha j^\alpha|j^\alpha i^\alpha) \quad (i^\beta j^\beta|j^\beta i^\beta)
\end{aligned} \tag{1.96}$$

hence, for a system with n_α spin α and n_β spin β electrons respectively, the expectation value of the Hamiltonian (1.78) can be written as the spin–resolved sum over spatial orbitals,

$$\begin{aligned}
\mathcal{E}(\Phi) = &\sum_i^{n_\alpha}(i^\alpha|\hat{h}|i^\alpha) + \frac{1}{2}\sum_i^{n_\alpha}\sum_j^{n_\alpha}\{(i^\alpha i^\alpha|j^\alpha j^\alpha) - (i^\alpha j^\alpha|j^\alpha i^\alpha)\} + \frac{1}{2}\sum_i^{n_\alpha}\sum_j^{n_\beta}\left(i^\alpha i^\alpha|j^\beta j^\beta\right) \\
&+ \sum_i^{n_\beta}(i^\beta|\hat{h}|i^\beta) + \frac{1}{2}\sum_i^{n_\beta}\sum_j^{n_\beta}\left\{\left(i^\beta i^\beta|j^\beta j^\beta\right) - \left(i^\beta j^\beta|j^\beta i^\beta\right)\right\} + \frac{1}{2}\sum_i^{n_\beta}\sum_j^{n_\alpha}\left(i^\beta i^\beta|j^\alpha j^\alpha\right).
\end{aligned} \tag{1.97}$$

Using the definitions of the Coulomb and exchange operators in (1.79) and applying the same rules of spin–integration, it follows that spin–resolved Fock operators can be defined as

$$\begin{aligned}
\hat{f}^\alpha(i) &= \hat{h}(i) + \sum_j^{n_\alpha}\left(\hat{\mathcal{J}}_j^\alpha(i) - \hat{\mathcal{K}}_j^\alpha(i)\right) + \sum_j^{n_\beta}\hat{\mathcal{J}}_j^\beta(i) = \hat{h}(i) + \left(\sum_j^{n_\alpha}\hat{\mathcal{J}}_j^\alpha(i) + \sum_j^{n_\beta}\hat{\mathcal{J}}_j^\beta(i)\right) - \sum_j^{n_\alpha}\hat{\mathcal{K}}_j^\alpha(i) \\
\hat{f}^\beta(i) &= \hat{h}(i) + \sum_j^{n_\beta}\left(\hat{\mathcal{J}}_j^\beta(i) - \hat{\mathcal{K}}_j^\beta(i)\right) + \sum_j^{n_\alpha}\hat{\mathcal{J}}_j^\alpha(i) = \hat{h}(i) + \left(\sum_j^{n_\alpha}\hat{\mathcal{J}}_j^\alpha(i) + \sum_j^{n_\beta}\hat{\mathcal{J}}_j^\beta(i)\right) - \sum_j^{n_\beta}\hat{\mathcal{K}}_j^\beta(i),
\end{aligned} \tag{1.98}$$

in which both the α and β Fock operators share the same one–electron operator and Coulomb operator, but differ in their exchange operator. Using the same method as in subsubsection 1.2.3.3, the canonical UHF equations can be shown to be

$$\hat{f}^\alpha|\varphi_i^\alpha\rangle = \varepsilon_i^\alpha|\varphi_i^\alpha\rangle \qquad\qquad \hat{f}^\beta|\varphi_i^\beta\rangle = \varepsilon_i^\beta|\varphi_i^\beta\rangle, \tag{1.99}$$

where $\{|\varphi_i^\alpha\rangle\}$ and $\{|\varphi_i^\beta\rangle\}$ are two independent sets of spatial orbitals, coupled through the Coulomb operator.

For a system in which the number of electrons is even and there are no unpaired electrons, all occupied spinorbitals can be arranged in pairs that share the same spatial function but opposite spin function – a condition enforced in RHF. Therefore, the Coulomb and exchange operators are the same for both α and β spin thus the Fock operator is the same for both spins and can be defined as

$$\hat{f}(i) = \hat{h}(i) + \sum_j^{n/2}\left(\hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i)\right) + \sum_j^{n/2}\hat{\mathcal{J}}_j(i) = \hat{h}(i) + \sum_j^{n/2}\left(2\hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i)\right), \tag{1.100}$$

where the summation limits are $n_\alpha = n_\beta = n/2$, hence each spatial function can be considered to account for two electrons. The canonical RHF equations are given by

$$\hat{f}|\varphi_i\rangle = \varepsilon_i|\varphi_i\rangle, \tag{1.101}$$

where each eigenfunction is a spatial orbital considered to be occupied by an α and a β electron, with the corresponding eigenvalue having equal values for α and β spin. The expectation value of the Hamiltonian can be written in terms of doubly–occupied spatial orbitals as

$$\mathcal{E}(\Phi) = 2\sum_i^{n/2}(i|\hat{h}|i) + \sum_i^{n/2}\sum_j^{n/2}\{2\,(ii|jj) - (ij|ji)\}. \tag{1.102}$$

For both the canonical RHF and UHF equations, the eigenfunctions of the Fock operator are a set of spatial functions which can be referred to as *molecular orbitals* (MOs), whilst the eigenvalues are the *canonical orbital energies*.

1.2.3.5 Linear Combination Of Atomic Orbitals

The canonical Hartree–Fock equations (1.93) are a set of coupled one–electron eigenvalue equations, which may be solved in practice for small systems using numerical methods such as the finite difference approach [15]. However,

because the exchange term is a function of the HF orbitals themselves & is non–local, its inclusion significantly increases the degrees of freedom of the problem. For a general system, the dimensions of the finite–difference problem can therefore rapidly become very large and eventually impractical to solve using direct numerical methods.

Generally, the HF equations are cast in a computationally tractable form by expanding the molecular orbitals in a finite basis of atom–centred *atomic orbitals* (AOs) as

$$\varphi_i(\mathbf{r}) = \sum_{\mu}^{n_\chi} c_{\mu i}\chi_\mu(\mathbf{r}), \tag{1.103}$$

where $\{\chi_\mu\}$ are a set of n_χ AOs with linear expansion coefficients $\{c_{\mu i}\}$. This basis–set expansion is the Roothaan–Hall approach[16,17]. Considering first with general spinorbitals, the expansion (1.103) can be substituted into the canonical HF equations (1.93) to give

$$\hat{f}\sum_{\nu}^{n_\chi} c_{\nu i}\left|\chi_\nu\right\rangle = \varepsilon_i \sum_{\nu}^{n_\chi} c_{\nu i}\left|\chi_\nu\right\rangle . \tag{1.104}$$

This can then be left–multiplied by $\langle\chi_\mu|$ to give

$$\sum_{\nu}^{n_\chi} c_{\nu i}\left\langle\chi_\mu|\hat{f}|\chi_\nu\right\rangle = \varepsilon_i \sum_{\nu}^{n_\chi} c_{\nu i}\left\langle\chi_\mu|\chi_\nu\right\rangle \quad\Longrightarrow\quad \sum_{\nu}^{n_\chi} c_{\nu i}F_{\mu\nu} = \varepsilon_i \sum_{\nu}^{n_\chi} S_{\mu\nu}c_{\nu i}, \tag{1.105}$$

where $F_{\mu\nu}$ is the matrix element of the Fock operator in AO basis and $S_{\mu\nu}$ the overlap matrix element between the μ^{th} and ν^{th} AOs, which are not necessarily orthogonal. Considering the MO coefficients for the i^{th} MO as the i^{th} column in the matrix $\mathbf{C}$ of MO coefficients, the Roothaan–Hall equations can be written as the generalised psuedo–eigenvalue equation,

$$\mathbf{FC} = \mathbf{SC}\varepsilon. \tag{1.106}$$

The expectation value of the energy can be written in the basis set expansion for this general case by substituting the linear expansion into (1.78) as

$$\begin{aligned}
\mathcal{E}(\Phi) &= \sum_i (i|\hat{h}|i) + \frac{1}{2}\sum_{ij}\{(ii|jj) - (ij|ji)\} \\
&= \sum_i\sum_{\mu\nu} c^*_{\mu i}c_{\nu i}(\mu|\hat{h}|\nu) + \frac{1}{2}\sum_{ij}\sum_{\mu\nu} c^*_{\mu i}c_{\nu i}\sum_{\lambda\sigma} c^*_{\lambda j}c_{\sigma j}\{(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\} \\
&= \sum_{\mu\nu}\sum_i c^*_{\mu i}c_{\nu i}\left[(\mu|\hat{h}|\nu) + \frac{1}{2}\sum_{\lambda\sigma}\sum_j c^*_{\lambda j}c_{\sigma j}\{(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\}\right] \\
&= \sum_{\mu\nu} D_{\mu\nu}\left[h_{\mu\nu} + \frac{1}{2}\sum_{\lambda\sigma} D_{\lambda\sigma}\{(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\}\right]
\end{aligned} \tag{1.107}$$

where the expression has been simplified by introducing the *density matrix*,

$$D_{\mu\nu} = \sum_i^n c^*_{\mu i}c_{\nu i}. \tag{1.108}$$

A similar analysis can be carried out for the Fock operator, of the form given in (1.80), to yield

$$\begin{aligned}
F_{\mu\nu} &= \langle\chi_\mu|\hat{h}|\chi_\nu\rangle + \sum_i \langle\chi_\mu|\hat{\mathcal{J}}_i - \hat{\mathcal{K}}_i|\chi_\nu\rangle \\
&= h_{\mu\nu} + \sum_i\{(\mu\nu|ii) - (\mu i|i\nu)\} \\
&= h_{\mu\nu} + \sum_{\lambda\sigma}\sum_i c^*_{\lambda i}c_{\sigma i}\{(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\} \\
&= h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma}\{(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\}
\end{aligned} \tag{1.109}$$

hence the expectation value of the energy in the Roothaan–Hall method can be concisely written as a function of the expansion coefficients for the molecular orbitals in the atom–centred basis,

$$\mathcal{E}(\{c_{\mu i}\}) = \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}\left(h_{\mu\nu} + F_{\mu\nu}\right). \tag{1.110}$$

These expressions are relatively straightforward to adapt for the spin–restricted and spin–unrestricted HF models:

(i) For **RHF**, there is only one set of MO coefficients so only one density and Fock matrix, given by

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma} \left\{ 2\,(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu) \right\} \qquad D_{\mu\nu} = \sum_{i}^{n/2} c_{\mu i}^{*} c_{\nu i} \tag{1.111}$$

hence the RHF problem is characterised by one psuedo–eigenvalue equation of the form (1.106) and its energy expectation value is given by

$$\mathcal{E}(\{c_{\mu i}\}) = \sum_{\mu\nu} D_{\mu\nu} \left(h_{\mu\nu} + F_{\mu\nu} \right). \tag{1.112}$$

(ii) For **UHF**, there is one set of MO coefficients for each spin state α and β, hence there are also two sets of Fock and density matrices given by

$$\begin{aligned} F_{\mu\nu}^{\alpha} &= h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\alpha} \left\{ (\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu) \right\} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\beta} \,(\mu\nu|\lambda\sigma) \qquad D_{\mu\nu}^{\alpha} = \sum_{i}^{n_\alpha} c_{\mu i}^{\alpha *} c_{\nu i}^{\alpha} \\ F_{\mu\nu}^{\beta} &= h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\beta} \left\{ (\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu) \right\} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\alpha} \,(\mu\nu|\lambda\sigma) \qquad D_{\mu\nu}^{\beta} = \sum_{i}^{n_\beta} c_{\mu i}^{\beta *} c_{\nu i}^{\beta} \end{aligned} \tag{1.113}$$

thus the UHF problem is characterised by a pair of psuedo–eigenvalue equations[18,19],

$$\mathbf{F}^{\alpha}\mathbf{C}^{\alpha} = \mathbf{S}\mathbf{C}^{\alpha}\varepsilon^{\alpha} \qquad\qquad \mathbf{F}^{\beta}\mathbf{C}^{\beta} = \mathbf{S}\mathbf{C}^{\beta}\varepsilon^{\beta} \tag{1.114}$$

and the expectation value of the energy is given by

$$\mathcal{E}(\{c_{\mu i}^{\alpha}, c_{\mu i}^{\beta}\}) = \frac{1}{2} \sum_{\mu\nu} \left\{ D_{\mu\nu}^{\alpha} \left(h_{\mu\nu} + F_{\mu\nu}^{\alpha} \right) + D_{\mu\nu}^{\beta} \left(h_{\mu\nu} + F_{\mu\nu}^{\beta} \right) \right\}. \tag{1.115}$$

Generally the atom–centred basis functions do not form an orthonormal set, however in order to re–cast this problem as a standard eigenvalue equation it is necessary to transform the basis into an orthonormal set $|\tilde{\mu}\rangle$,

$$|\tilde{\mu}\rangle = \sum_{\nu} X_{\nu\tilde{\mu}} \,|\nu\rangle \quad \text{such that} \quad \langle\tilde{\mu}|\tilde{\nu}\rangle = \delta_{\tilde{\mu}\tilde{\nu}}. \tag{1.116}$$

The transformation matrix $\mathbf{X}$ can be identified by substituting the original basis functions into the inner product as

$$\langle\tilde{\mu}|\tilde{\nu}\rangle = \sum_{\mu\nu} X_{\mu\tilde{\mu}}^{*} \,\langle\mu|\nu\rangle\, X_{\nu\tilde{\nu}} = \sum_{\mu\nu} X_{\mu\tilde{\mu}}^{*} S_{\mu\nu} X_{\nu\tilde{\nu}} = \delta_{\tilde{\mu}\tilde{\nu}} \qquad \Longrightarrow \qquad \mathbf{X}^{\dagger}\mathbf{S}\mathbf{X} = \mathbf{1}. \tag{1.117}$$

Given the Hermiticity of the overlap matrix $\mathbf{S}$, there must exist a unitary matrix $\mathbf{U}$ which diagonalises it as

$$\mathbf{U}^{\dagger}\mathbf{S}\mathbf{U} = \mathbf{s}, \tag{1.118}$$

where $\mathbf{s}$ is the vector of eigenvalues. Using the method of *canonical orthogonalisation*, $\mathbf{X}$ can thus be defined as

$$\mathbf{X} = \mathbf{U}\mathbf{s}^{-\frac{1}{2}} \qquad \Longrightarrow \qquad X_{\mu\nu} = U_{\mu\nu}/\sqrt{s_\nu}, \tag{1.119}$$

which is shown to be an orthogonalising transformation matrix by substitution back into (1.116) to give

$$\mathbf{X}^{\dagger}\mathbf{S}\mathbf{X} = (\mathbf{U}\mathbf{s}^{-\frac{1}{2}})^{\dagger}\mathbf{S}(\mathbf{U}\mathbf{s}^{-\frac{1}{2}}) = \mathbf{s}^{-\frac{1}{2}}\mathbf{U}^{\dagger}\mathbf{S}\mathbf{U}\mathbf{s}^{-\frac{1}{2}} = \mathbf{s}^{-\frac{1}{2}}\mathbf{s}\mathbf{s}^{-\frac{1}{2}} = \mathbf{1}. \tag{1.120}$$

Transforming the MO coefficients according to (1.116) yields an orthonormal MO coefficient matrix $\tilde{\mathbf{C}}$,

$$\tilde{\mathbf{C}} = \mathbf{X}^{-1}\mathbf{C} \qquad \Longrightarrow \qquad \mathbf{C} = \mathbf{X}\tilde{\mathbf{C}} \tag{1.121}$$

which may then be substituted into (1.106), then the resulting expression left–multiplied by $\mathbf{X}^{\dagger}$, to give

$$\mathbf{X}^{\dagger}\mathbf{F}\mathbf{X}\tilde{\mathbf{C}} = \mathbf{X}^{\dagger}\mathbf{S}\mathbf{X}\tilde{\mathbf{C}}\varepsilon. \tag{1.122}$$

Defining $\tilde{\mathbf{F}} = \mathbf{X}^{\dagger}\mathbf{F}\mathbf{X}$ and using the identity in (1.116) allows (1.106) to be written as a standard eigenvalue equation

$$\tilde{\mathbf{F}}\tilde{\mathbf{C}} = \tilde{\mathbf{C}}\varepsilon \tag{1.123}$$

which can be solved simply by diagonalisation of $\tilde{\mathbf{F}}$. The MO coefficients can be transformed back into the AO basis using the relations in (1.121). This is trivial to apply to RHF, whilst for UHF the same $\mathbf{X}$ is used to transform both α and β equations in (1.114).

To solve the Roothaan–Hall equations, the expectation value of the energy must be minimised with respect to the MO coefficients according to the variation principle. For practical calculations, the Roothaan–Hall method translates the Hartree–Fock problem into a series of matrix equations that are both convenient and efficient to compute. As the problem is non–linear, a process of iterative improvement is required to converge on a solution; a general outline of the *self–consistent field* approach to solving these equations is given below:

1. The one and two–electron integrals over the basis functions, $(\mu|\hat{h}|\nu)$ and $(\mu\nu|\lambda\sigma)$ respectively, are evaluated to allow the Fock matrix to be constructed. As it is the MO coefficients that are being optimised, it is not necessary to re–compute these integrals for each Fock matrix build, although to do so is sometimes either necessary, due to insufficient computational storage space, or even advantageous where computational activity is sufficiently faster than data storage & retrieval.

2. The orthogonalisation matrix defined in (1.119) must be computed from the overlap matrix so that the HF equations can be cast in a standard eigenvalue form.

3. Initially, the MO coefficient and density matrices are not defined, so the Fock matrix cannot be constructed. An initial guess of the MO coefficients is commonly obtained by transforming the one–electron Hamiltonian matrix $\mathbf{H}$ to the orthogonal basis and diagonalising it as

$$\begin{aligned} \tilde{\mathbf{H}} &= \mathbf{X}^\dagger \mathbf{H} \mathbf{X} \\ \tilde{\mathbf{H}}\tilde{\mathbf{C}} &= \tilde{\mathbf{C}}\varepsilon \\ \mathbf{C} &= \mathbf{X}\tilde{\mathbf{C}} \\ \mathbf{D} &= \mathbf{C}^\dagger \mathbf{C} \end{aligned} \tag{1.124}$$

4. The Fock matrix may now be constructed from $\mathbf{D}$, $\mathbf{H}$ and the two–electron integrals according to (1.111) or (1.113). This is then transformed and diagonalised in a similar way to similarly to (1.124) to obtain a new density matrix,

$$\begin{aligned} \tilde{\mathbf{F}} &= \mathbf{X}^\dagger \mathbf{F} \mathbf{X} \\ \tilde{\mathbf{F}}\tilde{\mathbf{C}} &= \tilde{\mathbf{C}}\varepsilon \\ \mathbf{C} &= \mathbf{X}\tilde{\mathbf{C}} \\ \mathbf{D} &= \mathbf{C}^\dagger \mathbf{C} \end{aligned} \tag{1.125}$$

5. The energy at this iteration is calculated using (1.112) or (1.115) and compared with the energy of the previous iteration. In addition, the modulus–squared of the change in the density matrix in this iteration may be calculated. If these are both below a certain threshold, the self–consistent procedure may be considered to have converged.

6. If these two differences are not below the required threshold, then a further iteration of this method is carried out; a new Fock matrix is constructed from the present density matrix, and transformed to the orthogonal space and diagonalised, to construct a further density matrix. These iterations continue until the changes in energy and in the density matrix fall below the required threshold for convergence.

The practical implementation of the HF method demonstrates how variational solutions may be obtained using iterative refinement. The HF method has become a standard and reliable tool used in electronic structure calculations, however the single–determinant restriction on the HF wavefunction has the consequence that the HF method fails to accurately describe many chemically important phenomena, notably *electron correlation*. Several theories have been developed that aim to address this failure by extending the HF method beyond the single Slater determinant; these will be discussed in Chapter 2.

2 Electron Correlation

As was discussed in subsections 1.2.2 and 1.2.3, the Hartree–Fock approximation only accounts for electron–electron interaction by a mean–field approach – where each electron interacts is considered to interact with the average of all the other electrons. This does not however constitute an accurate description of electron interactions as it omits the effects of instantaneous interactions between individual charge distributions and of near–degenerate electronic states. Thus, as is inherent in the variation principle, the exact ground–state energy of a system $\mathcal{E}_0$ will always be *lower* than that predicted by the HF approximation $\mathcal{E}_{\text{HF}}$. The difference between the two was defined by Löwdin as the *correlation energy* of the system[20],

$$\mathcal{E}_{\text{c}} = \mathcal{E}_0 - \mathcal{E}_{\text{HF}} \tag{2.1}$$

The correlation energy of a system typically comprises only a small percentage of its total energy, on the order of 1%. However it can often be the difference between a prediction that is qualitatively useful and one that has quantitative accuracy. Additionally, the correlation energy is often much larger in size as a proportion of the *energy differences* between a system in two states - such as that between a diatomic molecule and two independent atoms; treatment of electron correlation is therefore essential for making accurate predictions of chemical properties.

Electron correlation is often considered to have two constituent parts - *static correlation* and *dynamical correlation*. These are broadly defined as follows:

(i) The omission of instantaneous Coulomb interactions between the electrons in the HF model can be characterised as the origin of dynamical correlation, as this description implies an error in the treatment of the instantaneous interactions between electrons.

(ii) The limitations of modelling a wavefunction as a single Slater determinant, notably in systems with degenerate electronic configurations, are generally considered the origin of static correlation. For some systems, a single Slater determinant is a poor representation of the physical wavefunction; a linear combination of a few Slater determinants constitutes a far more accurate model. In contrast to dynamical correlation, it is not directly related to instantaneous electron interactions hence the description of static (or *nondynamical*) correlation.

The difference between static and dynamical correlation is a qualitative rather than quantitative one as they both arise from the same physical interaction - the electrostatic interaction between electrons, hence they can be difficult to identify individually.

The effects of both static and dynamical correlation can be taken into account by approximating the physical wavefunction Ψ as a linear combination of Slater determinants Φ_I, including the HF determinant Φ_0 as

$$\Psi \approx c_0\Phi_0 + \sum_I c_I\Phi_I, \tag{2.2}$$

where the coefficients $\sum_I c_I^* c_I = 1$. The relative values of the optimised coefficients can be used to characterise the dominant form of correlation present in the system in the following way:

(i) If $c_I \ll c_0$ for $I \geqslant 1$, whilst $c_0 \approx 1$, it implies a physical wavefunction well described by a single determinant for which the series of other determinants have small coefficients and constitute only a small correction to the HF wavefunction - this suggests the correlation is primarily dynamical.

(ii) If $c_0 < 1$ whilst $c_I \approx c_0$ for only a few determinants I and $c_I \approx 0$ otherwise, it implies a physical wavefunction poorly described by a single determinant and which exhibits degeneracy or near–degeneracy. The wavefunction is corrected by mixing a few determinants, including the HF determinant, with coefficients a similar order of magnitude in weight, suggesting the correlation is primarily static.

There are a range of *ab initio* methods commonly used to treat electron correlation. In practice, different methods are typically better suited to account primarily for either static or dynamical correlation.

Dynamical correlation can be treated by mixing the HF determinant with the manifold of *excited–state determinants* that can be constructed from it, each of which will have a relatively small coefficient. In contrast, static correlation is generally treated by constructing a *multi–configurational* reference wavefunction as the linear combination of a few significant Slater determinants, for example using the multi–configurational SCF (MCSCF) method, which can be subsequently corrected for dynamical correlation by mixing in excited state configurations. In the following sections, the main *ab initio* methods for including dynamical correlation effects are introduced, as multi–configurational methods were not used in the present work whereas the following methods were extensively utilised.

2.1 General Principles

In the form of the HF equations given in subsubsection 1.2.3.5, a number of basis functions m are used to model an equal number m of spatial orbitals - describing $2m$ spinorbitals. If the number of spinorbitals this yields is greater than the number of electrons n in the system, the ground–state determinant will contain an electron in the n spinorbitals with the lowest ε; these are the *occupied* spinorbitals. The remaining $(2m - n)$ spinorbitals with higher ε will not be occupied by an electron; these are the *virtual* spinorbitals. The following notation is used to distinguish occupied and virtual spinorbitals

$$\textbf{occupied}: \rightarrow (i, j, k, l, m, n, o) \qquad \textbf{virtual}: \rightarrow (a, b, c, d, e, f, g) \qquad \textbf{general}: \rightarrow (p, q, r, s, t, u, v) \tag{2.3}$$

thus excited–state configurations of a HF determinant can be obtained by exciting one or more electrons from the n occupied spinorbitals into the $(2m - n)$ virtual spinorbitals.

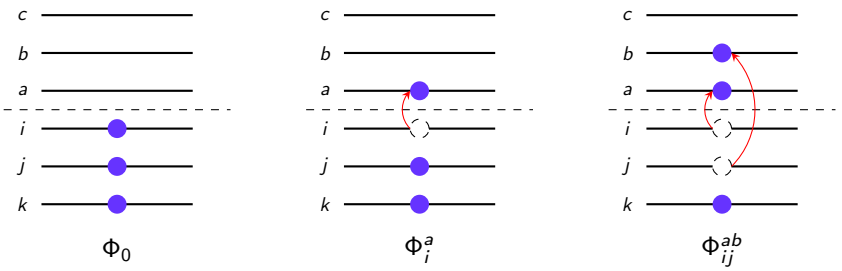

Figure 2.1: A schematic showing the energy levels of the occupied and virtual spinorbitals in the ground–state (left), singly–excited (centre) and doubly–excited (right) Hartree Fock determinants.

The schematic in Figure 2.1 demonstrates the notation used for the excited states of the HF determinant Φ_0; Φ_i^a indicates an excitation of an electron from occupied spinorbital i into virtual spinorbital a and Φ_{ij}^{ab} excitations from occupied spinorbitals i & j into virtual spinorbitals a & b.

Expressions for the matrix elements of the Hamiltonian between different states of the determinant wavefunction can be derived in a similar way to those for the expectation value of the Hamiltonian for the ground–state wavefunction, shown in subsubsections 1.2.3.1 and 1.2.3.2. The resulting expressions constitute the *Slater–Condon rules*, which identify the matrix elements of the Hamiltonian between determinants by the number of *non–coincident* spinorbitals; these are summarised in Table 2.1.

Non–coincidences	$\|\mathbf{I}\rangle$	$\|\mathbf{J}\rangle$	$\langle\mathbf{I}\|\hat{\mathcal{H}}\|\mathbf{J}\rangle$
0	Φ_0	Φ_0	$\sum_i \langle i\|\hat{h}\|i\rangle + \frac{1}{2}\sum_{ij}\langle ij\|\|ij\rangle$
1	Φ_0	Φ_p^q	$\langle p\|\hat{h}\|q\rangle + \sum_i \langle pi\|\|qi\rangle$
2	Φ_0	Φ_{pq}^{rs}	$\langle pq\|\|rs\rangle$
3	Φ_0	Φ_{pqr}^{stu}	0

Table 2.1: Matrix elements between determinants by the number of non–coincident spinorbitals between them.

As can be seen in Table 2.1, determinants with three or more non–coincident spinorbitals have vanishing matrix elements. These rules can be applied to any pair of determinants, ground–state or excited–state. An important instance of these rules can be seen for matrix elements of the ground–state determinant Φ_0 and singly–excited determinant Φ_i^a,

$$\langle\Phi_0|\hat{\mathcal{H}}|\Phi_i^a\rangle = \langle i|\hat{h}|a\rangle + \sum_j \langle ij||aj\rangle\,. \tag{2.4}$$

From (1.80), it is clear this expression amounts to the matrix element of the Fock operator $\langle i|\hat{f}|a\rangle$. However, matrix elements of the Fock operator can also be derived from the canonical HF equations (1.93), yielding the identity

$$\langle i|\hat{f}|a\rangle = \varepsilon_i \langle i|a\rangle = \varepsilon_i \delta_{ia} = 0. \tag{2.5}$$

It thus follows that the matrix element of the Hamiltonian between the ground–state HF determinant and the singly–excited state is zero; this is a statement of *Brillouin's theorem*[21].

2.2 Configuration Interaction

The method of *configuration interaction* (CI) is perhaps the most straightforward extension to the HF approximation; the physical wavefunction Ψ is represented as a linear combination of the HF determinant and its *excited state determinants*[22],

$$\Psi = c_0\Phi_0 + \sum_{ia} c_i^a \Phi_i^a + \frac{1}{4}\sum_{ijab} c_{ij}^{ab}\Phi_{ij}^{ab} + \frac{1}{36}\sum_{ijkabc} c_{ijk}^{abc}\Phi_{ijk}^{abc} + \ldots \tag{2.6}$$

which can be extended to the limit of all possible excited–state determinants. The expression for the CI wavefunction (2.6) can be simplified by grouping determinants and coefficients by excitation level,

$$|\Psi\rangle = c_0\,|0\rangle + c_S\,|S\rangle + c_D\,|D\rangle + c_T\,|T\rangle + \ldots = \sum_I c_I\,|I\rangle\,, \tag{2.7}$$

where $|0\rangle$ is the ground–state while $|S\rangle$, $|D\rangle$ and $|T\rangle$ are the blocks of singly, doubly and triply–excited determinants respectively. Assuming the CI wavefunction is normalised as $\sum_I c_I^* c_I = 1$, it can be substituted into the Schrödinger Equation of (1.1) to give

$$\hat{\mathcal{H}}\,|\Psi\rangle = \mathcal{E}\,|\Psi\rangle \qquad \Longrightarrow \qquad \hat{\mathcal{H}}\sum_I c_I\,|I\rangle = \mathcal{E}\sum_I c_I\,|I\rangle\,. \tag{2.8}$$

The resulting expression may be left–multiplied by $\langle J|$ and simplified as

$$\sum_I \langle J|\hat{\mathcal{H}}|I\rangle\, c_I = \mathcal{E}\sum_I \langle J|I\rangle\, c_I \quad \Longrightarrow \quad \sum_I H_{JI}c_I = \mathcal{E}\sum_I \delta_{JI}c_I = \mathcal{E}c_J \quad \Longrightarrow \quad \mathbf{Hc} = \mathcal{E}\mathbf{c}, \tag{2.9}$$

where $\mathbf{H}$ is the CI Hamiltonian matrix and $\mathbf{c}$ the expansion coefficients for the ground–state CI wavefunction with ground–state energy $\mathcal{E}$. This eigenvalue equation may be written explicitly in terms of the excitation–level blocks shown in (2.7) as

$$\begin{pmatrix} \langle 0|\hat{\mathcal{H}}|0\rangle & 0 & \langle 0|\hat{\mathcal{H}}|D\rangle & 0 & \ldots \\ 0 & \langle S|\hat{\mathcal{H}}|S\rangle & \langle S|\hat{\mathcal{H}}|D\rangle & \langle S|\hat{\mathcal{H}}|T\rangle & \ldots \\ \langle D|\hat{\mathcal{H}}|0\rangle & \langle D|\hat{\mathcal{H}}|S\rangle & \langle D|\hat{\mathcal{H}}|D\rangle & \langle D|\hat{\mathcal{H}}|T\rangle & \ldots \\ 0 & \langle T|\hat{\mathcal{H}}|S\rangle & \langle T|\hat{\mathcal{H}}|D\rangle & \langle T|\hat{\mathcal{H}}|T\rangle & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} c_0 \\ c_S \\ c_D \\ c_T \\ \vdots \end{pmatrix} = \mathcal{E}\begin{pmatrix} c_0 \\ c_S \\ c_D \\ c_T \\ \vdots \end{pmatrix} \tag{2.10}$$

where the Slater–Condon rules from Table 2.1 are used to eliminate the $\langle 0|\hat{\mathcal{H}}|T\rangle$ & $\langle T|\hat{\mathcal{H}}|0\rangle$ blocks in the Hamiltonian matrix, whilst Brillouin's theorem requires the $\langle 0|\hat{\mathcal{H}}|S\rangle$ & $\langle S|\hat{\mathcal{H}}|0\rangle$ blocks to be zero.

The expression (2.6) in principle extends to include every possible excitation of the reference determinant Φ_0; this is the *Full Configuration Interaction* (FCI) model. The HF determinant and the set of all excited–state determinants together constitute a complete basis in the space of one–electron orbitals hence the energy $\mathcal{E}$ obtained by diagonalising the FCI Hamiltonian is the exact energy in a given basis set; in the limit of a complete basis set, the FCI energy is considered to be exact.

In practice however, FCI calculations are only feasible for very small systems with a limited number of basis functions due to the very large number of determinants that make–up the full expansion. There are $\binom{2m}{n}$ possible arrangements of n electrons amongst $2m$ spinorbitals with the result that the size of the FCI problem scales as roughly $\sim m!$. Although many algorithms have been developed to try and make large–scale CI calculations more computationally manageable, see for example Refs. 23–25, it becomes necessary to truncate the expansion (2.6) in the CI treatment of larger systems.

A common level of truncation in general–purpose post–HF electronic structure calculations is to that of *configuration interaction singles and doubles* (CISD), in which only the singly and doubly excited Slater determinants are included in the CI expansion,

$$\Psi_{\text{CISD}} = c_0\Phi_0 + \sum_{ia} c_i^a\Phi_i^a + \frac{1}{4}\sum_{ijab} c_{ij}^{ab}\Phi_{ij}^{ab}. \tag{2.11}$$

This approximation is generally effective in accounting for correlation energy, recovering in the order of $\sim 90\%$ of that recovered in FCI.[i]

[i] Although this is highly dependent on basis–set size and the characteristics of the system itself.

However, truncated CI approximations have an important shortcoming - they are not *size–consistent*. For a method to be size–consistent, the total energy of a system AB composed of two non–interacting subsystems A & B must be equal to the sum of the energies of the separate subsystems; given their respective Hamiltonians $\hat{\mathcal{H}}$, wavefunctions Φ and energies $\mathcal{E}$ this relation is expressed as

$$\hat{\mathcal{H}}_A\Phi_A = \mathcal{E}_A\Phi_A \qquad \hat{\mathcal{H}}_B\Phi_B = \mathcal{E}_B\Phi_B \qquad \hat{\mathcal{H}}_{AB}\Phi_{AB} = \mathcal{E}_{AB}\Phi_{AB} \qquad \Longrightarrow \qquad \mathcal{E}_{AB} = \mathcal{E}_A + \mathcal{E}_B. \tag{2.12}$$

This can be seen to apply for the UHF method described in subsubsection 1.2.3.4, where the orbitals of the two non–interacting subsystems have no overlap hence $\Phi_{AB} = \Phi_A\Phi_B$. It follows that, for such a system, the Schrödinger Equation can be re–arranged as

$$\hat{\mathcal{H}}_{AB}\Phi_{AB} = (\mathcal{E}_A + \mathcal{E}_B)\Phi_A\Phi_B = (\hat{\mathcal{H}}_A\Phi_A)\Phi_B + \Phi_A(\hat{\mathcal{H}}_B\Phi_B) = (\hat{\mathcal{H}}_A + \hat{\mathcal{H}}_B)\Phi_{AB}, \tag{2.13}$$

demonstrating that the wavefunction operators are also additive.

It can be readily demonstrated that truncated CI is not necessarily size–consistent using the example of a Helium atom, for which the CISD model is equivalent to FCI as only two electrons are present hence only single and double excitations are possible. For a system of two Helium atoms, in which there are four electrons, the FCI wavefunction will contain triply–excited and quadruply–excited determinants however the CISD model will only contain up to doubly–excited determinants. The result is that the FCI Hamiltonian matrix,

$$\begin{pmatrix} \langle 0|\hat{\mathcal{H}}|0\rangle & 0 & \langle 0|\hat{\mathcal{H}}|D\rangle & 0 & 0 \\ 0 & \langle S|\hat{\mathcal{H}}|S\rangle & \langle S|\hat{\mathcal{H}}|D\rangle & \langle S|\hat{\mathcal{H}}|T\rangle & 0 \\ \langle D|\hat{\mathcal{H}}|0\rangle & \langle D|\hat{\mathcal{H}}|S\rangle & \langle D|\hat{\mathcal{H}}|D\rangle & \langle D|\hat{\mathcal{H}}|T\rangle & \langle D|\hat{\mathcal{H}}|Q\rangle \\ 0 & \langle T|\hat{\mathcal{H}}|S\rangle & \langle T|\hat{\mathcal{H}}|D\rangle & \langle T|\hat{\mathcal{H}}|T\rangle & \langle T|\hat{\mathcal{H}}|Q\rangle \\ 0 & 0 & \langle Q|\hat{\mathcal{H}}|D\rangle & \langle Q|\hat{\mathcal{H}}|T\rangle & \langle Q|\hat{\mathcal{H}}|Q\rangle \end{pmatrix} \tag{2.14}$$

contains significant blocks that are omitted in the CISD Hamiltonian matrix - these are shaded in (2.14). It follows that the FCI and CISD energies of He_2 differ, so the CISD energy of He_2 is not equal to twice that of He.

This shortcoming of truncated CI represents a significant problem, particularly for the accurate calculation of energy differences in bond formation or bond fission. Theories have been developed to partially correct this error[26], notably the Davidson correction[27] which estimates the energy difference between the CISD and CISDTQ approximations. Ultimately, more accurate results using CI are only obtained by including higher–order excitations in the CI expansion.

2.3 Perturbation Theory

An alternative approach to the treatment of electron correlation is by application of *perturbation theory*. Perturbation theory provides the means of approximating the solution to a problem using the principle that, if the solution to a similar problem is known, then the solution to the unknown problem can be approximated as the known solution plus a small correction.

This method is predicated on an assumption that the difference in energy between the known and perturbed system is small compared to the energy difference between states of the known system. In the context of electronic structure theory, the correlation energy of a system is often only around 1% its total energy, hence perturbation theory is a reasonable approach that may be taken to estimate correlation energy.

2.3.1 Rayleigh–Schrödinger Perturbation Theory

A convenient framework for the application of perturbation theory is that formalised by Rayleigh[28] and Schrödinger[1]: *Rayleigh–Schrödinger perturbation theory* (RSPT). According to RSPT, the Hamiltonian $\hat{\mathcal{H}}$ can be separated into the sum of a zeroth–order component $\hat{\mathcal{H}}_0$ and a perturbative operator $\hat{\mathcal{V}}$, the magnitude of which is scaled by some coefficient λ as

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \lambda\hat{\mathcal{V}}. \tag{2.15}$$

The Schrödinger Equation (1.1) can hence be written as

$$\hat{\mathcal{H}}\Psi_n = \left(\hat{\mathcal{H}}_0 + \lambda\hat{\mathcal{V}}\right)\Psi_n = \mathcal{E}_n\Psi_n. \tag{2.16}$$

The zeroth–order Hamiltonian $\hat{\mathcal{H}}_0$ is taken to have known eigenfunctions $\Psi_n^{(0)}$ and eigenvalues $\mathcal{E}_n^{(0)}$, satisfying the eigenvalue equation

$$\hat{\mathcal{H}}_0\Psi_n^{(0)} = \mathcal{E}_n^{(0)}\Psi_n^{(0)} \qquad \left\langle \Psi_m^{(0)} \middle| \Psi_n^{(0)} \right\rangle = \delta_{mn}. \tag{2.17}$$

For a perturbation $\lambda\hat{\mathcal{V}}$ that is small compared to the zeroth–order Hamiltonian, it can be reasonably inferred that $\Psi_n \approx \Psi_n^{(0)}$ and $\mathcal{E}_n \approx \mathcal{E}_n^{(0)}$. Therefore, the exact eigenfunctions and eigenvalues can be approximated as the series expansions in λ,

$$\begin{aligned} \Psi_n &= \Psi_n^{(0)} + \lambda\Psi_n^{(1)} + \lambda^2\Psi_n^{(2)} + \dots \\ \mathcal{E}_n &= \mathcal{E}_n^{(0)} + \lambda\mathcal{E}_n^{(1)} + \lambda^2\mathcal{E}_n^{(2)} + \dots \end{aligned} \tag{2.18}$$

where $\Psi_n^{(k)}$ and $\mathcal{E}_n^{(k)}$ are the k^{th} order corrections to the eigenfunctions and eigenvalues of $\hat{\mathcal{H}}_0$. Given that both the eigenfunctions of the zeroth–order Hamiltonian are orthonormal and those of the full Hamiltonian are orthonormal, the wavefunction correction terms can be constrained such that they are orthogonal to the zeroth–order wavefunction,

$$\left\langle \Psi_n^{(0)} \middle| \Psi_n^{(k)} \right\rangle = \delta_{0k}. \tag{2.19}$$

As a result, the zeroth–order and corrected wavefunctions are mutually orthonormal at arbitrary order – described as *intermediate normalisation*,

$$\left\langle \Psi_n^{(0)} \middle| \Psi_n \right\rangle = \left\langle \Psi_n^{(0)} \middle| \Psi_n^{(0)} \right\rangle + \lambda \left\langle \Psi_n^{(0)} \middle| \Psi_n^{(1)} \right\rangle + \lambda^2 \left\langle \Psi_n^{(0)} \middle| \Psi_n^{(2)} \right\rangle + \dots = 1. \tag{2.20}$$

The series expansions in (2.18) can be substituted into the Schrödinger Equation (2.17) to give

$$\left(\hat{\mathcal{H}}_0 + \lambda\hat{\mathcal{V}}\right)\left(\Psi_n^{(0)} + \lambda\Psi_n^{(1)} + \lambda^2\Psi_n^{(2)} + \dots\right) = \left(\mathcal{E}_n^{(0)} + \lambda\mathcal{E}_n^{(1)} + \lambda^2\mathcal{E}_n^{(2)} + \dots\right)\left(\Psi_n^{(0)} + \lambda\Psi_n^{(1)} + \lambda^2\Psi_n^{(2)} + \dots\right). \tag{2.21}$$

Collecting the terms in (2.21) by order in λ yields the set of equations

$$\begin{aligned} \lambda^0 : &\quad \left(\hat{\mathcal{H}}_0 - \mathcal{E}_n^{(0)}\right)\Psi_n^{(0)} = 0 \\ \lambda^1 : &\quad \left(\hat{\mathcal{H}}_0 - \mathcal{E}_n^{(0)}\right)\Psi_n^{(1)} = \left(\mathcal{E}_n^{(1)} - \hat{\mathcal{V}}\right)\Psi_n^{(0)} \\ \lambda^2 : &\quad \left(\hat{\mathcal{H}}_0 - \mathcal{E}_n^{(0)}\right)\Psi_n^{(2)} = \left(\mathcal{E}_n^{(1)} - \hat{\mathcal{V}}\right)\Psi_n^{(1)} + \mathcal{E}_n^{(2)}\Psi_n^{(0)}, \end{aligned} \tag{2.22}$$

each of which can be projected onto the bra–state of the zeroth–order wavefunction $\Psi_n^{(0)}$ and, using intermediate normalisation, the expressions for the k^{th} order energy contributions of (2.18) are obtained as

$$\begin{aligned} \lambda^0 : &\quad \mathcal{E}_n^{(0)} = \left\langle \Psi_n^{(0)} \middle| \hat{\mathcal{H}}_0 \middle| \Psi_n^{(0)} \right\rangle \\ \lambda^1 : &\quad \mathcal{E}_n^{(1)} = \left\langle \Psi_n^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle \\ \lambda^2 : &\quad \mathcal{E}_n^{(2)} = \left\langle \Psi_n^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(1)} \right\rangle. \end{aligned} \tag{2.23}$$

Taking the equations in (2.22) order–by–order in λ, expressions for the wavefunction correction terms in (2.18) can be determined by projecting each onto the bra–state of the set of zeroth–order wavefunctions $\Psi_m^{(0)}$ with $m \neq n$. For the first–order equation this gives

$$\mathcal{E}_m^{(0)} \left\langle \Psi_m^{(0)} \middle| \Psi_n^{(1)} \right\rangle + \left\langle \Psi_m^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle = \mathcal{E}_n^{(0)} \left\langle \Psi_m^{(0)} \middle| \Psi_n^{(1)} \right\rangle, \tag{2.24}$$

which can be re–arranged to identify the projection coefficient of $\Psi_n^{(1)}$ onto $\Psi_m^{(0)}$,

$$\left\langle \Psi_m^{(0)} \middle| \Psi_n^{(1)} \right\rangle = \frac{\left\langle \Psi_m^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle}{\mathcal{E}_n^{(0)} - \mathcal{E}_m^{(0)}}. \tag{2.25}$$

Using the intermediate normalisation to eliminate $\langle \Psi_m^{(0)} | \Psi_m^{(1)} \rangle$, the first–order correction to the wavefunction can be written generally as

$$\Psi_n^{(1)} = \sum_{m \neq n} \frac{\left\langle \Psi_m^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle}{\mathcal{E}_n^{(0)} - \mathcal{E}_m^{(0)}} \Psi_m^{(0)}. \tag{2.26}$$

This may be substituted into the expression for the second–order correction to the energy in (2.23), allowing it to be expressed in terms of only second–order quantities as

$$\mathcal{E}_n^{(2)} = \sum_{m \neq n} \frac{\left\langle \Psi_m^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle}{\mathcal{E}_n^{(0)} - \mathcal{E}_m^{(0)}} \left\langle \Psi_n^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_m^{(0)} \right\rangle = \sum_{m \neq n} \frac{\left| \left\langle \Psi_m^{(0)} \middle| \hat{\mathcal{V}} \middle| \Psi_n^{(0)} \right\rangle \right|^2}{\mathcal{E}_n^{(0)} - \mathcal{E}_m^{(0)}}. \tag{2.27}$$

Similar relations may be derived for higher orders in the perturbation expansion, however (2.27) is sufficient for the purposes of this work.

2.3.2 Møller–Plesset Perturbation Theory

The principle of RSPT can be applied to wavefunction–based electronic structure methods by considering the Hartree–Fock wavefunction and energy as the zeroth–order approximation to its exact form. This approach was taken by Møller and Plesset to formulate an approximation for the correlation energy of a system described by a HF wavefunction; the method is known as *Møller–Plesset* (MP) perturbation theory[29].

In MP perturbation theory, the zeroth–order operator is taken to be the sum of the Fock operators (1.80) for each electron, given by

$$\hat{\mathcal{H}}_0 = \sum_i \hat{f}(i) = \sum_i \hat{h}(i) + \sum_{ij} \left(\hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i) \right). \tag{2.28}$$

The perturbation operator is defined as the difference between the exact and zeroth–order Hamiltonians $\hat{\mathcal{H}} - \hat{\mathcal{H}}_0$, which is seen to contain only two–electron operators as

$$\hat{\mathcal{V}} = \left(\sum_i \hat{h}(i) + \sum_{i>j} \hat{w}(i,j) \right) - \left(\sum_i \hat{h}(i) + \sum_{ij} \left(\hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i) \right) \right) = \sum_{i>j} \hat{w}(i,j) - \sum_{ij} \left(\hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i) \right). \tag{2.29}$$

This perturbation is often referred to as the fluctuation potential as it describes the difference between instantaneous electron–electron interactions and the mean–field treatment characterised by HF approximation.

Using the expressions for the energy correction terms (2.23) derived using RSPT, the MP energy terms for a Slater determinant Φ_0 can be identified by order as

$$\begin{aligned}
\mathcal{E}_0^{(0)} &= \langle \Phi_0 | \hat{\mathcal{H}}_0 | \Phi_0 \rangle = \sum_i \langle \Phi_0 | \hat{f}(i) | \Phi_0 \rangle = \sum_i \langle i | \hat{f} | i \rangle = \sum_i \varepsilon_i \\
\mathcal{E}_0^{(1)} &= \langle \Phi_0 | \, \hat{\mathcal{V}} \, | \Phi_0 \rangle = \sum_{i>j} \langle \Phi_0 | \hat{w}(i,j) | \Phi_0 \rangle - \sum_{ij} \langle \Phi_0 | \hat{\mathcal{J}}_j(i) - \hat{\mathcal{K}}_j(i) | \Phi_0 \rangle = -\frac{1}{2} \sum_{ij} \langle ij || ij \rangle \\
\mathcal{E}_0^{(2)} &= \sum_{n \neq 0} \frac{|\langle \Phi_n | \hat{\mathcal{V}} | \Phi_0 \rangle|^2}{\mathcal{E}_0^{(0)} - \mathcal{E}_n^{(0)}} = \sum_{ia} \frac{|\langle \Phi_i^a | \hat{\mathcal{V}} | \Phi_0 \rangle|^2}{\varepsilon_i - \varepsilon_a} + \frac{1}{4} \sum_{ijab} \frac{|\langle \Phi_{ij}^{ab} | \hat{\mathcal{V}} | \Phi_0 \rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} = \frac{1}{4} \sum_{ijab} \frac{|\langle ab || ij \rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}
\end{aligned} \tag{2.30}$$

where the second–order energy correction includes only contributions from doubly–excited determinants, as the HF determinant does not interact with singly–excited determinants due to Brillouin's theorem nor with triply–excited or higher determinants according to the Slater–Condon rules given in Table 2.1.

It can be seen that the sum of the zeroth and first–order Møller–Plesset energy terms simply yields the ground–state HF energy,

$$\mathcal{E}_0^{(0)} + \mathcal{E}_0^{(1)} = \sum_i \varepsilon_i - \frac{1}{2} \sum_{ij} \langle ij || ij \rangle = \sum_i h_{ii} + \sum_{ij} \langle ij || ij \rangle - \frac{1}{2} \sum_{ij} \langle ij || ij \rangle = \sum_i h_{ii} + \frac{1}{2} \sum_{ij} \langle ij || ij \rangle = \mathcal{E}_{\text{HF}}, \tag{2.31}$$

hence corrections to the HF energy first appear at second–order perturbation theory. These perturbative corrections to the HF energy describe correlation hence the simplest estimate of the correlation energy is given by the second–order energy correction of MP perturbation theory, commonly referred to as the *MP2* correlation energy,

$$\mathcal{E}_{\text{c}}^{\text{MP2}} = \frac{1}{4} \sum_{ijab} \frac{|\langle ij || ab \rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{2.32}$$

The MP2 treatment of correlation is relatively inexpensive computationally, and typically $\sim 80\%$ of the correlation energy of the system is accounted for by (2.32). Although in principle higher–orders of perturbation theory could be used in the same way to systematically improve the description of correlation, the series of MPN correlation energies can behave unpredictably at higher orders[30], as well as becoming significantly more computationally expensive. Rather than expand the MP series to higher and higher orders, alternative post–HF methods are usually preferable for a more advanced treatment of correlation.

2.4 Coupled-Cluster Theory

The final *ab initio* treatment of electron correlation to be discussed here is the *coupled–cluster* (CC) method which has, for many purposes, become the method of choice for reliable and accurate modelling of correlation effects in electronic systems. The CC method was initially proposed by Čížek in 1966[31] however it wasn't until some years later, after further work by a number of other research groups, that CC became established as an attractive alternative to MP & CI methods[32–38].

In a similar way to CI, the CC method accounts for electron correlation by modelling the exact wavefunction as a series in the HF reference determinant and the manifold of its excited–states. In contrast to CI however, the CC wavefunction is constructed using the *exponential ansatz*[31,32,39]

$$\Psi = e^{\hat{T}}\Phi_0 \tag{2.33}$$

where Φ is the HF reference determinant and $\hat{T}$ is the *cluster operator*, comprising the sum of excitation operators up to a given order,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots = \sum_{\lambda} \hat{T}_\lambda. \tag{2.34}$$

The excitation operator for a given order is generally defined using the *second–quantisation* formalism as

$$\hat{T}_1 = \sum_i \sum_a t_i^a \, \hat{a}^\dagger \hat{i} \qquad \hat{T}_2 = \left(\frac{1}{2}\right)^2 \sum_{ij} \sum_{ab} t_{ij}^{ab} \, \hat{a}^\dagger \hat{i} \, \hat{b}^\dagger \hat{j} \qquad \ldots \qquad \hat{T}_\lambda = \left(\frac{1}{\lambda!}\right)^2 \sum_{ij\ldots} \sum_{ab\ldots} t_{ij\ldots}^{ab\ldots} \, \hat{a}^\dagger \hat{i} \, \hat{b}^\dagger \hat{j} \ldots \tag{2.35}$$

where $\hat{a}^\dagger$ is a *creation* operator for spinorbital a, $\hat{i}$ is an *annihilation* operator for spinorbital i and $t_{ij\ldots}^{ab\ldots}$ are the excitation amplitudes. [ii] The action of each cluster operator in (2.35) on the HF determinant yields contributes to the CC wavefunction (2.33) of the form

$$\hat{T}_1\Phi_0 = \sum_i \sum_a t_i^a \Phi_i^a \qquad \hat{T}_2\Phi_0 = \left(\frac{1}{2}\right)^2 \sum_{ij} \sum_{ab} t_{ij}^{ab} \Phi_{ij}^{ab} \qquad \ldots \qquad \hat{T}_\lambda\Phi_0 = \left(\frac{1}{\lambda!}\right)^2 \sum_{ij\ldots} \sum_{ab\ldots} t_{ij\ldots}^{ab\ldots} \Phi_{ij\ldots}^{ab\ldots}. \tag{2.36}$$

The cluster operator in (2.34) can theoretically be extended to include all possible excitations in the system in the same way that the CI series (2.6) is extended to include all possible excited–state determinants in FCI.

In Section 2.2 it was shown that truncating the CI expansion and not including all possible terms yields a method that is not size–consistent. By contrast, in CC theory the cluster operator may be arbitrarily truncated and still yield a size–consistent wavefunction. Using the terms defined in (2.12), the identity in (2.13) can be applied to the CC method, given cluster operators $\hat{T}_\mathrm{A}$ & $\hat{T}_\mathrm{B}$ for subsystems A & B respectively, the overall cluster operator for the system AB must simply be the sum of those for each subsystem; $\hat{T}_\mathrm{AB} = \hat{T}_\mathrm{A} + \hat{T}_\mathrm{B}$. Substituting Φ_AB & $\hat{T}_\mathrm{AB}$ into (2.33) yields the CC wavefunction for the combined system, which can be re–arranged as

$$\Psi_\mathrm{AB} = e^{\hat{T}_\mathrm{AB}}\Phi_\mathrm{AB} = e^{\hat{T}_\mathrm{A}+\hat{T}_\mathrm{B}}\Phi_\mathrm{A}\Phi_\mathrm{B} = e^{\hat{T}_\mathrm{A}}\Phi_\mathrm{A}e^{\hat{T}_\mathrm{B}}\Phi_\mathrm{B} = \Psi_\mathrm{A}\Psi_\mathrm{B}. \tag{2.37}$$

The size–consistency of the CC model can thus be seen as a direct result of its exponential *ansatz*, as the CC wavefunction for a combined system is equal to the product of those of its constituent non–interacting subsystems.

The advantage of the CC method over the CI method may be demonstrated by comparing their second–order truncated forms; CC singles and doubles (CCSD)[44] and CISD respectively. In CCSD, the cluster operator contains only single and double–excitations and is given by $\hat{T} = \hat{T}_1 + \hat{T}_2$. Substituting this into (2.33) and using the Taylor series expansion of the exponential [iii], the CCSD wavefunction can be written as

$$\begin{aligned}\Psi_\mathrm{CCSD} = e^{\hat{T}}\Phi_0 &= \left(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \ldots\right)\Phi_0 \\ &= \Phi_0 + \sum_i \sum_a t_i^a \Phi_i^a + \frac{1}{4}\sum_{ij}\sum_{ab} t_{ij}^{ab}\Phi_{ij}^{ab} + \frac{1}{2}\sum_{ij}\sum_{ab} t_i^a t_j^b \Phi_{ij}^{ab} \\ &\quad + \frac{1}{4}\sum_{ijk}\sum_{abc} t_{ij}^{ab} t_k^c \Phi_{ijk}^{abc} + \frac{1}{32}\sum_{ijkl}\sum_{abcd} t_{ij}^{ab} t_{kl}^{cd} \Phi_{ijkl}^{abcd} + \ldots\end{aligned} \tag{2.38}$$

[ii] The second–quantisation is a theoretical framework in which quantum many–body systems can be described, using the creation and annihilation operators to represent the occupancy and vacancy respectively of one–particle states such as spinorbitals relative to a ground–state such as the HF state. The principles of the second–quantisation are beyond the scope of the present work, however the approach well documented in the literature, for example in Refs. 40–43.

[iii] $e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \ldots$

where contributions to the wavefunction linear in the cluster operator, i.e. $\hat{\mathcal{T}}_1\Phi_0$ and $\hat{\mathcal{T}}_2\Phi_0$, are referred to as *connected clusters* and those that involve the products of cluster operators, i.e. $\frac{1}{2}\hat{\mathcal{T}}_1^2\Phi$ and $\hat{\mathcal{T}}_1\hat{\mathcal{T}}_2\Phi$, are referred to as *disconnected clusters*. The relevance of this distinction becomes apparent on comparison with CISD, which may be written in a comparable form to (2.38) using 'CI excitation operators' $\hat{\mathcal{C}}$ as

$$\begin{aligned}\Psi_{\text{CISD}} &= \left(1 + \hat{\mathcal{C}}_1 + \hat{\mathcal{C}}_2\right)\Phi_0 \\ &= \Phi_0 + \sum_i \sum_a c_i^a \Phi_i^a + \frac{1}{4}\sum_{ij}\sum_{ab} c_{ij}^{ab}\Phi_{ij}^{ab},\end{aligned} \tag{2.39}$$

where it can be seen that connected clusters in the CC method give rise to terms similar to those appearing in the equivalent CI wavefunction whereas the disconnected clusters result in terms absent from the CI wavefunction. The disconnected clusters introduce higher–order excited determinants than those present in the CI expansion; it is by virtue of these disconnected clusters that CC is in practice size–consistent.

Although the CC method is not variational, the correlation energy is calculated through solving the CC equations, which requires an iterative determination of the cluster amplitudes, rather than through determining a set of coefficients for excited–state determinants as in CI.

2.4.1 Coupled–Cluster Equations

In order to compute the CC energy expectation value, it is first necessary to determine the cluster amplitudes $\mathbf{t} = (t_i^a, t_{ij}^{ab}, \ldots)$. This may be accomplished by firstly substituting the CC wavefunction (2.33) into the Schrödinger Equation to yield

$$\hat{\mathcal{H}}e^{\hat{\mathcal{T}}}\Phi_0 = \mathcal{E}_{\text{CC}}e^{\hat{\mathcal{T}}}\Phi_0 \qquad \Longrightarrow \qquad (\hat{\mathcal{H}} - \mathcal{E}_{\text{CC}})e^{\hat{\mathcal{T}}}\Phi_0 = 0, \tag{2.40}$$

onto which the manifold of excited–state determinants can be projected to yield a set of non–linear equations. In the derivation of these equations, it is convenient to introduce the *normal–ordered Hamiltonian*[iv]

$$\hat{\mathcal{H}}_{\text{N}} = \hat{\mathcal{H}} - \langle\Phi_0|\hat{\mathcal{H}}|\Phi_0\rangle, \tag{2.41}$$

in which the expectation value of the Hamiltonian on the HF determinant can be identified as the HF energy. Substituting the definition of correlation energy in (2.1) into (2.41) and re–arranging yields the identity

$$\mathcal{E}_{\text{CC}} = \langle\Phi_0|\hat{\mathcal{H}}|\Phi_0\rangle + \mathcal{E}_{\text{C}} \qquad \Longrightarrow \qquad \hat{\mathcal{H}}_{\text{N}} - \mathcal{E}_{\text{c}} = \hat{\mathcal{H}} - \mathcal{E}_{\text{CC}} \qquad \Longrightarrow \qquad (\hat{\mathcal{H}}_{\text{N}} - \mathcal{E}_{\text{c}})e^{\hat{\mathcal{T}}}\Phi_0 = 0 \tag{2.42}$$

This may be further transformed by multiplication on both sides with $e^{-\hat{\mathcal{T}}}$, simplifying as

$$e^{-\hat{\mathcal{T}}}\hat{\mathcal{H}}_{\text{N}}e^{\hat{\mathcal{T}}}\Phi_0 = e^{-\hat{\mathcal{T}}}\mathcal{E}_{\text{c}}e^{\hat{\mathcal{T}}}\Phi_0 \qquad \Longrightarrow \qquad e^{-\hat{\mathcal{T}}}\hat{\mathcal{H}}_{\text{N}}e^{\hat{\mathcal{T}}}\Phi_0 = \mathcal{E}_{\text{c}}\Phi_0, \tag{2.43}$$

which can be then projected onto the HF bra–state to yield an expression for the correlation energy expectation value

$$\langle\Phi_0|e^{-\hat{\mathcal{T}}}\hat{\mathcal{H}}_{\text{N}}e^{\hat{\mathcal{T}}}|\Phi_0\rangle = \mathcal{E}_{\text{c}}\langle\Phi_0|\Phi_0\rangle = \mathcal{E}_{\text{c}}. \tag{2.44}$$

A useful feature of the similarity–transformed normal ordered Hamiltonian $e^{-\hat{\mathcal{T}}}\hat{\mathcal{H}}_{\text{N}}e^{\hat{\mathcal{T}}}$ is that it may be represented by the Baker–Campbell–Hausdorff (BCH) commutator expansion

$$e^{-\hat{\mathcal{T}}}\hat{\mathcal{H}}_{\text{N}}e^{\hat{\mathcal{T}}} = \hat{\mathcal{H}}_{\text{N}} + [\hat{\mathcal{H}}_{\text{N}}, \hat{\mathcal{T}}] + \frac{1}{2}[[\hat{\mathcal{H}}_{\text{N}}, \hat{\mathcal{T}}], \hat{\mathcal{T}}] + \frac{1}{6}[[[\hat{\mathcal{H}}_{\text{N}}, \hat{\mathcal{T}}], \hat{\mathcal{T}}], \hat{\mathcal{T}}] + \frac{1}{24}[[[[\hat{\mathcal{H}}_{\text{N}}, \hat{\mathcal{T}}], \hat{\mathcal{T}}], \hat{\mathcal{T}}], \hat{\mathcal{T}}] \tag{2.45}$$

which terminates at the four–fold commutator level due to the fact that the Hamiltonian contains only one and two–electron operators and can only be contracted with up to four $\hat{\mathcal{T}}$ terms.[v] This formulation provides a convenient

[iv] A normal–ordered operator in the second–quantisation is one in which all annihilation operators are to the right of the creation operators. The key feature of normal–ordered operators is that their expectation value over a completely unoccupied state is zero.

[v] In the second–quantised form, the normal–ordered Hamiltonian is given by

$$\hat{\mathcal{H}}_{\text{N}} = \sum_{pq} f_{pq}\{\hat{p}^\dagger\,\hat{q}\} + \frac{1}{4}\sum_{pqrs}\langle pq||rs\rangle\,\{\hat{p}^\dagger\,\hat{q}^\dagger\,\hat{s}\,\hat{r}\}$$

where f_{pq} are the matrix elements of the Fock operator $f_{pq} = \langle p|\hat{h}|q\rangle + \sum_i \langle pi||qi\rangle$ and $\{\cdot\}$ indicates that the operators are normal–ordered. In the BCH expansion, products of $\hat{\mathcal{T}}$ & $\hat{\mathcal{H}}_{\text{N}}$ only yield nonzero terms where at least one creation/annihilation operator in $\hat{\mathcal{T}}$ is contracted with an operator in $\hat{\mathcal{H}}_{\text{N}}$; in other words the second–quantised form of $\hat{\mathcal{H}}_{\text{N}}$ & $\hat{\mathcal{T}}$ are *connected* by common indices. Evidently, the normal–ordered Hamiltonian can only be connected to a maximum of four $\hat{\mathcal{T}}$ terms, hence the BCH expansion terminates at fourth–order.

means by which the amplitude equations can be obtained; in the same way as in (2.44) for the correlation energy, (2.43) may be projected against excited–state determinants to yield the series of simultaneous equations

$$\langle \Phi_i^a | e^{-\hat{T}} \hat{\mathcal{H}}_{\mathrm{N}} e^{\hat{T}} | \Phi_0 \rangle = 0 \tag{2.46}$$

$$\left\langle \Phi_{ij}^{ab} \right| e^{-\hat{T}} \hat{\mathcal{H}}_{\mathrm{N}} e^{\hat{T}} \left| \Phi_0 \right\rangle = 0 \tag{2.47}$$

$$\vdots$$

$$\left\langle \Phi_{ij\ldots}^{ab\ldots} \right| e^{-\hat{T}} \hat{\mathcal{H}}_{\mathrm{N}} e^{\hat{T}} \left| \Phi_0 \right\rangle = 0 \tag{2.48}$$

which can be expanded using the principles of the second–quantisation to produce non–linear equations for the cluster amplitudes.

The CCSD approximation is extensively used in this work as an accurate reference method; an expression for the CCSD correlation energy can be derived from substitution of the CCSD cluster operator $\hat{T} = \hat{T}_1 + \hat{T}_2$ into (2.44) as

$$\mathcal{E}_{\mathrm{c}}^{\mathrm{CCSD}} = \sum_{ia} t_i^a f_{ia} + \frac{1}{4} \sum_{ijab} \tau_{ij}^{ab} \langle ij||ab \rangle \tag{2.49}$$

in which τ is an intermediate comprised of the single and double–excitation amplitudes,

$$\tau_{ij}^{ab} = t_{ij}^{ab} + \mathcal{P}_{(ab)} t_i^a t_j^b \tag{2.50}$$

and $\mathcal{P}_{(pq)}$ is an 'antisymmetrisation' operator, transforming terms as $\mathcal{P}_{(pq)} X(p,q) = X(p,q) - X(q,p)$.

In order to determine the cluster amplitudes required to evaluate (2.50), it is necessary to solve (2.46) & (2.47) with the CCSD cluster operator. Considering first the projection against singly–excited determinants in (2.46), a non–linear equation can be elucidated for the single–excitation amplitudes as

$$\begin{aligned} t_i^a \Delta_i^a = f_{ai} &+ \sum_e t_i^e \mathcal{F}_{ae} - \sum_m t_m^a \mathcal{F}_{mi} + \sum_{me} t_{im}^{ae} \mathcal{F}_{me} \\ &- \sum_{me} t_m^e \langle ma||ie \rangle + \frac{1}{2} \sum_{mef} t_{im}^{ef} \langle am||ef \rangle - \frac{1}{2} \sum_{mne} t_{mn}^{ae} \langle mn||ie \rangle \,, \end{aligned} \tag{2.51}$$

where $\Delta_i^a = f_{ii} - f_{aa}$ and $\mathcal{F}_{pq}$ are one–electron intermediate terms given in (2.53). Likewise, projection against doubly–excited determinants in (2.47) results in an equation for the double–excitation cluster amplitudes,

$$\begin{aligned} t_{ij}^{ab} \Delta_{ij}^{ab} = \langle ij||ab \rangle &+ \mathcal{P}_{(ab)} \left\{ \sum_e t_{ij}^{ae} \left(\mathcal{F}_{be} - \frac{1}{2} \sum_m t_m^b \mathcal{F}_{me} \right) - \sum_m t_m^a \mathcal{W}_{mbij} \right\} \\ &- \mathcal{P}_{(ij)} \left\{ \sum_m t_{im}^{ab} \left(\mathcal{F}_{mj} + \frac{1}{2} \sum_e t_j^e \mathcal{F}_{me} \right) - \sum_e t_i^e \langle ab||ej \rangle \right\} + \frac{1}{2} \sum_{mn} \mathcal{W}_{mnij} \tau_{mn}^{ab} \\ &+ \frac{1}{2} \sum_{ef} \tau_{ij}^{ef} \langle ef||ab \rangle + \mathcal{P}_{(ij)} \mathcal{P}_{(ab)} \left\{ \sum_{me} t_{im}^{ae} \mathcal{W}_{mbej} + \sum_m \left(\sum_e \langle mb||ej \rangle t_i^e \right) t_m^a \right\} , \end{aligned} \tag{2.52}$$

where $\Delta_{ij}^{ab} = f_{ii} + f_{jj} - f_{aa} - f_{bb}$ and $\mathcal{W}_{pqrs}$ are two–electron intermediates given in (2.54).

The amplitude equations are factorised using one and two–electron intermediates, in a similar way to those given by Stanton *et. al.* in Ref. 37, as this significantly simplifies the problem and results in practical implementations that are of superior efficiency to the fully expanded form. The one–electron intermediates are constructed as

$$\begin{aligned} \mathcal{F}_{me} &= f_{me} + \sum_{nf} t_n^f \langle mn||ef \rangle \\ \mathcal{F}_{ae} &= (1 - \delta_{ae}) f_{ae} - \frac{1}{2} \sum_m f_{me} t_m^a + \sum_{mf} t_m^f \langle ma||fe \rangle - \frac{1}{2} \sum_{mnf} \tilde{\tau}_{mn}^{af} \langle mn||ef \rangle \\ \mathcal{F}_{mi} &= (1 - \delta_{mi}) f_{mi} + \frac{1}{2} \sum_e f_{me} t_i^e + \sum_{ne} t_n^e \langle mn||ie \rangle + \frac{1}{2} \sum_{nef} \tilde{\tau}_{in}^{ef} \langle mn||ef \rangle \,, \end{aligned} \tag{2.53}$$

where $\tilde{\tau}$ is a modified form of (2.50) given by $\tilde{\tau}_{ij}^{ab} = t_{ij}^{ab} + \frac{1}{2} \mathcal{P}_{(ab)} t_i^a t_j^b$. Similarly, the two–electron intermediates are

defined as

$$\begin{aligned}
\mathcal{W}_{mbij} &= \langle mb||ij\rangle + \frac{1}{2}\sum_{ef} \tau_{ij}^{ef} \langle mb||ef\rangle \\
\mathcal{W}_{mnij} &= \langle mn||ij\rangle + \mathcal{P}_{(ij)} \sum_{e} t_j^e \langle mn||ie\rangle + \frac{1}{2}\sum_{ef} \tau_{ij}^{ef} \langle mn||ef\rangle \\
\mathcal{W}_{mbej} &= \langle mb||ej\rangle + \sum_{f} t_j^f \langle mb||ef\rangle - \sum_{n} t_n^b \langle mn||ej\rangle - \sum_{nf} \left(\frac{1}{2} t_{jn}^{fb} + t_j^f t_n^b\right) \langle mn||ef\rangle ,
\end{aligned} \tag{2.54}$$

thus providing the complete set of intermediate terms required in (2.51) & (2.52). The CC amplitude equations are solved iteratively, with initial values of the cluster amplitudes given by $t_i^a = 0$ & $t_{ij}^{ab} = \langle ij||ab\rangle / \Delta_{ij}^{ab}$[vi], terminating when the change in correlation energy (2.49) between iterations is below a required threshold.

As with CI, the quality of the predictions made by CC increases with the number of excitations considered, with the inclusion of triple–excitations in the cluster amplitude constituting the next increment in accuracy above CCSD. However, the inclusion of higher order excitations in the model leads to a steeply–scaling computational cost that rapidly leads to practical calculations becoming unfeasible to execute.

One popular technique for improving the accuracy of CCSD whilst minimising the increase in cost is the CCSD(T) method[45,46]. In this scheme, a full CC treatment of single and double excitations is followed by a perturbative treatment of triple excitations, which is often found to give similar results to the full triples calculation CCSDT, but with significantly lower computational cost. The success of this approach is such that it is often referred to as the 'gold standard' of quantum chemical calculations.

2.5 Analytical Derivatives

For many chemical applications, calculation of the ground–state energy can be of limited utility as it does not correspond directly to an experimentally observable chemical property; it is essential that electronic structure methods can be used to compute *molecular properties*.

In general, molecular properties can be defined for a system in a given electronic state as its *response* to a perturbation, in which the energy is represented as a series expansion in the perturbation with molecular properties given by the expansion coefficients. For example, given a ground–state energy $\mathcal{E}^{(0)}$, the energy of a molecular system subject to perturbation γ is given by

$$\mathcal{E}(\gamma) = \mathcal{E}^{(0)} + \mathcal{E}^{(1)}\gamma + \frac{1}{2}\mathcal{E}^{(2)}\gamma^2 + \ldots , \tag{2.55}$$

where $\mathcal{E}^{(1)}$ & $\mathcal{E}^{(2)}$ denote first–order and second–order molecular properties respectively. For a time–independent perturbation, these molecular properties may be computed as derivatives of the energy with respect to the perturbation,

$$\mathcal{E}^{(1)} = \left.\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma}\right|_{\gamma=0} \qquad \mathcal{E}^{(2)} = \left.\frac{\mathrm{d}^2\mathcal{E}}{\mathrm{d}\gamma^2}\right|_{\gamma=0} , \tag{2.56}$$

where first–order properties require the gradient of the energy with respect to the perturbation whilst second–order properties relate to the second derivative of the energy with respect to the perturbation.

For a given quantum chemical method, derivatives of the energy can be computed in one of two ways: by numerical approximation or by analytical evaluation. Numerical approximation by finite difference is the most straightforward approach as this only requires calculation of the energy of the system over a range of γ. However, finite difference is of limited accuracy and the requirement to evaluate the energy over many values of γ can restrict the applicability of approach due to the computational demand this entails.

The analytical method overcomes this limitation by directly evaluating the derivative of the energy itself rather than the energy at different values of the perturbation γ. The main drawback is that it requires the derivation and implementation of a new set of expressions for the analytical derivatives for each quantum chemical method. However, due to its accuracy and wider applicability, the analytical method is generally preferable to the numerical method of computing molecular derivatives.

[vi] The initial value of the double–excitation cluster amplitude t_{ij}^{ab} can be recognised as part of the MP2 energy expression (2.32), which may be re–written in terms of t_{ij}^{ab} as $\mathcal{E}_c^{\mathrm{MP2}} = 1/4 \sum_{ijab} t_{ij}^{ab} \langle ij||ab\rangle$.

For a general unperturbed system, the expectation value of the energy can be expressed as a functional of the wavefunction parameters pertaining to the system - for example, the expectation value of the HF energy can be expressed as a functional of the molecular orbital coefficients, as shown in (1.110).

For a general system subject to perturbation γ, the expectation value of the energy also becomes dependent on the perturbation, as demonstrated in (2.55). However, in addition to this explicit dependence on γ, there is also an implicit dependence as the wavefunction parameters themselves are dependent on γ in a general system. Therefore, given a set of perturbation dependent wavefunction parameters $\boldsymbol{\kappa}(\gamma)$, the expression for the first derivative of the energy is determined by the chain rule as

$$\mathcal{E} = \mathcal{E}(\gamma, \boldsymbol{\kappa}(\gamma)) \qquad \Longrightarrow \qquad \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma} = \left(\frac{\partial \mathcal{E}}{\partial \gamma}\right) + \left(\frac{\partial \mathcal{E}}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial \boldsymbol{\kappa}}{\partial \gamma}\right), \tag{2.57}$$

where the term $\partial\boldsymbol{\kappa}/\partial\gamma$ is referred to as the *response* of the wavefunction to the perturbation. If the wavefunction is variationally determined, such as the HF wavefunction, the energy is stationary with respect to changes in the wavefunction parameters thus $\partial\mathcal{E}/\partial\gamma = 0$ for all $\boldsymbol{\kappa}$. As such, the derivative of the energy with respect to γ simplifies as $\mathrm{d}\mathcal{E}/\mathrm{d}\gamma = \partial\mathcal{E}/\partial\gamma$. This is not the case however for non–variational wavefunctions such as the CCSD wavefunction; the response of the wavefunction parameters to the perturbation must also be computed.

2.5.1 Derivatives In Hartree–Fock Theory

The HF wavefunction is variationally determined as described in subsection 1.2.3 and its energy expectation value can be evaluated in the general LCAO form as

$$\mathcal{E}(\gamma, \mathbf{c}(\gamma)) = \sum_{\mu\nu}\sum_{i} c^*_{\mu i} h_{\mu\nu} c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i} c_{\nu i} c^*_{\lambda j} c_{\sigma j} \langle \mu\lambda || \nu\sigma \rangle, \tag{2.58}$$

where γ is an external perturbation parameter, $\mathbf{c}$ the MO coefficients and all other terms as defined in subsubsection 1.2.3.5. As the expectation value of the energy is determined variationally, it is stationary with respect to variations in the MO coefficients. Therefore, the derivative of (2.58) with respect to γ can be written as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma} = &\sum_{\mu\nu}\sum_{i} c^*_{\mu i}\frac{\partial h_{\mu\nu}}{\partial\gamma} c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i} c_{\nu i} c^*_{\lambda j} c_{\sigma j}\frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} + \sum_{\mu\nu}\sum_{i}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma} h_{\mu\nu} c_{\nu i} + c^*_{\mu i} h_{\mu\nu}\frac{\partial c_{\nu i}}{\partial\gamma}\right\} \\ &+ \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma} c_{\nu i} c^*_{\lambda j} c_{\sigma j} + c^*_{\mu i}\frac{\partial c_{\nu i}}{\partial\gamma} c^*_{\lambda j} c_{\sigma j} + c^*_{\mu i} c_{\nu i}\frac{\partial c^*_{\lambda j}}{\partial\gamma} c_{\sigma j} + c^*_{\mu i} c_{\nu i} c^*_{\lambda j}\frac{\partial c_{\sigma j}}{\partial\gamma}\right\}\langle \mu\lambda || \nu\sigma \rangle,\end{aligned} \tag{2.59}$$

which may be further simplified as shown by Pople *et. al.*[47] as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma} &= \sum_{\mu\nu}\sum_{i} c^*_{\mu i}\frac{\partial h_{\mu\nu}}{\partial\gamma} c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i} c_{\nu i} c^*_{\lambda j} c_{\sigma j}\frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} \\ &+ \sum_{\mu\nu}\sum_{i}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma} c_{\nu i} + c^*_{\mu i}\frac{\partial c_{\nu i}}{\partial\gamma}\right\}\left\{h_{\mu\nu} + \sum_{\lambda\sigma}\sum_{j} c^*_{\lambda j} c_{\sigma j}\langle \mu\lambda || \nu\sigma \rangle\right\} \\ &= \sum_{\mu\nu}\sum_{i} c^*_{\mu i}\frac{\partial h_{\mu\nu}}{\partial\gamma} c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i} c_{\nu i} c^*_{\lambda j} c_{\sigma j}\frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} + \sum_{\mu\nu}\sum_{i}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma} F_{\mu\nu} c_{\nu i} + c^*_{\mu i} F_{\mu\nu}\frac{\partial c_{\nu i}}{\partial\gamma}\right\} \\ &= \sum_{\mu\nu}\sum_{i} c^*_{\mu i}\frac{\partial h_{\mu\nu}}{\partial\gamma} c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i} c_{\nu i} c^*_{\lambda j} c_{\sigma j}\frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} + \sum_{\mu\nu}\sum_{i}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma}\varepsilon_i S_{\mu\nu} c_{\nu i} + c^*_{\mu i}\varepsilon_i S_{\mu\nu}\frac{\partial c_{\nu i}}{\partial\gamma}\right\}.\end{aligned} \tag{2.60}$$

Given that the MOs in a variationally–determined HF wavefunction must constitute an orthonormal basis, they must satisfy the identity

$$\sum_{\mu\nu} c^*_{\mu i} S_{\mu\nu} c_{\nu i} = \delta_{ij}, \tag{2.61}$$

the derivative of which with resect to the perturbation γ is given by

$$\sum_{\mu\nu}\left\{\frac{\partial c^*_{\mu i}}{\partial\gamma} S_{\mu\nu} c_{\nu i} + c^*_{\mu i}\frac{\partial S_{\mu\nu}}{\partial\gamma} c_{\nu i} + c^*_{\mu i} S_{\mu\nu}\frac{\partial c_{\nu i}}{\partial\gamma}\right\} = 0. \tag{2.62}$$

This relation may be substituted into the final term in (2.60) to eliminate the derivatives of the MO coefficients, yielding the final expression for the first derivative of the HF energy in the AO basis

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma} = \sum_{\mu\nu}\sum_{i} c^*_{\mu i}\frac{\partial h_{\mu\nu}}{\partial\gamma}c_{\nu i} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}\sum_{ij} c^*_{\mu i}c_{\nu i}c^*_{\lambda j}c_{\sigma j}\frac{\partial\left\langle\mu\lambda||\nu\sigma\right\rangle}{\partial\gamma} - \sum_{\mu\nu}\sum_{i}\varepsilon_i c^*_{\mu i}\frac{\partial S_{\mu\nu}}{\partial\gamma}c_{\nu i}, \tag{2.63}$$

where the derivatives of the overlap matrix, core Hamiltonian matrix and two–electron integrals are computed as appropriate.

2.5.2 Derivatives In Coupled–Cluster Theory

The procedure for computing the analytical derivatives of correlated wavefunctions is somewhat more complex than that for the HF wavefunction as there is an additional set of parameters that must be considered, such as the expansion coefficients in CI or the excitation amplitudes in CC. This is further complicated by the fact that, for non–variational wavefunctions such as CC, the energy expectation value is not stationary with respect to these additional parameters thus it would appear that their derivatives would be required according to (2.57).

It is however possible to formulate a method for evaluating the derivative of the CC energy that does not require derivatives of the excitation amplitudes by instead constructing the *CC Lagrangian*[38,48],

$$\mathcal{L}_{\mathrm{CC}} = \langle\Phi_0|e^{-\hat{T}}\hat{\mathcal{H}}_{\mathrm{N}}e^{\hat{T}}|\Phi_0\rangle + \sum_{I}\lambda_I\langle\Phi_I|e^{-\hat{T}}\hat{\mathcal{H}}_{\mathrm{N}}e^{\hat{T}}|\Phi_0\rangle, \tag{2.64}$$

where the first term is the correlation energy expression from (2.44) and the second term is a summation of the amplitude equations in (2.48) multiplied by a set of Lagrange multipliers $\boldsymbol{\lambda}$. Recognising that the similarity–transformed normal ordered Hamiltonian is not Hermitian, the CC Lagrangian (2.64) may be re–written as

$$\mathcal{L}_{\mathrm{CC}} = \langle\Phi_0|(1+\hat{\Lambda})e^{-\hat{T}}\hat{\mathcal{H}}_{\mathrm{N}}e^{\hat{T}}|\Phi_0\rangle \qquad \text{where} \qquad \langle\Phi_0|\hat{\Lambda} = \sum_{I}\lambda_I\langle\Phi_I|, \tag{2.65}$$

hence the operator $\hat{\Lambda}$ is a *de–excitation operator* defined by analogy to the cluster operators in (2.35) as[36,49]

$$\hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2 + \ldots = \sum_{i}\sum_{a}\lambda^i_a\,\hat{i}^\dagger\,\hat{a} + \frac{1}{4}\sum_{ij}\sum_{ab}\lambda^{ij}_{ab}\,\hat{i}^\dagger\,\hat{a}\,\hat{j}^\dagger\,\hat{b} + \ldots. \tag{2.66}$$

However, in contrast to the cluster operator, the $\hat{\Lambda}$ operators only appear linearly in (2.65) and in that sense share some similarities with the 'CI operators' in (2.39).

It can be readily deduced that the expectation value of the correlation energy is independent of the Lagrange multipliers $\boldsymbol{\lambda}$ by considering the stationary point of the Lagrangian with respect to $\boldsymbol{\lambda}$,

$$\frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial\lambda_I} = \langle\Phi_I|e^{-\hat{T}}\hat{\mathcal{H}}_{\mathrm{N}}e^{\hat{T}}|\Phi_0\rangle = 0, \tag{2.67}$$

which yields the CC amplitude equations of (2.48). Therefore, the excitation amplitudes $\mathbf{t}$ and thus the energy expectation value are independent of $\boldsymbol{\lambda}$. It follows that the Lagrange multipliers $\boldsymbol{\lambda}$ are defined as those that yield a Lagrangian stationary with respect to the excitation amplitudes $\mathbf{t}$, obtained by solving the CC λ*–equations*

$$\frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial t_I} = \langle\Phi_0|(1+\hat{\Lambda})(e^{-\hat{T}}\hat{\mathcal{H}}_{\mathrm{N}}e^{\hat{T}} - \mathcal{E}_{\mathrm{c}})|\Phi_I\rangle = 0. \tag{2.68}$$

Considering (2.68) for the CCSD model, in which $\hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2$, algebraic forms of the λ–equations can be derived in a similar way to the equations for the excitation amplitudes in subsection 2.4.1. For the single de–excitation amplitudes λ^a_i, the λ–equations take the form

$$\begin{aligned}\lambda^a_i\Delta^a_i = \mathcal{F}_{ia} &+ \sum_{e}\lambda^e_i\mathcal{F}_{ea} - \sum_{m}\lambda^a_m\mathcal{F}_{im} + \sum_{me}\lambda^e_m\mathcal{W}_{ieam} + \frac{1}{2}\sum_{mef}\lambda^{ef}_{im}\mathcal{W}_{efam} \\ &- \frac{1}{2}\sum_{mne}\lambda^{ae}_{mn}\mathcal{W}_{iemn} - \sum_{ef}\mathcal{G}_{ef}\mathcal{W}_{eifa} - \sum_{mn}\mathcal{G}_{mn}\mathcal{W}_{mina},\end{aligned} \tag{2.69}$$

whilst the corresponding equations for the double de–excitation amplitudes λ_{ij}^{ab} are given by

$$\begin{aligned}\lambda_{ij}^{ab}\Delta_{ij}^{ab} &= \langle ij||ab\rangle + \frac{1}{2}\sum_{mn}\lambda_{mn}^{ab}\mathcal{W}_{ijmn} + \frac{1}{2}\sum_{ef}\lambda_{ij}^{ef}\langle ab||ef\rangle + \frac{1}{2}\sum_{mn}\mathcal{V}_{ijmn}\langle mn||ab\rangle \\ &+ \mathcal{P}_{(ab)}\left\{\sum_{e}\lambda_{ij}^{ae}\mathcal{F}_{eb} + \sum_{e}\mathcal{G}_{be}\langle ij||ae\rangle - \frac{1}{2}\sum_{me}\left(\sum_{f}\lambda_{ij}^{fe}t_{m}^{f}\right)\langle me||ab\rangle + \sum_{m}\lambda_{m}^{b}\mathcal{W}_{ijma}\right\} \\ &+ \mathcal{P}_{(ij)}\left\{\sum_{m}\lambda_{jm}^{ab}\mathcal{F}_{im} + \sum_{e}\lambda_{i}^{e}\mathcal{W}_{ejab} + \sum_{m}\mathcal{G}_{mi}\langle jm||ab\rangle\right\} + \mathcal{P}_{(ij)}\mathcal{P}_{(ab)}\left\{\sum_{me}\lambda_{im}^{ae}\mathcal{W}_{jebm} + \lambda_{i}^{a}\mathcal{F}_{jb}\right\},\end{aligned} \tag{2.70}$$

where the definitions of Δ_{i}^{a}, Δ_{ij}^{ab}, $\tilde{\tau}_{ij}^{ab}$ and τ_{ij}^{ab} are the same as for the amplitude equations in subsection 2.4.1. As for (2.51) & (2.52), the CCSD λ–equations have been factorised[37,50,51] for computational efficiency, with the one–electron intermediates used in (2.69) & (2.70) defined as

$$\begin{aligned}\mathcal{F}_{me} &= f_{me} + \sum_{nf}t_{n}^{f}\langle mn||ef\rangle \\ \mathcal{F}_{ae} &= (1-\delta_{ae})f_{ae} - \sum_{m}f_{me}t_{m}^{a} + \sum_{mf}t_{m}^{f}\langle ma||fe\rangle - \frac{1}{2}\sum_{mnf}\tilde{\tau}_{mn}^{af}\langle mn||ef\rangle \\ \mathcal{F}_{mi} &= (1-\delta_{mi})f_{mi} + \sum_{e}f_{me}t_{i}^{e} + \sum_{ne}t_{n}^{e}\langle mn||ie\rangle + \frac{1}{2}\sum_{nef}\tilde{\tau}_{in}^{ef}\langle mn||ef\rangle\end{aligned} \tag{2.71}$$

and the two–electron intermediates given by

$$\begin{aligned}\mathcal{W}_{mnie} &= \langle mn||ie\rangle + \sum_{f}t_{i}^{f}\langle mn||fe\rangle \\ \mathcal{W}_{amef} &= \langle am||ef\rangle - \sum_{n}t_{n}^{a}\langle nm||ef\rangle \\ \mathcal{W}_{mnij} &= \langle mn||ij\rangle + \mathcal{P}_{(ij)}\sum_{e}t_{j}^{e}\langle mn||ie\rangle + \frac{1}{2}\sum_{ef}\tau_{ij}^{ef}\langle mn||ef\rangle \\ \mathcal{W}_{mbej} &= \langle mb||ej\rangle + \sum_{f}t_{j}^{f}\langle mb||ef\rangle - \sum_{n}t_{n}^{b}\langle mn||ej\rangle - \sum_{nf}\left(t_{jn}^{fb} + t_{j}^{f}t_{n}^{b}\right)\langle mn||ef\rangle \\ \mathcal{W}_{mbij} &= \langle mb||ij\rangle - \sum_{e}t_{ij}^{be}\mathcal{F}_{me} - \sum_{n}t_{n}^{b}\mathcal{W}_{mnij} + \frac{1}{2}\sum_{ef}\tau_{ij}^{ef}\langle mb||ef\rangle \\ &+ \mathcal{P}_{(ij)}\left\{\sum_{ne}t_{jn}^{be}\langle mn||ie\rangle - \sum_{e}\left(\langle mb||ei\rangle - \sum_{nf}t_{ni}^{bf}\right)t_{j}^{e}\right\} \\ \mathcal{W}_{abei} &= \langle ab||ei\rangle - \sum_{m}t_{mi}^{ab}\mathcal{F}_{me} + \sum_{f}t_{i}^{f}\langle ab||ef\rangle + \frac{1}{2}\sum_{mn}\tau_{mn}^{ab}\left(\langle mn||ei\rangle + \sum_{f}\langle mn||ef\rangle t_{i}^{f}\right) \\ &+ \mathcal{P}_{(ab)}\left\{\sum_{mf}\langle mb||ef\rangle t_{mi}^{af} + \sum_{m}\left(\langle mb||ei\rangle + \sum_{f}\langle mb||ef\rangle t_{i}^{f} - \sum_{nf}\langle mn||ef\rangle t_{ni}^{bf}\right)t_{m}^{a}\right\}.\end{aligned} \tag{2.72}$$

As these intermediates are independent of $\boldsymbol{\lambda}$, they need only be computed once and can be re–used at each stage in the iterative solution of (2.68). In addition to these, (2.69) & (2.70) also involve the set of intermediates

$$\mathcal{G}_{ae} = -\frac{1}{2}\sum_{mnf}\lambda_{mn}^{af}t_{mn}^{ef} \qquad \mathcal{G}_{mi} = \frac{1}{2}\sum_{mef}\lambda_{in}^{ef}t_{mn}^{ef} \qquad \mathcal{V}_{ijmn} = \frac{1}{2}\sum_{ef}\lambda_{ij}^{ef}t_{mn}^{ef} \tag{2.73}$$

however these must be re–computed at each iteration as they are functions of $\boldsymbol{\lambda}$. In practical terms, the λ–equations are solved in the same way as the amplitude equations - by a process of iterative optimisation until convergence of the norm of $\boldsymbol{\lambda}$ to within some threshold.

Having solved (2.67) & (2.68) to yield a set of converged $\mathbf{t}$ and $\boldsymbol{\lambda}$ amplitudes respectively, the Lagrangian is stationary with respect to these two wavefunction parameters. It can be seen that this eliminates derivatives of the $\mathbf{t}$ and $\boldsymbol{\lambda}$ amplitudes from the first derivative of the energy by substituting the CC Lagrangian into (2.57) as

$$\mathcal{L}_{\mathrm{CC}} = \mathcal{L}_{\mathrm{CC}}(\gamma, \mathbf{t}(\gamma), \boldsymbol{\lambda}(\gamma)) \quad \Longrightarrow \quad \frac{\mathrm{d}\mathcal{L}_{\mathrm{CC}}}{\mathrm{d}\gamma} = \left(\frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial\gamma}\right) + \underbrace{\left(\frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial\mathbf{t}}\right)}_{=0}\left(\frac{\partial\mathbf{t}}{\partial\gamma}\right) + \underbrace{\left(\frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial\boldsymbol{\lambda}}\right)}_{=0}\left(\frac{\partial\boldsymbol{\lambda}}{\partial\gamma}\right) = \frac{\partial\mathcal{L}_{\mathrm{CC}}}{\partial\gamma} \tag{2.74}$$

and, noting that at the stationary point of the Lagrangian it is identical to the energy expectation value[38,48], substituting (2.65) into the result of (2.74) to yield the expression

$$\frac{\mathrm{d}\mathcal{E}_{\mathrm{c}}}{\mathrm{d}\gamma} = \frac{\mathrm{d}\mathcal{L}_{\mathrm{CC}}}{\mathrm{d}\gamma} = \frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \gamma} = \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} \frac{\partial \hat{\mathcal{H}}_{\mathrm{N}}}{\partial \gamma} e^{\hat{T}} | \Phi_0 \rangle \,. \tag{2.75}$$

This expression can be re–cast in a convenient form by substituting in the second–quantised form of the normal–ordered Hamiltonian [v] and resolving into a one and two–particle components as[52,53]

$$\begin{aligned} \frac{\mathrm{d}\mathcal{E}_{\mathrm{c}}}{\mathrm{d}\gamma} &= \sum_{pq} \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} \{ \hat{p}^\dagger \, \hat{q} \} e^{\hat{T}} | \Phi_0 \rangle \frac{\partial f_{pq}}{\partial \gamma} + \frac{1}{4} \sum_{pqrs} \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} \{ \hat{p}^\dagger \, \hat{q}^\dagger \, \hat{s} \, \hat{r} \} e^{\hat{T}} | \Phi_0 \rangle \frac{\partial \langle pq || rs \rangle}{\partial \gamma} \\ &= \sum_{pq} D^{\mathrm{CC}}_{pq} \frac{\partial f_{pq}}{\partial \gamma} + \frac{1}{4} \sum_{pqrs} \Gamma^{\mathrm{CC}}_{pqrs} \frac{\partial \langle pq || rs \rangle}{\partial \gamma}, \end{aligned} \tag{2.76}$$

where D^{CC}_{pq} and $\Gamma^{\mathrm{CC}}_{pqrs}$ are the correlation components of the one and two–particle response densities respectively[54–56].

Algebraic expressions for the non–zero blocks of D^{CC}_{pq} and $\Gamma^{\mathrm{CC}}_{pqrs}$ have been developed from their second–quantised forms[50,57], allowing their computation from the converged $\mathbf{t}$ and $\boldsymbol{\lambda}$ amplitudes. The occupied–occupied and virtual–virtual blocks of the one–particle term are given by

$$D^{\mathrm{CC}}_{ij} = -\sum_{e} \lambda^{e}_{j} t^{e}_{i} - \frac{1}{2} \sum_{mef} \lambda^{ef}_{jm} t^{ef}_{im} \qquad D^{\mathrm{CC}}_{ab} = \sum_{m} \lambda^{a}_{m} t^{b}_{m} + \frac{1}{2} \sum_{mne} \lambda^{ae}_{mn} t^{be}_{mn}, \tag{2.77}$$

whilst the non–zero blocks of the two–particle density are constructed as

$$\begin{aligned} \Gamma^{\mathrm{CC}}_{ijkl} &= \frac{1}{8} \sum_{ef} \lambda^{ef}_{kl} \tau^{ef}_{ij} \qquad \Gamma^{\mathrm{CC}}_{akij} = -\frac{1}{4} \sum_{e} \lambda^{ae}_{ij} t^{e}_{k} \qquad \Gamma^{\mathrm{CC}}_{abij} = \frac{1}{4} \lambda^{ab}_{ij} \\ \Gamma^{\mathrm{CC}}_{abcd} &= \frac{1}{8} \sum_{mn} \lambda^{ab}_{mn} \tau^{cd}_{mn} \qquad \Gamma^{\mathrm{CC}}_{abci} = \frac{1}{4} \sum_{m} \lambda^{ab}_{mi} t^{c}_{m} \qquad \Gamma^{\mathrm{CC}}_{ajib} = \frac{1}{4} \sum_{me} \lambda^{ae}_{im} \left(t^{eb}_{mj} - t^{e}_{j} t^{b}_{m} \right) + \frac{1}{4} \lambda^{a}_{i} t^{b}_{j} \\ \Gamma^{\mathrm{CC}}_{ciab} &= \frac{1}{4} \sum_{m} \lambda^{c}_{m} \tau^{ab}_{mi} - \frac{1}{8} \sum_{mne} \lambda^{ce}_{mn} t^{e}_{i} \tau^{ab}_{mn} - \frac{1}{4} \mathcal{P}_{(ab)} \sum_{mne} \lambda^{ce}_{mn} t^{b}_{m} t^{ae}_{in} + \frac{1}{8} \mathcal{P}_{(ab)} \sum_{mne} \lambda^{ce}_{mn} t^{b}_{i} t^{ae}_{mn} \\ \Gamma^{\mathrm{CC}}_{ijka} &= -\frac{1}{4} \sum_{e} \lambda^{e}_{k} \tau^{ea}_{ij} + \frac{1}{8} \sum_{mef} \lambda^{ef}_{km} t^{a}_{m} \tau^{ef}_{ij} + \frac{1}{4} \mathcal{P}_{(ij)} \sum_{mef} \lambda^{ef}_{mk} t^{f}_{j} t^{ae}_{im} - \frac{1}{8} \mathcal{P}_{(ij)} \sum_{mef} \lambda^{ef}_{km} t^{a}_{j} t^{ef}_{im} \\ \Gamma^{\mathrm{CC}}_{ijab} &= \frac{1}{4} \tau^{ab}_{ij} + \frac{1}{16} \sum_{mn} \sum_{ef} \lambda^{ef}_{mn} \tau^{ef}_{ij} \tau^{ab}_{mn} - \frac{1}{8} \mathcal{P}_{(ij)} \sum_{mn} \sum_{ef} \lambda^{ef}_{mn} t^{ef}_{in} \tau^{ab}_{mj} - \frac{1}{4} \mathcal{P}_{(ij)} \sum_{me} \lambda^{e}_{m} t^{e}_{i} \tau^{ab}_{mj} \\ &\quad - \frac{1}{8} \mathcal{P}_{(ab)} \sum_{mn} \sum_{ef} \lambda^{ef}_{mn} t^{af}_{mn} \tau^{eb}_{ij} - \frac{1}{4} \mathcal{P}_{(ab)} \sum_{me} \lambda^{e}_{m} t^{a}_{m} \tau^{eb}_{ij} - \frac{1}{8} \mathcal{P}_{(ij)} \mathcal{P}_{(ab)} \sum_{mn} \sum_{ef} \lambda^{ef}_{mn} \left(t^{ae}_{mi} + 2 t^{e}_{i} t^{a}_{m} \right) t^{bf}_{jn} \\ &\quad - \frac{1}{4} \mathcal{P}_{(ij)} \mathcal{P}_{(ab)} \sum_{me} \lambda^{e}_{m} \left(t^{ae}_{mi} + 2 t^{e}_{i} t^{a}_{m} \right) t^{b}_{j} + \frac{3}{4} \mathcal{P}_{(ij)} \mathcal{P}_{(ab)} \sum_{me} \lambda^{e}_{m} t^{a}_{i} t^{e}_{j} t^{b}_{m}. \end{aligned} \tag{2.78}$$

As the CC method is defined with respect to a HF reference, the overall electronic gradient of a system requires both the CC derivative term (2.76) and the HF derivative term (2.63). However, for this approach to be correct the CC Lagrangian must also be stationary with respect to the HF wavefunction parameters. This condition can be satisfied through the addition of further terms to (2.65) to form an *extended Lagrangian*,

$$\begin{aligned} \mathcal{L}_{\mathrm{CC}} &= \langle \Phi_0 | (1 + \hat{\Lambda}) e^{-\hat{T}} \hat{\mathcal{H}}_{\mathrm{N}} e^{\hat{T}} | \Phi_0 \rangle + \sum_{ai} Z_{ai} f_{ai} + \sum_{pq} I_{pq} (S_{pq} - \delta_{pq}) \\ &= \sum_{pq} D^{\mathrm{CC}}_{pq} f_{pq} + \frac{1}{4} \sum_{pqrs} \Gamma^{\mathrm{CC}}_{pqrs} \langle pq || rs \rangle + \sum_{ai} Z_{ai} f_{ai} + \sum_{pq} I_{pq} (S_{pq} - \delta_{pq}), \end{aligned} \tag{2.79}$$

in which Z_{ai} are the Lagrange multipliers pertaining to the Brillouin condition (2.5) and I_{pq} are the Lagrange multipliers ensuring orthonormality of the MOs (2.61).

In determining the stationary point of the CC Lagrangian (2.79) with respect to the MO coefficients $\mathbf{c}$, it is convenient to express a general set coefficients as the product of the unperturbed HF coefficients $\mathbf{c}^0$ and an *orbital rotation* matrix $\mathbf{U}$,

$$c_{\mu p} = \sum_{q} c^{0}_{\mu q} U_{qp}. \tag{2.80}$$

The expression for the stationary point of the Lagrangian with respect to $\mathbf{c}$ can thus be re–cast as

$$\left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \mathbf{c}}\right|_{\mathbf{c}=\mathbf{c}^0} = \left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \mathbf{U}}\right|_{\mathbf{U}=\mathbf{1}} = 0, \tag{2.81}$$

for which an explicit form can be elucidated by re–writing the Lagrangian (2.79) in terms of $\mathbf{U}$,

$$\begin{aligned}
\mathcal{L}_{\mathrm{CC}} &= \sum_{pq} D^{\mathrm{CC}}_{pq} \sum_{\mu\nu} c^*_{\mu p} c_{\nu q} \left\{ h_{\mu\nu} + \sum_{\lambda\sigma} \sum_{j} c^*_{\lambda j} c_{\sigma j} \langle \mu\lambda || \nu\sigma \rangle \right\} \frac{1}{4} \sum_{pqrs} \Gamma^{\mathrm{CC}}_{pqrs} \sum_{\mu\nu\lambda\sigma} c^*_{\mu p} c_{\nu r} c^*_{\lambda q} c_{\sigma s} \langle \mu\lambda || \nu\sigma \rangle \\
&\quad + \sum_{ai} Z_{ai} \sum_{\mu\nu} c^*_{\mu a} c_{\nu i} \left\{ h_{\mu\nu} + \sum_{\lambda\sigma} \sum_{j} c^*_{\lambda j} c_{\sigma j} \langle \mu\lambda || \nu\sigma \rangle \right\} + \sum_{pq} I_{pq} \left(\sum_{\mu\nu} c^*_{\mu p} S_{\mu\nu} c_{\nu q} - \delta_{pq} \right) \\
&= \sum_{pq} D^{\mathrm{CC}}_{pq} \sum_{tu} U^*_{tp} U_{uq} \left\{ h_{tu} + \sum_{vw} \sum_{j} U^*_{vj} U_{wj} \langle tv || uw \rangle \right\} + \frac{1}{4} \sum_{pqrs} \Gamma^{\mathrm{CC}}_{pqrs} \sum_{tuvw} U^*_{tp} U_{vr} U^*_{uq} U_{ws} \langle tu || vw \rangle \\
&\quad + \sum_{ai} Z_{ai} \sum_{tu} U^*_{ta} U_{ui} \left\{ h_{tu} + \sum_{vw} \sum_{j} U^*_{vj} U_{wj} \langle tv || uw \rangle \right\} + \sum_{pq} I_{pq} \left(\sum_{tu} U^*_{tp} U_{uq} S_{tu} - \delta_{tu} \right).
\end{aligned} \tag{2.82}$$

The derivative of this expression at the stationary point can then be taken with respect to elements of the rotation matrix as

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial U_{ij}}\right|_{U_{ij}=\delta_{ij}} &= \sum_{q} D^{\mathrm{CC}*}_{jq} f^*_{iq} + \sum_{p} D^{\mathrm{CC}}_{pj} f_{pi} + \sum_{pq} D^{\mathrm{CC}*}_{pq} \langle pi || qj \rangle^* + \sum_{pq} D^{\mathrm{CC}}_{pq} \langle pj || qi \rangle \\
&\quad + \frac{1}{4} \sum_{qrs} \Gamma^{\mathrm{CC}*}_{jqrs} \langle iq || rs \rangle^* + \frac{1}{4} \sum_{prs} \Gamma^{\mathrm{CC}*}_{pjrs} \langle pi || rs \rangle^* + \frac{1}{4} \sum_{pqs} \Gamma^{\mathrm{CC}}_{pqjs} \langle pq || is \rangle + \frac{1}{4} \sum_{pqr} \Gamma^{\mathrm{CC}}_{pqrj} \langle pq || ri \rangle \\
&\quad + \sum_{a} Z_{aj} f_{ai} + \sum_{ak} Z^*_{ak} \langle ai || kj \rangle^* + \sum_{ak} Z_{ak} \langle aj || ki \rangle + \sum_{q} I^*_{jq} S^*_{iq} + \sum_{p} I_{pj} S_{pi} = 0
\end{aligned} \tag{2.83}$$

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial U_{ab}}\right|_{U_{ab}=\delta_{ab}} &= \sum_{q} D^{\mathrm{CC}*}_{bq} f^*_{aq} + \sum_{p} D^{\mathrm{CC}}_{pb} f_{pa} + \frac{1}{4} \sum_{qrs} \Gamma^{\mathrm{CC}*}_{bqrs} \langle aq || rs \rangle^* + \frac{1}{4} \sum_{prs} \Gamma^{\mathrm{CC}*}_{pbrs} \langle pa || rs \rangle^* \\
&\quad + \frac{1}{4} \sum_{pqs} \Gamma^{\mathrm{CC}}_{pqbs} \langle pq || as \rangle + \frac{1}{4} \sum_{pqr} \Gamma^{\mathrm{CC}}_{pqrb} \langle pq || ra \rangle + \sum_{j} Z^*_{bj} f^*_{aj} + \sum_{q} I^*_{bq} S^*_{aq} + \sum_{p} I_{pb} S_{pa} = 0
\end{aligned} \tag{2.84}$$

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial U_{ia}}\right|_{U_{ia}=\delta_{ia}} &= \sum_{q} D^{\mathrm{CC}*}_{aq} f^*_{iq} + \sum_{p} D^{\mathrm{CC}}_{pa} f_{pi} + \frac{1}{4} \sum_{qrs} \Gamma^{\mathrm{CC}*}_{aqrs} \langle iq || rs \rangle^* + \frac{1}{4} \sum_{prs} \Gamma^{\mathrm{CC}*}_{pars} \langle pi || rs \rangle^* \\
&\quad + \frac{1}{4} \sum_{pqs} \Gamma^{\mathrm{CC}}_{pqas} \langle pq || is \rangle + \frac{1}{4} \sum_{pqr} \Gamma^{\mathrm{CC}}_{pqra} \langle pq || ri \rangle + \sum_{j} Z^*_{aj} f^*_{ij} + \sum_{q} I^*_{aq} S^*_{iq} + \sum_{p} I_{pa} S_{pi} = 0
\end{aligned} \tag{2.85}$$

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial U_{ai}}\right|_{U_{ai}=\delta_{ai}} &= \sum_{q} D^{\mathrm{CC}*}_{iq} f^*_{aq} + \sum_{p} D^{\mathrm{CC}}_{pi} f_{pa} + \sum_{pq} D^{\mathrm{CC}*}_{pq} \langle pa || qi \rangle^* + \sum_{pq} D^{\mathrm{CC}}_{pq} \langle pi || qa \rangle \\
&\quad + \frac{1}{4} \sum_{qrs} \Gamma^{\mathrm{CC}*}_{iqrs} \langle aq || rs \rangle^* + \frac{1}{4} \sum_{prs} \Gamma^{\mathrm{CC}*}_{pirs} \langle pa || rs \rangle^* + \frac{1}{4} \sum_{pqs} \Gamma^{\mathrm{CC}}_{pqis} \langle pq || as \rangle + \frac{1}{4} \sum_{pqr} \Gamma^{\mathrm{CC}}_{pqri} \langle pq || ra \rangle \\
&\quad + \sum_{b} Z_{bi} f_{ba} + \sum_{bj} Z^*_{bj} \langle ba || ji \rangle^* + \sum_{bj} Z_{bj} \langle bi || ja \rangle + \sum_{q} I^*_{iq} S^*_{aq} + \sum_{p} I_{pi} S_{pa} = 0.
\end{aligned} \tag{2.86}$$

These expressions may be re-arranged using the identity $S_{pq} = \delta_{pq}$ to give the expressions

$$\begin{aligned}
I_{ij} + I^*_{ji} &= -\sum_{p} \left(D^{\mathrm{CC}*}_{jp} f^*_{ip} + D^{\mathrm{CC}}_{pj} f_{pi} \right) - \sum_{pq} \left(D^{\mathrm{CC}*}_{pq} \langle pi || qj \rangle^* + D^{\mathrm{CC}}_{pq} \langle pj || qi \rangle \right) - \sum_{a} Z_{aj} f_{ai} \\
&\quad - \sum_{ak} \left(Z^*_{ak} \langle ai || kj \rangle^* + Z_{ak} \langle aj || ki \rangle \right) - \frac{1}{2} \sum_{pqr} \left(\Gamma^{\mathrm{CC}*}_{jpqr} \langle ip || qr \rangle^* + \Gamma^{\mathrm{CC}}_{pqjr} \langle pq || ir \rangle \right)
\end{aligned} \tag{2.87}$$

$$I_{ab} + I^*_{ba} = -\sum_{p} \left(D^{\mathrm{CC}*}_{bp} f^*_{ap} + D^{\mathrm{CC}}_{pb} f_{pa} \right) - \sum_{j} Z^*_{bj} f^*_{aj} - \frac{1}{2} \sum_{pqr} \left(\Gamma^{\mathrm{CC}*}_{bpqr} \langle ap || qr \rangle^* + \Gamma^{\mathrm{CC}}_{pqbr} \langle pq || ar \rangle \right) \tag{2.88}$$

$$I_{ia} + I_{ai}^{*} = -\sum_{p} \left(D_{ap}^{\mathrm{CC}*} f_{ip}^{*} + D_{pa}^{\mathrm{CC}} f_{pi} \right) - \sum_{j} Z_{aj}^{*} f_{ij}^{*} - \frac{1}{2} \sum_{pqr} \left(\Gamma_{apqr}^{\mathrm{CC}*} \langle ip||qr \rangle^{*} + \Gamma_{pqar}^{\mathrm{CC}} \langle pq||ir \rangle \right) \tag{2.89}$$

$$\begin{aligned} I_{ai} + I_{ia}^{*} = &- \sum_{p} \left(D_{ip}^{\mathrm{CC}*} f_{ap}^{*} + D_{pi}^{\mathrm{CC}} f_{pa} \right) - \sum_{pq} \left(D_{pq}^{\mathrm{CC}*} \langle pa||qi \rangle^{*} + D_{pq}^{\mathrm{CC}} \langle pi||qa \rangle \right) - \sum_{b} Z_{bi} f_{ba} \\ &- \sum_{bj} \left(Z_{bj}^{*} \langle ba||ji \rangle^{*} + Z_{bj} \langle bi||ja \rangle \right) - \frac{1}{2} \sum_{pqr} \left(\Gamma_{ipqr}^{\mathrm{CC}*} \langle ap||qr \rangle^{*} + \Gamma_{pqir}^{\mathrm{CC}} \langle pq||ar \rangle \right). \end{aligned} \tag{2.90}$$

Recognising that the complex conjugate of (2.89) is equal to (2.90), the former can be subtracted from the latter to eliminate the I_{pq} terms

$$\begin{aligned} I_{ai} + I_{ia}^{*} - I_{ia}^{*} - I_{ai} = &- \sum_{p} \left(D_{ip}^{\mathrm{CC}*} f_{ap}^{*} + D_{pi}^{\mathrm{CC}} f_{pa} \right) - \sum_{pq} \left(D_{pq}^{\mathrm{CC}*} \langle pa||qi \rangle^{*} + D_{pq}^{\mathrm{CC}} \langle pi||qa \rangle \right) - \sum_{b} Z_{bi} f_{ba} \\ &- \sum_{bj} \left(Z_{bj}^{*} \langle ba||ji \rangle^{*} + Z_{bj} \langle bi||ja \rangle \right) - \frac{1}{2} \sum_{pqr} \left(\Gamma_{ipqr}^{\mathrm{CC}*} \langle ap||qr \rangle^{*} + \Gamma_{pqir}^{\mathrm{CC}} \langle pq||ar \rangle \right) \\ &+ \sum_{p} \left(D_{pa}^{\mathrm{CC}*} f_{pi}^{*} + D_{ap}^{\mathrm{CC}} f_{ip} \right) + \sum_{j} Z_{aj} f_{ij} + \frac{1}{2} \sum_{pqr} \left(\Gamma_{pqar}^{\mathrm{CC}*} \langle pq||ir \rangle^{*} + \Gamma_{apqr}^{\mathrm{CC}} \langle ip||qr \rangle \right), \end{aligned} \tag{2.91}$$

which may thus be re–arranged to yield linear equations for Z_{ai}

$$\sum_{bj} \left\{ Z_{bj}^{*} \langle ba||ji \rangle^{*} + Z_{bj} \left(\langle bi||ja \rangle + \delta_{ij} f_{ba} - \delta_{ab} f_{ij} \right) \right\} = \mathcal{X}_{ai}, \tag{2.92}$$

where $\mathcal{X}_{ai}$ is the *orbital rotation gradient*, defined as

$$\begin{aligned} \mathcal{X}_{ai} = &\sum_{p} \left(D_{pa}^{\mathrm{CC}*} f_{pi}^{*} + D_{ap}^{\mathrm{CC}} f_{ip} - D_{ip}^{\mathrm{CC}*} f_{ap}^{*} - D_{pi}^{\mathrm{CC}} f_{pa} \right) + \frac{1}{2} \sum_{pqr} \left(\Gamma_{pqar}^{\mathrm{CC}*} \langle pq||ir \rangle^{*} + \Gamma_{apqr}^{\mathrm{CC}} \langle ip||qr \rangle \right) \\ &- \frac{1}{2} \sum_{pqr} \left(\Gamma_{ipqr}^{\mathrm{CC}*} \langle ap||qr \rangle^{*} + \Gamma_{pqir}^{\mathrm{CC}} \langle pq||ar \rangle \right) - \sum_{pq} \left(D_{pq}^{\mathrm{CC}*} \langle pa||qi \rangle^{*} + D_{pq}^{\mathrm{CC}} \langle pi||qa \rangle \right). \end{aligned} \tag{2.93}$$

The linear equation (2.92) is often known as a *Z–vector equation*; it is a convenient method for computing the orbital relaxation terms in an analytical derivative calculation introduced by Handy & Schaefer[58], based on the interchange theorem of Dalgarno & Stewart[59–61], which permits this term to be computed independently of the external perturbation.

The solution to the Z–vector equation can be used to construct the I_{pq} term in (2.81) according to the expressions (2.87), (2.88), (2.89) & (2.90), with the identity

$$I_{pq} = \frac{1}{2} \left(I_{pq} + I_{qp}^{*} \right), \tag{2.94}$$

whilst the solution vector itself constitutes the occupied-virtual and virtual-occupied blocks of the CC one–particle density[43,50,62],

$$D_{pq}^{\mathrm{CC}} = \begin{pmatrix} D_{ij}^{\mathrm{CC}} & Z_{ia}^{*} \\ Z_{ai} & D_{ab}^{\mathrm{CC}} \end{pmatrix}. \tag{2.95}$$

This yields a CC Lagrangian that is stationary with respect to all relevant wavefunction parameters,

$$\frac{\mathrm{d}\mathcal{L}_{\mathrm{CC}}}{\mathrm{d}\gamma} = \left(\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \gamma} \right) + \underbrace{\left(\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \mathbf{t}} \right)}_{=0} \left(\frac{\partial \mathbf{t}}{\partial \gamma} \right) + \underbrace{\left(\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \boldsymbol{\lambda}} \right)}_{=0} \left(\frac{\partial \boldsymbol{\lambda}}{\partial \gamma} \right) + \underbrace{\left(\frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \mathbf{c}} \right)}_{=0} \left(\frac{\partial \mathbf{c}}{\partial \gamma} \right) = \frac{\partial \mathcal{L}_{\mathrm{CC}}}{\partial \gamma}, \tag{2.96}$$

from which, by analogy to (2.76), the expression for the analytical gradient of the CC correlation energy is derived as

$$\frac{\mathrm{d}\mathcal{E}_{\mathrm{c}}}{\mathrm{d}\gamma} = \sum_{pq} D_{pq}^{\mathrm{CC}} \frac{\partial f_{pq}}{\partial \gamma} + \frac{1}{4} \sum_{pqrs} \Gamma_{pqrs}^{\mathrm{CC}} \frac{\partial \langle pq||rs \rangle}{\partial \gamma} + \sum_{pq} I_{pq} \frac{\partial S_{pq}}{\partial \gamma}. \tag{2.97}$$

The overall analytical gradient is obtained by adding the Hartree–Fock reference to (2.97), which is achieved by replacing the CC correlation density matrices with the effective one and two–particle density matrices[43],

$$\begin{aligned} D_{pq} &= D_{pq}^{\mathrm{CC}} + \delta_{pq}^{occ} \\ \Gamma_{pqrs} &= \Gamma_{pqrs}^{\mathrm{CC}} + \frac{1}{4} \delta_{pr}^{occ} \delta_{qs}^{occ} + \frac{1}{4} \delta_{pr}^{occ} D_{qs}^{\mathrm{CC}} + \frac{1}{4} \delta_{qs}^{occ} D_{pr}^{\mathrm{CC}} - \frac{1}{4} \delta_{ps}^{occ} \delta_{qr}^{occ} - \frac{1}{4} \delta_{ps}^{occ} D_{qr}^{\mathrm{CC}} - \frac{1}{4} \delta_{qr}^{occ} D_{ps}^{\mathrm{CC}}. \end{aligned} \tag{2.98}$$

For ease of computation, the expression for the gradient may be transformed into AO basis and evaluated as

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\gamma} = \sum_{\mu\nu} D_{\mu\nu} \left\{ \frac{\partial h_{\mu\nu}}{\partial\gamma} + \sum_{\lambda\sigma} D^{\mathrm{HF}}_{\lambda\sigma} \frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} \right\} + \frac{1}{4} \sum_{\mu\nu\lambda\sigma} \Gamma_{\mu\nu\lambda\sigma} \frac{\partial \langle \mu\lambda || \nu\sigma \rangle}{\partial\gamma} + \sum_{\mu\nu} I_{\mu\nu} \frac{\partial S_{\mu\nu}}{\partial\gamma}, \tag{2.99}$$

where D_{pq}, Γ_{pqrs} & I_{pq} are transformed into AO basis as

$$D_{\mu\nu} = \sum_{pq} c^*_{\mu p} D_{pq} c_{\nu q} \qquad \Gamma_{\mu\nu\lambda\sigma} = \sum_{pqrs} c^*_{\mu p} c^*_{\nu q} \Gamma_{pqrs} c_{\lambda r} c_{\sigma s} \qquad I_{\mu\nu} = \sum_{pq} c^*_{\mu p} I_{pq} c_{\nu q}. \tag{2.100}$$

2.6 Review

In this Chapter, some of the most well–established *ab initio* methods for treating correlation have been discussed. Although techniques such as CCSD(T) and FCI can be used to obtain highly accurate predictions of many–electron systems, the high computational cost and very poor scaling associated with post-HF methods limit their practical applicability to the study of systems with a few tens of atoms only.

In Chapter 3, density functional theory is introduced; this is a more efficient approach than those discussed in the present Chapter and may be used to model systems with many thousands of atoms. However, its underlying principles are such that it is not clear how systematic improvements can be made to improve its accuracy (as is the case with *ab initio* methods, i.e. by increasing the CI space or including higher–order excitations in CC).

The main scientific purpose of this work is to make use of the most accurate *ab initio* methods available, such as CCSD(T), to construct local models for correlation energy distributions within small systems. For this purpose, many of the methods discussed in this chapter have been implemented for general, spin–unrestricted systems in the quantum chemical rapid development code QUEST[63], in particular for the energies and analytical derivatives of the CCSD method as presented here but also the simpler CCD model (in which single–excitation amplitudes are omitted) and the CCSD(T) model. The relevance of these *ab initio* methods to the investigation of correlation in density functional theory will be fully discussed in Chapter 5.

3 Density Functional Theory

The methods of obtaining approximate solutions to the Schrödinger Equation discussed in the previous chapter have been successfully used to make accurate predictions for a wide range of molecular systems for many decades. However, these *wavefunction* methods share one main disadvantage, which is that their computational cost scales very unfavourably with system size. The table below lists some of these methods and their formal scaling with system-size, N (i.e. number of basis functions, which is dependant on both system size and basis–set size).

Method	Formal Scaling	Maximum System Size
HF	$\mathcal{O}(N^4)$	$\sim$ 100's of atoms
MP2	$\mathcal{O}(N^5)$	$\sim$ 100's of atoms
MP3, CISD, CCSD	$\mathcal{O}(N^6)$	$\sim$ 100 atoms
MP4, CCSD(T)	$\mathcal{O}(N^7)$	$\sim$ 10's of atoms
MP5, CISDT, CCSDT	$\mathcal{O}(N^8)$	$\sim$ 10's of atoms
FCI	$\mathcal{O}(N!)$	$\sim$ 5 atoms

Table 3.1: Computational scaling & system–size limitations for common *ab initio* methods.

As this demonstrates, the computational cost of these methods increases by orders of magnitude with more advanced treatments of correlation. This leads to a limit on the size of the system that may be feasibly modelled using modern technology and software. However, this is also dependent on the basis–set chosen for the expansion of the one–electron orbitals; it is well known that the estimates obtained from post–HF methods converge slowly towards the true value with increasing basis–set size. This means that, where extremely accurate predictions are required, it is necessary to use both a higher–order post–HF method and a very large basis–set, further restricting the number of electrons that can be treated.

One very different approach to electronic–structure calculations is to describe systems entirely in terms of their *electron density* $\rho(\mathbf{r})$ – this is the principle of *Density Functional Theory* (DFT). The immediate advantage of this approach is that the number of variables required to describe a system is reduced from the spatial & spin coordinates of every electron as in the many–electron wavefunction, to simply three spatial coordinates; this would present a much lower computational cost than any of the *ab initio* methods and would have the potential to scale much more favourably with system size.

Initially, it was thought that DFT was not an exact theory, and couldn't be used to exactly solve a quantum mechanical problem, as the only models based on this approach at the time were formulated as approximate methods. This perception was changed entirely in 1964 when Hohenberg & Kohn (HK) published their landmark paper[64] in which they proved that the ground–state of an electronic system in an external potential can be described completely by the one–electron density ρ. A further ground–breaking paper was published a year later by Kohn & Sham (KS)[65], in which they showed that from the ground–state *non–interacting* one particle density it is possible to fully account for the effects of exchange & correlation through addition of an exchange–correlation term, or *xc–functional*, the physical form of which would allow exact solutions to be yielded by DFT calculations.

These developments showed that DFT was, in principle, a method that could give exact solutions to the many–electron problem, but that its accuracy would be determined by the quality of the exchange–correlation term in the KS formalism. In their 1965 paper KS noted that the exact form of the xc–functional was unknown, however they proposed a simple approximation to this based–on the density – this became known as the *local density approximation* (LDA). Although computationally inexpensive, it was found that the KS scheme with the LDA functional was poor when it came to the accuracy of its predictions - often inferior to HF for molecules, although was found to be accurate for many metallic systems.

It wasn't until some years later that approximations forms of the *xc*–functional were derived which made the accuracy

of DFT calculations competitive with those of many post-HF methods, yet far cheaper computationally. Most quantum chemistry software today has the facility to execute DFT calculations within the KS scheme, with one or more of the many hundreds of *xc*–functionals available being used to account for exchange and correlation. However, in the pursuit of higher accuracy with lower cost, there remains the need for improved forms of the *xc*–functional; the aim of the work presented in later chapters is to make progress in this direction. This chapter will outline the main theories on which these investigations are fundamentally based.

3.1 The Rayleigh–Ritz Variation Principle

The problem that DFT endeavours to solve is the same as that of *ab initio* methods; the non–relativistic time independent electronic Schrödinger Equation. The standard Hamiltonian (1.10) may be re–cast in the more general form

$$\begin{aligned}\hat{\mathcal{H}}_\lambda(v) &= -\frac{1}{2}\sum_i \nabla_i^2 + \sum_i v(\mathbf{r}_i) + \sum_{i>j}\frac{\lambda}{|\mathbf{r}_i-\mathbf{r}_j|} \\ &= \hat{T} + \hat{V} + \hat{W}_\lambda,\end{aligned} \tag{3.1}$$

where $\hat{T}$ is the *kinetic energy* operator, $\hat{W}_\lambda$ the *electron–electron repulsion* operator scaled by a interaction–strength coefficient $\lambda \in [0,1]$ such that physical systems are represented at $\lambda = 1$ and $\hat{V}$ a multiplicative *external potential*, typically describing the nuclear potential, but also including contributions from other system-dependent sources where present.[i]

It can be readily seen from (3.1) that the kinetic energy & electron–electron repulsion components of the Hamiltonian will be equivalent for any N–electron system. The only term in the Hamiltonian that is not necessarily universal for any N–electron system is the functional dependence on an external potential $v(\mathbf{r})$. Considering only multiplicative potentials that can bind an N–electron ground state, the set of *ρ–representable potentials* $\mathcal{V}_N$

$$\mathcal{V}_N = \left\{ v \;\middle|\; \hat{\mathcal{H}}_\lambda(v) \text{ has an } N\text{-electron ground state } \Psi_v \right\}, \tag{3.2}$$

it can be shown that there is a ground–state N–electron wavefunction that satisfies the Schrödinger Equation

$$\hat{\mathcal{H}}_\lambda(v)\Psi_v = \mathcal{E}_\lambda(v)\Psi_v \qquad \forall v \in \mathcal{V}_N \tag{3.3}$$

in which $\mathcal{E}_\lambda(v)$ is the ground–state energy at a given λ. The Schrödinger Equation (3.3) thus implies that the ground–state energy may be interpreted as a mapping of the set of potentials $\mathcal{V}_N$ to the set of real numbers $\mathbb{R}$,

$$\mathcal{E}_\lambda : \mathcal{V}_N \mapsto \mathbb{R}. \tag{3.4}$$

Considering only wavefunctions Ψ_v belonging to the set $\mathcal{W}_N$ of normalised, antisymmetric N–electron functions for which the expectation value of $\hat{\mathcal{H}}_\lambda(v)$ is finite,

$$\mathcal{W}_N = \left\{ \Psi_v \;\middle|\; \langle\Psi_v|\Psi_v\rangle = 1 \;;\; \left\langle \Psi_v \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_v \right\rangle < \infty \right\}, \tag{3.5}$$

the ground–state solution to (3.3) is defined by the *Rayleigh–Ritz variation principle*.

Theorem 3.1: Rayleigh–Ritz Variation Principle

For all pairs of potentials and wavefunctions $(v, \Psi) \in \mathcal{V}_N \times \mathcal{W}_N$, the Rayleigh–Ritz inequality is satisfied

$$\mathcal{E}_\lambda(v) \leqslant \left\langle \Psi \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi \right\rangle. \tag{3.6}$$

Furthermore for each $v \in \mathcal{V}_N$, one or more ground–state wavefunctions $\Psi_v \in \mathcal{W}_N$ exist as the global minimisers of the Rayleigh–Ritz variation principle, with ground state energy $\mathcal{E}_\lambda(v)$,

$$\mathcal{E}_\lambda(v) = \left\langle \Psi_v \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_v \right\rangle \;,\; \Psi_v \in \underset{\Psi\in\mathcal{W}_N}{\arg\min} \left\langle \Psi \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi \right\rangle. \tag{3.7}$$

For a system with n–fold degenerate wavefunctions, the global minimiser Ψ_v determined according to Theorem 3.1 can be constructed from a linear combination of the n normalised, degenerate ground–state eigenfunctions.

[i] The scaling of the electron–electron repulsion operator can be used to provide additional insight into DFT and is to be discussed in detail in Section 3.6.

Given that all ρ–representable potentials $v \in \mathcal{V}_N$ have an associated electronic ground state Ψ_v, they must also have an associated ground–state density ρ_v given by

$$\rho_v(\mathbf{r}_1) = N \int_{\mathbb{R}^3_\Sigma} \dots \int_{\mathbb{R}^3_\Sigma} \Psi_v^*(\tau_1, \tau_2, \dots, \tau_N)\Psi_v(\tau_1, \tau_2, \dots, \tau_N)\, \mathrm{d}\sigma_1\, \mathrm{d}\tau_2 \dots\, \mathrm{d}\tau_N \tag{3.8}$$

at a point $\mathbf{r}_1$ in space, where τ & σ are as defined in subsection 1.2.1. With this identity, the expectation value for the ground–state energy in (3.7) can be re–cast in terms of the density as

$$\mathcal{E}_\lambda(v) = \left\langle \Psi_v \left| \hat{T} + \hat{W}_\lambda \right| \Psi_v \right\rangle + \int_{\mathbb{R}^3} v(\mathbf{r})\rho_v(\mathbf{r})\mathrm{d}\mathbf{r} \;=\; \left\langle \Psi_v \left| \hat{\mathcal{H}}_\lambda(0) \right| \Psi_v \right\rangle + (v|\rho_v) \tag{3.9}$$

where the $(v|\rho_v)$ term describes the interaction between the external potential acting on the system and its electron density. The relation between potential and density in DFT was first described by Hohenberg & Kohn for the set of potentials $v \in \mathcal{V}_N$ that can bind a ground–state of an N–electron system[64]. However, subsequent work by Levy[66] & Lieb[67] generalised the theory further effectively removing this limitation.

3.2 The Hohenberg-Kohn Theorems

Theories for modelling many–electron systems using the one–electron density are well established and in the case of the Thomas–Fermi theory[68,69], predate the formulation of the Hartree–Fock method. However, it was not until 1964 that an exact theory for doing so was developed by Hohenberg & Kohn[64], in which it was mathematically shown that the one–electron density is sufficient to determine the ground–state wavefunction of a many–electron system.

The principles set out by Hohenberg & Kohn are generally considered in terms of two fundamental theorems: the *first Hohenberg–Kohn theorem* and the *second Hohenberg–Kohn theorem*.

3.2.1 The First Hohenberg-Kohn Theorem

The first Hohenberg–Kohn theorem states that the external potential v can be uniquely determined to within an additive constant from the ground–state density ρ of the N–electron system. To demonstrate this, it is necessary to consider the effects of shifting the potential by a constant on the eigenstates of the Schrödinger Equation (3.3) and the expectation value of the energy (3.9).

A gauge transformation can be applied to an external potential v by introducing an additive constant c,

$$u = v + c \quad , \quad c \in \mathbb{R}. \tag{3.10}$$

The N–electron Hamiltonian is thus transformed as

$$\hat{\mathcal{H}}_\lambda(u) = \hat{\mathcal{H}}_\lambda(v + c) = \hat{\mathcal{H}}_\lambda(0) + \sum_i (v(\mathbf{r}_i) + c) = \hat{\mathcal{H}}_\lambda(v) + Nc, \tag{3.11}$$

for which the expectation value (3.7) becomes

$$\begin{aligned} \mathcal{E}_\lambda(u) = \mathcal{E}_\lambda(v + c) &= \inf_{\Psi \in \mathcal{W}_N} \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) + Nc \right| \Psi \right\rangle \\ &= \inf_{\Psi \in \mathcal{W}_N} \left\{ \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) \right| \Psi \right\rangle + Nc \right\} = \mathcal{E}_\lambda(v) + Nc. \end{aligned} \tag{3.12}$$

It can be seen therefore that the eigenvalues of the Schrödinger Equation (3.3) are invariant to such a shift in gauge of the potential,

$$\begin{aligned} \hat{\mathcal{H}}_\lambda(u)\Psi = \hat{\mathcal{H}}_\lambda(v + c)\Psi &= \left(\hat{\mathcal{H}}_\lambda(v) + Nc\right)\Psi \\ &= \left(\mathcal{E}_\lambda(v) + Nc\right)\Psi = \mathcal{E}_\lambda(v + c)\Psi. \end{aligned} \tag{3.13}$$

It can furthermore be shown that two different Hamiltonians can only share a ground–state wavefunction if the external potentials of the two Hamiltonians are equal to within an additive constant. Given two external potentials v_1 and v_2 such that $v_1, v_2 \in \mathcal{V}_N$ assumed to share the same ground–state wavefunction Ψ, the corresponding Schrödinger equations are

$$\hat{\mathcal{H}}_\lambda(v_1)\Psi = \mathcal{E}_\lambda(v_1)\Psi \qquad \hat{\mathcal{H}}_\lambda(v_2)\Psi = \mathcal{E}_\lambda(v_2)\Psi. \tag{3.14}$$

Subtracting the Schrödinger Equation of the second system from that of the first, and eliminating the terms $\hat{T}$ and $\hat{W}_\lambda$ from the Hamiltonian, as these operators only vary between systems if the number of electrons varies, leads to the expression

$$\left[\hat{\mathcal{H}}_\lambda(v_1) - \hat{\mathcal{H}}_\lambda(v_2)\right]\Psi = \sum_i \left[v_1(\mathbf{r}_i) - v_2(\mathbf{r}_i)\right]\Psi = \left[\mathcal{E}_\lambda(v_1) - \mathcal{E}_\lambda(v_2)\right]\Psi \tag{3.15}$$

This may be simplified by eliminating Ψ, leaving an identity relating the two potentials and the ground–state energies of Ψ as

$$\sum_i \left[v_1(\mathbf{r}_i) - v_2(\mathbf{r}_i)\right] = \mathcal{E}_\lambda(v_1) - \mathcal{E}_\lambda(v_2) \qquad \Longrightarrow \qquad v_1(\mathbf{r}_i) - v_2(\mathbf{r}_i) = c \in \mathbb{R}, \tag{3.16}$$

assuming $\Psi \neq 0$ over all space. Thus the two potentials v_1 and v_2 must be equal to within a constant $c \in \mathbb{R}$ if they are to yield wavefunctions that are identical in regions of non–zero measure. Any pair of potentials differing by more than an additive constant yield linearly independent wavefunctions.

This analysis can then be applied to densities which, in Hohenberg–Kohn theory, are considered to pertain to ground–states. The set of ground–state densities for potentials $v \in \mathcal{V}_N$ are said to be *v–representable*,

$$\mathcal{A}_N = \left\{\rho \;\middle|\; \rho \text{ is yielded from an } N\text{-electron ground state of } \hat{\mathcal{H}}_\lambda(v)\,,\; v \in \mathcal{V}_N\right\}. \tag{3.17}$$

Given a ground–state density $\rho \in \mathcal{A}_N$ corresponding to potential $v \in \mathcal{V}_N$ and applying the Rayleigh–Ritz variational principle, it can be seen for all $u \in \mathcal{V}_N$ that

$$\begin{aligned}\mathcal{E}_\lambda(u) = \left\langle \Psi_u \middle| \hat{\mathcal{H}}_\lambda(u) \middle| \Psi_u \right\rangle &\leqslant \left\langle \Psi_v^\rho \middle| \hat{\mathcal{H}}_\lambda(u) \middle| \Psi_v^\rho \right\rangle \\ &\leqslant \left\langle \Psi_v^\rho \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_v^\rho \right\rangle + (u - v|\rho) = \mathcal{E}_\lambda(v) + (u - v|\rho)\,,\end{aligned} \tag{3.18}$$

where $\Psi_v^\rho \in \mathcal{W}_N$ is a ground–state wavefunction of $\hat{\mathcal{H}}_\lambda(v)$ with density ρ whilst $\Psi_u \in \mathcal{W}_N$ is a ground–state wavefunction of $\hat{\mathcal{H}}_\lambda(u)$.

Considering then a density $\rho \in \mathcal{A}_N$ that is not a ground–state density of $v \in \mathcal{V}_N$, it must therefore be the ground–state density of some potential $u \in \mathcal{V}_N$ since it is v–representable. However, $u - v \neq c \in \mathbb{R}$ else the ground–state wavefunctions would not be linearly independent. It can be seen, from applying the Rayleigh–Ritz variation principle, that

$$\begin{aligned}\mathcal{E}_\lambda(u) = \left\langle \Psi_u^\rho \middle| \hat{\mathcal{H}}_\lambda(u) \middle| \Psi_u^\rho \right\rangle &= \left\langle \Psi_u^\rho \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_u^\rho \right\rangle + (u - v|\rho) \\ &> \left\langle \Psi_v \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_v \right\rangle + (u - v|\rho) = \mathcal{E}_\lambda(v) + (u - v|\rho)\,.\end{aligned} \tag{3.19}$$

Furthermore, given the relation of (3.16), it can be shown that for a potential $u - v = c \in \mathbb{R}$ the inequality (3.18) becomes saturated as the two wavefunctions must be equal in all regions of non–zero measure,

$$\begin{aligned}\mathcal{E}_\lambda(u) = \left\langle \Psi_u \middle| \hat{\mathcal{H}}_\lambda(u) \middle| \Psi_u \right\rangle &= \left\langle \Psi_v^\rho \middle| \hat{\mathcal{H}}_\lambda(u) \middle| \Psi_v^\rho \right\rangle \\ &= \left\langle \Psi_v^\rho \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi_v^\rho \right\rangle + (u - v|\rho) = \mathcal{E}_\lambda(v) + (c|\rho)\end{aligned} \tag{3.20}$$

where $(c|\rho) = c\int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = cN$, thus yielding the same identity as (3.12). The results (3.18), (3.19) & (3.20) may be summarised as the *energy supergradient inequality*.

Definition 3.1: The Energy Supergradient Inequality

A density $\rho \in \mathcal{A}_N$ is a ground–state density of the potential $v \in \mathcal{V}_N$ if and only if

$$\mathcal{E}_\lambda(u) \leqslant \mathcal{E}_\lambda(v) + (u - v|\rho) \qquad \forall u \in \mathcal{V}_N. \tag{3.21}$$

The inequality is saturated if and only if $u - v = c \in \mathbb{R}$, giving the result

$$\mathcal{E}_\lambda(v + c) = \mathcal{E}_\lambda(v) + cN \qquad \forall c \in \mathbb{R}. \tag{3.22}$$

In considering the energy supergradient inequality, (3.21) may be viewed as comprising a linear approximation to the ground–state energy, exact for the external potential v, which thus never underestimates the exact ground–state energy, as illustrated in Figure 3.1.

The first Hohenberg–Kohn theorem can now be elucidated from the energy supergradient inequality. Given two potentials $v_1, v_2 \in \mathcal{V}_N$ with ground–state densities $\rho_1, \rho_2 \in \mathcal{A}_N$ respectively, the relations

$$\mathcal{E}_\lambda(v_1) \leqslant \mathcal{E}_\lambda(v_2) + (v_1 - v_2|\rho_2) \qquad \& \qquad \mathcal{E}_\lambda(v_2) \leqslant \mathcal{E}_\lambda(v_1) - (v_1 - v_2|\rho_1)\,, \tag{3.23}$$

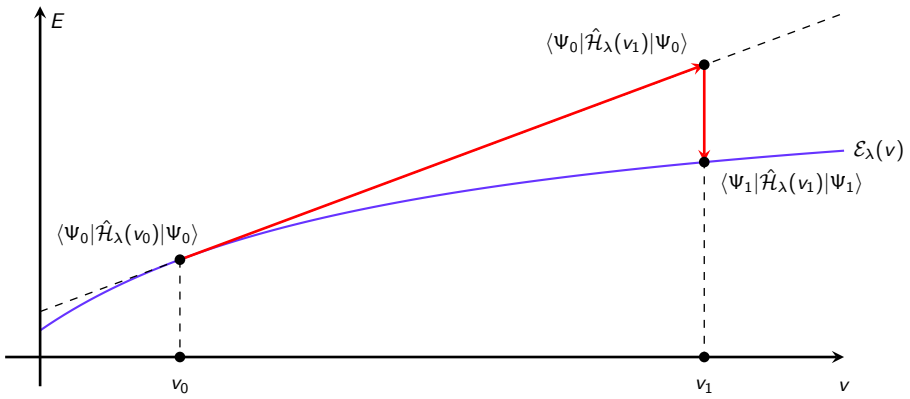

Figure 3.1: An illustration of the linearity of the Hamiltonian in the potential v. Assuming potentials v_0 & v_1 have respective ground–state wavefunctions Ψ_0 & Ψ_1, as the potential varies $v_0 \to v_1$ with fixed wavefunction Ψ_0 the energy varies linearly. If the wavefunction then relaxes from Ψ_0 to its ground state Ψ_1 at potential v_1, the energy decreases. The ground–state energy $\mathcal{E}_\lambda(v)$ is thus concave in v.

follow from (3.21). Taking the sum of these two inequalities and simplifying yields

$$\mathcal{E}_\lambda(v_1) + \mathcal{E}_\lambda(v_2) \leqslant \mathcal{E}_\lambda(v_1) + \mathcal{E}_\lambda(v_2) + (v_1 - v_2|\rho_2 - \rho_1) \quad \Longrightarrow \quad (v_1 - v_2|\rho_2 - \rho_1) \leqslant 0, \tag{3.24}$$

from which it can be deduced that

$$\begin{aligned} (v_1 - v_2|\rho_2 - \rho_1) &= 0 \quad \text{if} \quad v_1 - v_2 = c \in \mathbb{R} \\ (v_1 - v_2|\rho_2 - \rho_1) &< 0 \quad \text{if} \quad v_1 - v_2 \neq c \in \mathbb{R}. \end{aligned} \tag{3.25}$$

In other words, if the two potentials differ by more than a scalar term, the difference between their respective ground–state densities must be non–zero as otherwise the inner product $(v_1 - v_2|\rho_2 - \rho_1)$ would vanish, thus proving the first Hohenberg–Kohn theorem.

Theorem 3.2: The First Hohenberg–Kohn Theorem

A ground–state density determines the external potential uniquely up to an additive constant.

$$v_1 - v_2 \neq c \quad \Longrightarrow \quad \rho_1 - \rho_2 \neq 0 \qquad \rho_1, \rho_2 \in \mathcal{A}_N \;\; \& \;\; v_1, v_2 \in \mathcal{V}_N \tag{3.26}$$

It follows from this that there is a $1:1$ correspondence between families of ρ–representable potentials, defined to within an additive constant, and sets of degenerate v–representable densities. In other words, the density determines the potential to within an additive constant whilst the potential determines the density to within a degenerate set.

3.2.2 The Second Hohenberg–Kohn Theorem

The second Hohenberg–Kohn theorem follows directly from the first, utilising the $1:1$ correspondence between ρ–representable potentials & v–representable densities to construct an exact expression for the ground–state energy of a system in terms of the density instead of the wavefunction. The first step is to introduce the *Hohenberg–Kohn universal density functional* $F_{\text{HK}} : \mathcal{A}_N \mapsto \mathbb{R}$

$$F_{\text{HK}}(\rho) = \mathcal{E}_\lambda(v_\rho) - (v_\rho|\rho)\,, \tag{3.27}$$

where $\rho \in \mathcal{A}_N$ is a ground–state density of $v_\rho \in \mathcal{V}_N$[ii]. Applying the energy supergradient inequality in Definition 3.1, it can be seen that the inequality

$$\mathcal{E}_\lambda(v) \leqslant \mathcal{E}_\lambda(v_\rho) - (v_\rho|\rho) + (v|\rho) \tag{3.28}$$

is satisfied for all $v \in \mathcal{V}_N$ if and only if the density ρ is a ground–state density of v_ρ. Substituting the Hohenberg–Kohn functional (3.27) into this inequality yields the *Hohenberg–Kohn inequality*; for each $v \in \mathcal{V}_N$ and each $\rho \in \mathcal{A}_N$, the inequality

$$\mathcal{E}_\lambda(v) \leqslant F_{\text{HK}}(\rho) + (v|\rho) \tag{3.29}$$

[ii]The functional F_{HK} is well defined despite v_ρ only being defined to within a constant since $\mathcal{E}_\lambda(v_\rho + c) - (v_\rho + c|\rho) = \mathcal{E}_\lambda(v_\rho) - (v_\rho|\rho)$

is satisfied, and is saturated if and only if ρ is a ground–state density of v. Following directly from this, the second Hohenberg–Kohn theorem is elucidated, comprising the fundamental variational principles of DFT.

Theorem 3.3: Hohenberg–Kohn & Lieb Variation Principles

For $v \in \mathcal{V}_N$ & $\rho \in \mathcal{A}_N$, the following variation principles can be defined

$$\textbf{Hohenberg–Kohn Variation Principle:} \qquad \mathcal{E}_\lambda(v) = \min_{\rho \in \mathcal{A}_N} \{F_{\text{HK}}(\rho) + (v|\rho)\} \tag{3.30}$$

$$\textbf{Lieb Variation Principle:} \qquad F_{\text{HK}}(\rho) = \max_{v \in \mathcal{V}_N} \{\mathcal{E}_\lambda(v) - (v|\rho)\} \tag{3.31}$$

in which the minimisers of the Hohenberg–Kohn variation principle are the ground–state densities of v whilst the maximisers of the Lieb variation principle are the potentials with ground–state density ρ.

The Lieb variation principle (3.31) as defined here is equivalent to the Hohenberg–Kohn variation principle. It is however of material importance to the present work, as will be discussed in detail in Section 3.4.

The significance of the Hohenberg–Kohn variation principle is that it provides the theoretical means by which the ground–state energy $\mathcal{E}_\lambda$ for a given potential v may be computed without reference to the many–electron wavefunction but instead using the Hohenberg–Kohn functional, which is defined for a ground–state wavefunction Ψ_ρ yielding density ρ as

$$F_{\text{HK}}(\rho) = \left\langle \Psi_\rho \middle| \hat{\mathcal{H}}_\lambda(0) \middle| \Psi_\rho \right\rangle = \left\langle \Psi_\rho \middle| \hat{T} + \hat{W}_\lambda \middle| \Psi_\rho \right\rangle . \tag{3.32}$$

To do so however requires an accurate functional form of $F_{\text{HK}}(\rho)$ such that it may be evaluated directly from the density; its explicit form is unknown although approximations of it can be constructed.

Furthermore, the Hohenberg–Kohn theory is only derived for the sets of v–representable densities $\mathcal{A}_N$ & ρ–representable potentials $\mathcal{V}_N$, neither of which are easily characterised thus preventing Theorem 3.3 from being applied in practice. In the following sections, the generalisation of the Hohenberg–Kohn functional to explicitly known sets of densities & potentials, allowing direct application of the theory, will be discussed. This will provide the fundamental background & theoretical basis for the results discussed in Chapter 5.

3.3 The Levy–Lieb Constrained Search

Although the Hohenberg–Kohn theorems were groundbreaking, the poorly characterised domains of the densities and potentials over which the energy is to be optimised makes it difficult to apply in practice to electronic systems. The work of Hohenberg & Kohn was further developed independently by Levy[70] and Lieb[67] to improve its definition and make it more useful. What resulted was the *Levy–Lieb Constrained–Search Functional* $F_{\text{LL}}(\rho)$, with which a quantum mechanical problem can be posed as an optimisation over known domains.

To discuss the significance of this development further in a way that is coherent, and also to make this document reasonably self–contained, it is first necessary to outline some of the key mathematical principles on which the work is based. However, it is far beyond the scope of this discussion to undertake an comprehensive mathematical derivation. A full exposition of the principles set out here may be found in the detailed works of Eschrig[71] & Helgaker[72], whilst a detailed survey of mathematical analysis can be found in the work of Lieb & Loss[73].

3.3.1 Sets & Operators

Sets

A *set* is the general term for a collection of elements. A set may contain

(i) Zero elements for example the *empty set* $\emptyset$,

(ii) A finite number of elements for which the number is known as the *cardinality* of the set,

(iii) An infinite number of elements such as the set of natural numbers $\mathbb{N}$.

An element x may be identified as a member of the set $\mathcal{A}$ or not a member of $\mathcal{A}$ with the notation $x \in \mathcal{A}$ or $x \notin \mathcal{A}$ respectively.

A set $\mathcal{A}$ is labelled a *subset* of another set $\mathcal{B}$ if every element in $\mathcal{A}$ is also a member of $\mathcal{B}$,

$$\mathcal{A} \subset \mathcal{B} \quad \text{if} \quad x \in \mathcal{B} \ \ \forall x \in \mathcal{A}. \tag{3.33}$$

If the two sets are not identical, i.e. there are elements in $\mathcal{B}$ that are not members of $\mathcal{A}$ then the set $\mathcal{A}$ is a *proper subset* of $\mathcal{B}$; $\mathcal{A} \subsetneq \mathcal{B}$.

Set Operations

For two sets $\mathcal{A}$ & $\mathcal{B}$, several *set operations* can be defined as

(i) *Union*: $\mathcal{A} \cup \mathcal{B} = \{x \mid x \in \mathcal{A} \text{ or } x \in \mathcal{B}\}$,

(ii) *Intersection*: $\mathcal{A} \cap \mathcal{B} = \{x \mid x \in \mathcal{A} \text{ and } x \in \mathcal{B}\}$,

(iii) *Complement*: $\mathcal{A} \setminus \mathcal{B} = \{x \mid x \in \mathcal{A} \text{ and } x \notin \mathcal{B}\}$,

(iv) *Cartesian product*: $\mathcal{A} \times \mathcal{B} = \{(x, y) \mid x \in \mathcal{A}\ ,\ y \in \mathcal{B}\}$.

If the intersection $\mathcal{A} \cap \mathcal{B} = \emptyset$ the two sets are said to be *disjoint* whereas if $\mathcal{A} \cap \mathcal{B} \neq \emptyset$ then the set $\mathcal{A}$ is said to *intersect* $\mathcal{B}$.

Set Bounds

An *equivalence relation* $\sim$ can be defined between pairs of elements x, y, z of a set $\mathcal{A}$ satisfying the conditions: $x \sim x$, $x \sim y \implies y \sim x$ and $x \sim y$ & $y \sim z \implies x \sim z$. The subset of $\mathcal{A}$ containing all elements equivalent to an element $x \in \mathcal{A}$ is defined as the *equivalence class* $[x]$.

A *partial order* such as $\leqslant$ is a relation between pairs of elements x, y, z in set $\mathcal{A}$ that satisfy: $x \sim y$, $x \leqslant y$ & $y \leqslant x \implies x = y$ and $x \leqslant y$ & $y \leqslant z \implies x \leqslant z$. A set equipped with this relation is termed a *partially ordered set*. For a partially ordered set $\mathcal{A}$ with subset $\mathcal{S} \subset \mathcal{A}$ and element $a \in \mathcal{A}$,

(i) a is the *lower bound* of $\mathcal{S}$ if $a \leqslant x$ for all $x \in \mathcal{S}$,

(ii) a is the *upper bound* of $\mathcal{S}$ if $a \geqslant x$ for all $x \in \mathcal{S}$,

for which $\mathcal{S}$ is said to be bounded from above or bounded from below respectively.[iii] The element a may be thus characterised in one of two ways:

(i) If the subset $\mathcal{S} \subset \mathcal{A}$ has a lower bound a and $a \in \mathcal{S}$, then a is the *minimum* of $\mathcal{S}$. Likewise, if $\mathcal{S} \subset \mathcal{A}$ has an upper bound $a \in \mathcal{S}$, then a is its *maximum*.

(ii) If the subset $\mathcal{S} \subset \mathcal{A}$ has lower bounds a_l, the element $a \in \mathrm{a}_l$ for which $a \geqslant \alpha \in \mathrm{a}_l$ is the greatest lower bound, or *infimum* of $\mathcal{S}$. Equally, if $\mathcal{S}$ has upper bounds a_u, the element $a \in \mathrm{a}_u$ for which $a \leqslant \alpha \in \mathrm{a}_u$ is the least upper bound, or *supremum* of $\mathcal{S}$.

3.3.2 Vector Spaces

Fields

In mathematical terms, a *field* $\mathbb{F}$ is an arbitrary set of elements that satisfy the six *field axioms* for the addition and multiplication operators. These axioms are outlined for arbitrary elements a, b & $c \in \mathbb{F}$ in Table 3.2[74]. The individual

Axiom	Addition	Multiplication
Associativity	$(a + b) + c = a + (b + c)$	$(ab)c = a(bc)$
Commutativity	$a + b = b + a$	$ab = ba$
Distributivity	$a(b + c) = ab + ac$	$(a + b)c = ac + bc$
Identity	$a + 0 = 0 + a = a$	$a \cdot 1 = 1 \cdot a = a$
Inverses	$a + (-a) = (-a) + a = 0$	$aa^{-1} = a^{-1}a = 1 \quad (a \neq 0)$

Table 3.2: The six field axioms that characterise a scalar field.

elements of a field are simply *scalars*. A common example of a scalar field would be the set of real numbers $\mathbb{R}$.

[iii] If the subset $\mathcal{S}$ is bounded from above and below, it is simply referred to as *bounded*.

Vector Spaces

A *vector space* $\mathcal{X}$ over a scalar field $\mathbb{F}$ is a set that satisfies the operations of vector addition and multiplication by scalars. For the vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$ and scalars $\alpha, \beta \in \mathbb{F}$, the following conditions must be satisfied:

(i) Associativity of addition: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$

(ii) Commutativity of addition: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(iii) Zero vector: $\exists\, \mathbf{0} \in \mathcal{X}$ where $\mathbf{x} + \mathbf{0} = \mathbf{x}$

(iv) Additive inverse: $\forall\, \mathbf{x} \in \mathcal{X}$, $\exists\, \mathbf{y} \in \mathcal{X}$ where $\mathbf{x} + \mathbf{y} = \mathbf{0}$

(v) Associativity of multiplication: $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$

(vi) Distributivity: $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$

(vii) Multiplicative identity: $\exists\, 1 \in \mathbb{F}$ where $1 \cdot \mathbf{x} = \mathbf{x} \quad \forall\, \mathbf{x} \in \mathcal{X}$

Normed Spaces

A *norm* on a vector space $\mathcal{X}$ is a mapping $\|\cdot\| : \mathcal{X} \mapsto \mathbb{R}$ that satisfies the conditions

(i) $\|\mathbf{x}\| \geqslant 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$

(ii) $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$

(iii) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\alpha \in \mathbb{F}$.

A vector space in which a norm is defined is known as a *normed vector space*. A vector space may be equipped with more than one norm, for example the k–dimensional real coordinate space $\mathbb{R}^k$ is equipped with the norms

(i) *Manhattan Norm:* $\|\mathbf{x}\|_1 = \sum_{i=1}^{k} |x_i|$

(ii) *Euclidean Norm:* $\|\mathbf{x}\|_2 = \left[\sum_{i=1}^{k} |x_i|^2\right]^{\frac{1}{2}}$

(iii) *Maximum Norm:* $\|\mathbf{x}\|_\infty = \max_{1 \leqslant i \leqslant k} |x_i|$

for all $\mathbf{x} \in \mathbb{R}^k$. A normed space can further be identified as a *metric space* if, in addition to a norm, it is equipped with the induced metric

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathcal{X}. \tag{3.34}$$

The induced metric may be used to characterise members of the vector space $\mathcal{X}$; a *sequence* $\{\mathbf{x}_n\}_{n=1}^{\infty}$ converges to $\mathbf{x} \in \mathcal{X}$ if, for any $\delta > 0$, there exists an $N \in \mathbb{N}$ such that $d(\mathbf{x}, \mathbf{x}_n) < \delta \;\; \forall n \geqslant N$.

The sequence $\{\mathbf{x}_n\}$ may be defined as a *Cauchy sequence* if, for each $\delta > 0$, there exists an $N \in \mathbb{N}$ for which $d(\mathbf{x}_m, \mathbf{x}_n) < \delta \;\; \forall\, m, n \geqslant N$. All convergent sequences are Cauchy sequences, however a sequence may be a Cauchy sequence but not convergent. A metric space is said to be *complete* if and only if each Cauchy sequence in $\mathcal{X}$ converges to a point in that space. A metric space that is complete with respect the vector norm is known as a *Banach Space*.

Inner–Product Spaces

An *inner product* on the vector space $\mathcal{X}$ over a field $\mathbb{F}$ is a mapping $(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{F}$ such that the conditions

(i) $(\mathbf{x}, \mathbf{x}) \geqslant 0$ and $(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$

(ii) $(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z})$

(iii) $(\alpha\mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y})$

(iv) $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})^*$

are satisfied $\forall\, \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$ and $\forall \alpha \in \mathbb{F}$. A vector space in which an inner product is defined is an *inner–product space*. All elements of an inner–product space $\mathcal{X}$ satisfy the *Cauchy–Schwarz inequality*,

$$|(\mathbf{x}, \mathbf{y})| \leqslant (\mathbf{x}, \mathbf{x})^{\frac{1}{2}} \cdot (\mathbf{y}, \mathbf{y})^{\frac{1}{2}} \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathcal{X}. \tag{3.35}$$

The inner–product space is normed if it is equipped with the norm $\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})^{\frac{1}{2}}$ for all $\mathbf{x} \in \mathcal{X}$. If the inner–product space is complete in this norm, it is known as a *Hilbert space* - in other words, a Hilbert space is simply a Banach space equipped with an inner product.

Functions

A *function* $f : \mathcal{X} \mapsto \mathcal{Y}$ assigns to each element $\mathbf{x}$ in its *domain* $\mathcal{X}$ an element $f(\mathbf{x})$ of its *codomain* $\mathcal{Y}$. For $\mathcal{X}_0 \subset \mathcal{X}$ and $\mathcal{Y}_0 \subset \mathcal{Y}$, it is possible to define

(i) The *image* $f(\mathcal{X}_0)$ of $\mathcal{X}_0$ under f: $f(\mathcal{X}_0) = \{\mathbf{y} \in \mathcal{Y} \mid \mathbf{y} = f(\mathbf{x}) \text{ for at least one } \mathbf{x} \in \mathcal{X}_0\}$,

(ii) The *preimage* $f^{-1}(\mathcal{Y}_0)$ of $\mathcal{Y}_0$ under f: $f^{-1}(\mathcal{Y}_0) = \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}) \in \mathcal{Y}_0\}$.

If $\mathcal{S}$ is a subset of $\mathcal{X}$, then the *restriction* of $f : \mathcal{X} \mapsto \mathcal{Y}$ to $\mathcal{S}$ is the function $f|_{\mathcal{S}} : \mathcal{S} \mapsto \mathcal{Y}$, which is defined by $f|_{\mathcal{S}}(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{S}$.

The function $f : \mathcal{X} \mapsto \mathcal{Y}$ can be characterised as

(i) *Injective* if $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$ for $\mathbf{x}_1 \neq \mathbf{x}_2$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$,

(ii) *Surjective* onto $\mathcal{Y}$ if, for every $\mathbf{y} \in \mathcal{Y}$ there is some $\mathbf{x} \in \mathcal{X}$ such that $f(\mathbf{x}) = \mathbf{y}$,

(iii) *Bijective* if it is both injective & surjective, meaning that there is a one–to–one correspondence between elements in the domain and codomain of the function.

All bijective functions are *invertible*, meaning that if $f : \mathcal{X} \mapsto \mathcal{Y}$ is bijective then its inverse exists as $f^{-1} : \mathcal{Y} \mapsto \mathcal{X}$ such that $f^{-1}(\mathbf{y}) = \mathbf{x}$ if $f(\mathbf{x}) = \mathbf{y}$.

Linear Transformations

A *linear transformation* is a mapping between two vector spaces which preserves the operations of vector addition and scalar multiplication; given the vector spaces $\mathcal{X}$ & $\mathcal{Y}$ defined on the field $\mathbb{F}$, a function $f : \mathcal{X} \mapsto \mathcal{Y}$ is a linear transformation if

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \tag{3.36}$$

for all $x, y \in \mathcal{X}$ and $\alpha, \beta \in \mathbb{F}$. The set of functions to which the linear transformation $f : \mathcal{X} \mapsto \mathcal{Y}$ belongs can be denoted $F(\mathcal{X}, \mathcal{Y})$; this constitutes a vector space over the field $\mathbb{F}$ if the following operations are preserved:

(i) *Vector addition*: $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$,

(ii) *Scalar multiplication*: $(\alpha f)(\mathbf{x}) = \alpha f(\mathbf{x})\ \forall \alpha \in \mathbb{F}$.

for all $f, g \in F(\mathcal{X}, \mathcal{Y})$. A linear transformation $f : \mathcal{X} \mapsto \mathbb{F}$ from a vector space to its underlying scalar field is generally referred to as a *linear functional*.

A linear transformation $f \in F(\mathcal{X}, \mathcal{Y})$ from the normed space $\mathcal{X}$ to the normed space $\mathcal{Y}$ is said to be *bounded* if there exists a nonnegative real number k such that

$$\|f(\mathbf{x})\|_{\mathcal{Y}} \leqslant k\|\mathbf{x}\|_{\mathcal{X}} \ \ \forall \mathbf{x} \in \mathcal{X}. \tag{3.37}$$

If the set of linear transformations $F(\mathcal{X}, \mathcal{Y})$ is equipped with the *operator norm*

$$\|f\| = \sup_{\mathbf{x} \in \mathcal{X} \setminus \emptyset} \frac{\|f(\mathbf{x})\|}{\|\mathbf{x}\|}, \tag{3.38}$$

then $F(\mathcal{X}, \mathcal{Y})$ can be identified as a normed space, on which the *operator norm inequality* is satisfied,

$$\|f(\mathbf{x})\| \leqslant \|f\| \cdot \|\mathbf{x}\|. \tag{3.39}$$

If the space $\mathcal{Y}$ is a Banach space, then the set of linear transformations $F(\mathcal{X}, \mathcal{Y})$ must also constitute a Banach space - henceforth denoted $B(\mathcal{X}, \mathcal{Y})$.

Dual & Bidual Spaces

A bounded linear transformation f from a vector space $\mathcal{X}$ to its underlying scalar field $\mathbb{F}$ is simply a bounded linear functional. The set of such bounded linear functionals $B(\mathcal{X}, \mathbb{F})$ can be characterised as the *dual space* of $\mathcal{X}$, denoted $\mathcal{X}^*$

$$\mathcal{X}^* = B(\mathcal{X}, \mathbb{F}), \tag{3.40}$$

in which $\mathcal{X}$ is sometimes referred to as the *predual* of $\mathcal{X}^*$. As this constitutes a normed space, the dual of $\mathcal{X}^*$ can be defined as

$$\mathcal{X}^{**} = B(\mathcal{X}^*, \mathbb{F}), \tag{3.41}$$

which is more commonly known as the *bidual* of $\mathcal{X}$. Given that both $\mathcal{X}^*$ and $\mathcal{X}^{**}$ are equipped with the operator norm in the form (3.38), it follows that both the dual and bidual of a normed space are themselves Banach spaces. A normed space can be characterised further from its dual & bidual spaces in the following way:

(i) If each element in $\mathcal{X}^{**}$ can be associated with an element in $\mathcal{X}$, then $\mathcal{X}$ is said to be *reflexive*: $\mathcal{X} = \mathcal{X}^{**}$.

(ii) If $\mathcal{X}^{**}$ contains elements not associated with any member of $\mathcal{X}$ in this manner, then $\mathcal{X}$ is said to be *nonreflexive*: $\mathcal{X} \subsetneq \mathcal{X}^{**}$.

Only Banach spaces are reflexive, however a Banach space $\mathcal{X}$ is only reflexive if its dual $\mathcal{X}^*$ is also reflexive.

Lebesgue Spaces

Linear functionals, as defined above, are essential in the description of density functional theory; the density and the external potential are both linear functionals as they map a vector in the three–dimensional Euclidean space to a scalar value, $\rho : \mathbb{R}^3 \mapsto \mathbb{R}$ & $v : \mathbb{R}^3 \mapsto \mathbb{R}$ respectively. The vector space of linear functionals is identified as a normed space if it is equipped with the *p–norm*,

$$\|f\|_p = \left(\int_{\mathcal{X}} |f(\mathbf{x})|^p \mathrm{d}\mathbf{x} \right)^{\frac{1}{p}} \quad , \quad 1 \leqslant p < \infty, \tag{3.42}$$

in other words, the functions must be *measurable*[iv] and be *Lebesgue p–integrable*[v]. The space of Lebesgue p–integrable functions is generally known as an L^p space, with the definition

$$\mathrm{L}^\mathrm{p}(\mathcal{X}) = \left\{ f : \mathcal{X} \mapsto \mathbb{R} \;\middle|\; \|f\|_p < +\infty \right\}, \tag{3.43}$$

and which are themselves also Banach spaces. To show that this is the case, the p–norm (3.42) must satisfy the conditions for a norm on a vector space; $\forall\, f, g \in \mathrm{L}^\mathrm{p}(\mathcal{X})$,

(i) $\|f\|_p \geqslant 0$ and $\|f\|_p = 0$ if and only if $f = 0$,

(ii) $\|\alpha f\|_p = |\alpha| \|f\|_p \;\; \forall \alpha \in \mathbb{F}$,

(iii) $\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$.

The first two of these follow fairly straightforwardly from the definition of the p–norm and the preservation of the vector addition & scalar multiplication operations $\forall\, f, g \in \mathrm{L}^\mathrm{p}(\mathcal{X})$. The third condition is less straightforward and requires the introduction of the *Hölder inequality* of Lebesgue p–integrable functions.

Definition 3.2: The Hölder Inequality

For the measurable functions $f \in \mathrm{L}^\mathrm{p}(\mathcal{X})$ and $g \in \mathrm{L}^\mathrm{q}(\mathcal{X})$, $1 \leqslant p, q < \infty$ then

$$\|f + g\|_1 \leqslant \|f\|_p \cdot \|g\|_q \qquad \forall \;\; \frac{1}{p} + \frac{1}{q} = 1. \tag{3.44}$$

[iv] The term *measurable* here refers to *measure theory*. Whilst a discussion of measure theory is beyond the scope of the present work, it is noted that a *measure* is the general term for a function on a set that assigns a number to its appropriate subsets. An example of this is the *Lebesgue measure*, which assigns appropriate values to the subsets of the *n*–dimensional Euclidean space; length, area and volume for the 1, 2 and 3–dimensional Euclidean spaces respectively.

[v] Technically the p–norm is defined for *equivalence classes* of linear functionals $[f]$ - sets of linear functionals that are identical in regions of non–zero measure, rather than individual functionals themselves.

The requirement that $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$ can then be shown as

$$\begin{aligned}\|f+g\|_p^p &= \int_{\mathcal{X}} |f(\mathbf{x})+g(\mathbf{x})|^p \mathrm{d}\mathbf{x} \\ &= \int_{\mathcal{X}} |f(\mathbf{x})+g(\mathbf{x})| \cdot |f(\mathbf{x})+g(\mathbf{x})|^{p-1} \mathrm{d}\mathbf{x} \\ &\leqslant \int_{\mathcal{X}} |f(\mathbf{x})| \cdot |f(\mathbf{x})+g(\mathbf{x})|^{p-1} \mathrm{d}\mathbf{x} + \int_{\mathcal{X}} |g(\mathbf{x})| \cdot |f(\mathbf{x})+g(\mathbf{x})|^{p-1} \mathrm{d}\mathbf{x},\end{aligned} \tag{3.45}$$

where the final step is obtained by recognising that $|f+g| \leqslant |f| + |g|$. Applying the Hölder inequality to each of the two terms on the right–hand side of (3.45) yields the following

$$\begin{aligned}\|f+g\|_p^p &\leqslant (\|f\|_p + \|g\|_p) \cdot \||f+g|^{p-1}\|_{\frac{p}{p-1}} \\ &\leqslant (\|f\|_p + \|g\|_p) \cdot \|f+g\|_p^p \cdot \|f+g\|_p^{-1},\end{aligned} \tag{3.46}$$

which may be re–arranged to yield the required result, known as the *Minkowski's inequality*,

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p. \tag{3.47}$$

A special case of the L^p spaces can be defined for $p = \infty$ which satisfies the same requirements as the general L^p spaces, however uses the *supremum norm*

$$\|f\|_\infty = \inf\left\{C \geqslant 0 \;\middle|\; |f(\mathbf{x})| \leqslant C \text{ almost everywhere in } \mathcal{X}\right\}. \tag{3.48}$$

Thus the normed space L^∞ of all such *essentially bounded functions* is a Banach space defined as

$$\mathrm{L}^\infty(\mathcal{X}) = \left\{f : \mathcal{X} \mapsto \mathbb{F} \;\middle|\; \|f\|_\infty < +\infty\right\}. \tag{3.49}$$

Dual Of Lebesgue Spaces

The characteristics of dual and bidual spaces defined previously for Banach spaces can be applied to Lebesgue spaces, providing the means by which the spaces of densities and potentials may be linked in DFT.

Given the functions $f \in \mathrm{L}^p$ and $g \in \mathrm{L}^q$, with $1 \leqslant p, q < \infty$, where p and q are conjugate indices satisfying $p^{-1} + q^{-1} = 1$, the function $T_g : \mathrm{L}^p \mapsto \mathbb{F}$ given by

$$T_g(f) = \int_{\mathcal{X}} g(\mathbf{x}) \cdot f(\mathbf{x}) \mathrm{d}\mathbf{x} = (g, f) \tag{3.50}$$

can be shown to belong to the space L^{p*} and have the norm $\|T_g\| = \|g\|_p$ by application of the Hölder inequality, Definition 3.2. Therefore, between the elements of L^q & L^{p*} there exists a norm–preserving one–to–one correspondence for all $p \neq \infty$.

By extension, it can be determined that the dual of L^p satisfies $\mathrm{L}^{p*} = \mathrm{L}^q$ with $p^{-1} + q^{-1} = 1$ when $1 \leqslant p < \infty$ and $\mathrm{L}^{\infty *} \supsetneq \mathrm{L}^1$. Clearly there exist bounded linear functionals T on $\mathrm{L}^{\infty *}$ that cannot be represented by $T(f) = (g, f)$ for $g \in \mathrm{L}^1$. Every Lebesgue space L^p is therefore reflexive for $1 < p < \infty$, whilst L^1 & L^∞ are nonreflexive. This relationship between a Lebesgue space and its dual ensures that the pairing (g, f) is finite for all $f \in \mathrm{L}^p$ and $g \in \mathrm{L}^q$.

The L^p spaces are essential to the mathematical formulation of DFT discussed in the present work, as can be demonstrated by the two examples:

(i) The $\mathrm{L}^1(\mathbb{R}^3)$ space of absolutely integrable functions can be seen to contain the density $\rho(\mathbf{r})$ since it must satisfy the identity $\int_{\mathbb{R}^3} |\rho(\mathbf{r})| \mathrm{d}\mathbf{r} = N$,

(ii) The $\mathrm{L}^2(\mathbb{R}^3)$ space of square integrable functions contains the bound–state one–electron wavefunctions $\Psi(\mathbf{r})$ since its normalisation requires that $\int_{\mathbb{R}^3} |\Psi(\mathbf{r})|^2 \mathrm{d}\mathbf{r} = 1$.

However, not all elements of $\mathrm{L}^1(\mathbb{R}^3)$ & $\mathrm{L}^2(\mathbb{R}^3)$ correspond to appropriate densities or wavefunctions respectively, for example densities for which $\rho(\mathbf{r}) < 0$ at some $\mathbf{r}$ would still belong to the $\mathrm{L}^1(\mathbb{R}^3)$ space but would not be appropriate as densities. Using the concepts outlined here, a more precise definition of the spaces of densities and potentials appropriate in DFT will be discussed in the next section.

3.3.3 Admissible N–Electron Wavefunctions

In Section 3.1, it was required that the external potential in the Hamiltonian be ρ–representable to ensure the existence of a ground–state wavefunction. Considering instead the vector space of all *admissible potentials* $\mathcal{U}$ that yield a finite expectation value of the Hamiltonian, the ground–state energy can be defined as an *infimum* over admissible N–electron wavefunctions at any given $v \in \mathcal{U}$,

$$\mathcal{E}_\lambda(v) = \inf_{\Psi \in \mathcal{W}_N} \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) \right| \Psi \right\rangle \quad \forall v \in \mathcal{U}. \tag{3.51}$$

Given the generalised Rayleigh–Ritz variation principle (3.51), the set of *admissible N–electron wavefunctions* $\mathcal{W}_N$ can be characterised as the set of normalised, antisymmetric N–electron wavefunctions with a finite expectation value of the Hamiltonian,

$$\mathcal{W}_N = \left\{ \Psi \;\middle|\; \|\Psi\|_2 = 1\,,\; |\langle\Psi|\hat{\mathcal{H}}_\lambda(v)|\Psi\rangle| < \infty\,,\; v \in \mathcal{V}_\mathrm{C}\,,\; N\text{–electron antisymmetric } \Psi \right\}, \tag{3.52}$$

where $\mathcal{V}_\mathrm{C}$ is the set of all point–charge *Coulomb potentials*,

$$\mathcal{V}_\mathrm{C} = \left\{ v \;\middle|\; v(\mathbf{r}_i) = -\sum_I^{N_\mathrm{C}} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}\,,\; N_\mathrm{C} \in \mathbb{N}\,,\; Z_I \in \mathbb{R} \right\}, \tag{3.53}$$

for a system of N_C point–charge nuclei with charges Z_I, $1 \leqslant I \leqslant N_\mathrm{C}$. As it is possible for $v = 0$ if the charges $Z_I = 0$, it is necessary for the expectation value of the kinetic and two–electron operators to be finite for all admissible wavefunctions,

$$|\langle\Psi|\hat{\mathcal{H}}_\lambda(0)|\Psi\rangle| < \infty \quad \Longrightarrow \quad \langle\Psi|\hat{T}|\Psi\rangle < \infty \quad \& \quad \langle\Psi|\hat{W}_\lambda|\Psi\rangle < \infty, \tag{3.54}$$

where the Hamiltonian is decomposed according to (3.1). Each of these can be considered in turn to determine the requirements for the set of admissible wavefunctions.

For a wavefunction Ψ that vanishes asymptotically, the kinetic energy expectation value can be expressed, through partial integration, as

$$T(\Psi) = -\frac{1}{2}\sum_i \left\langle \Psi \left| \nabla_i^2 \right| \Psi \right\rangle = \frac{1}{2}\sum_i \langle \nabla_i\Psi | \nabla_i\Psi \rangle = \frac{1}{2}\sum_i \|\nabla_i\Psi\|_2^2 > 0, \tag{3.55}$$

meaning that the gradients of admissible N–electron wavefunctions must be square–integrable, $\nabla_i\Psi \in \mathrm{L}^2(\mathbb{R}^{3N})$. Additionally, as described in the previous section, a wavefunction of a finite system which is normalised must belong to the space of square–integrable functions; $\Psi \in \mathrm{L}^2(\mathbb{R}^{3N})$. It therefore follows that admissible N–electron wavefunctions satisfying these two criteria must belong to the *first–order Sobolev space* $\mathcal{H}^1_N$,

$$\mathcal{H}^1_N = \left\{ f \;\middle|\; f \in \mathrm{L}^2(\mathbb{R}^{3N})\,,\; \nabla f \in \mathrm{L}^2(\mathbb{R}^{3N}) \right\}, \tag{3.56}$$

which is defined as a Hilbert space equipped with the finite inner product

$$(f, g) = \int_{\mathbb{R}^{3N}} f(\mathbf{r}) \cdot g(\mathbf{r})\, \mathrm{d}\mathbf{r} + \int_{\mathbb{R}^{3N}} \nabla f(\mathbf{r}) \cdot \nabla g(\mathbf{r})\, \mathrm{d}\mathbf{r} \quad \forall\, f, g \in \mathcal{H}^1_N. \tag{3.57}$$

It is thus determined that admissible N–electron wavefunctions must belong to the first–order Sobolev space $\Psi \in \mathcal{H}^1_N$. It can be seen that this requirement also ensures that the expectation value of the two–electron operator is finite; by re–expressing the two–electron operator as a commutator with the gradient and using partial integration, its expectation value is re–written as

$$W_\lambda(\Psi) = \sum_{i>j} \left\langle \Psi \left| \frac{\lambda}{|\mathbf{r}_{ij}|} \right| \Psi \right\rangle = \frac{1}{2}\sum_{i>j} \left\langle \Psi \left| \left[\nabla_i, \lambda \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \right] \right| \Psi \right\rangle = -\sum_{i>j} \mathrm{Re} \left\langle \nabla_i\Psi \left| \lambda \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \right| \Psi \right\rangle, \tag{3.58}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. Resolving this expression as the sum of three Cartesian components α, a *generalised Hölder inequality* can be used to determine an upper bound to the matrix elements of the two–electron expectation value,

$$\left| \left\langle \Psi \left| \frac{\lambda}{|\mathbf{r}_{ij}|} \right| \Psi \right\rangle \right| \leqslant \sum_\alpha \left| \left\langle \nabla_{\alpha i}\Psi \left| \lambda \frac{\mathbf{r}_{\alpha ij}}{|\mathbf{r}_{ij}|} \right| \Psi \right\rangle \right| \leqslant \lambda \sum_\alpha \left(\|\nabla_{\alpha i}\Psi\|_2 \cdot \left\| \frac{\mathbf{r}_{\alpha ij}}{|\mathbf{r}_{ij}|} \right\|_\infty \cdot \|\Psi\|_2 \right) > 0 \;\; (\text{for } \lambda \neq 0). \tag{3.59}$$

As for the kinetic energy, the terms $\|\nabla_{\alpha i}\Psi\|_2 < \infty$ and $\|\Psi\|_2 < \infty$ for $\Psi \in \mathcal{H}^1_N$, whilst the additional term satisfies the identity $\|\mathbf{r}_{\alpha ij}/|\mathbf{r}_{ij}|\|_\infty = 1$. This confirms that the expectation value of the two–electron operator is finite for for $\Psi \in \mathcal{H}^1_N$.

Furthermore, it is apparent that substituting the two–electron operator $\lambda/|\mathbf{r}_{ij}|$ for the Coulomb potential as given in (3.53), i.e. $Z_I/|\mathbf{r}_{iI}|$, yields an equivalent result to (3.59). Therefore, the expectation value of the Coulomb potential $|\langle\Psi|\sum_i v(\mathbf{r}_i)|\Psi\rangle|$ is also finite for $\Psi \in \mathcal{H}^1_N$, hence this condition ensures that the expectation value of the Hamiltonian is finite for all $v \in \mathcal{V}_C$.

Given that both $T(\Psi)$ and $W_\lambda(\Psi)$ are non–negative for $\Psi \in \mathcal{H}^1_N$, shown in (3.55) and (3.59) respectively, the expectation value of $\hat{\mathcal{H}}_\lambda(v)$ itself is bounded from below for all $v \in \mathcal{V}_C$. The ground–state energy is therefore well defined by the generalised Rayleigh–Ritz variation principle for all Coulomb potentials and $\Psi \in \mathcal{H}^1_N$.

Definition 3.3: Ground–State Energy For Coulomb Potentials

The ground–state energy for all Coulomb potentials $\mathcal{V}_C$ is finite and well–defined by the generalised Rayleigh–Ritz variation principle,

$$\mathcal{E}_\lambda(v) = \inf_{\Psi\in\mathcal{W}_N} \left\langle \Psi \middle| \hat{\mathcal{H}}_\lambda(v) \middle| \Psi \right\rangle \quad \forall v \in \mathcal{V}_C, \tag{3.60}$$

where $\mathcal{W}_N$ is the set of admissible N–electron wavefunctions

$$\mathcal{W}_N = \left\{ \Psi \in \mathcal{H}^1_N \;\middle|\; \|\Psi\|_2 = 1\,,\ \text{antisymmetric}\,\Psi \right\}. \tag{3.61}$$

3.3.4 Admissible N–Electron Densities

In the Hohenberg–Kohn variation principle (3.30), N–electron densities are required to belong to the set of v–representable densities; these are non–negative, integrate over all space to N and result from an N–electron ground–state of the Hamiltonian as described in subsection 3.2.1. The first two of these conditions implies that N–electron densities must belong to the space of absolutely integrable functions

$$\mathcal{L}^1_N = \left\{ \rho \in \mathrm{L}^1(\mathbb{R}^3) \;\middle|\; \rho \geqslant 0\,,\ \int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = N \right\}. \tag{3.62}$$

However, this definition itself does not necessarily ensure that a density $\rho \in \mathcal{L}^1_N$ is yielded by an admissible N–electron wavefunction $\Psi \in \mathcal{W}_N$; the densities that are yielded from such wavefunctions form the set of *N–representable densities*,

$$\mathcal{I}_N = \left\{ \rho \;\middle|\; \rho \text{ can be obtained from some } \Psi \in \mathcal{W}_N \right\}. \tag{3.63}$$

The set of N–representable densities must be characterised so that the densities $\rho \in \mathcal{L}^1_N$ for which $\rho \in \mathcal{I}_N$ can be identified without recourse to the wavefunction.

Given (3.61), it follows that only densities pertaining to antisymmetric wavefunctions are N–representable. It can however be shown that each $\rho \in \mathcal{L}^1_N$ is the density of an N–electron antisymmetric wavefunction in the form of a Slater determinant. This is demonstrated by considering a set of N complex–valued orbitals for a density $\rho \in \mathcal{L}^1_N$,

$$\phi_k(\mathbf{r}) = \sqrt{\frac{\rho(\mathbf{r})}{N}} e^{2\pi\imath(k-1)q(x)}\,,\ k = 1, 2 \ldots N \tag{3.64}$$

where $q(x)$ is a real–valued, non–negative function defined in terms of ρ as

$$q(x) = \frac{1}{N}\int_{-\infty}^{x}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(t, y, z)\mathrm{d}t\,\mathrm{d}y\,\mathrm{d}z. \tag{3.65}$$

The orbitals (3.64) can be shown to constitute an orthonormal set,

$$\int \phi^*_k(\mathbf{r})\phi_l(\mathbf{r})\mathrm{d}\mathbf{r} = \frac{1}{N}\int e^{-2\pi\imath(k-l)q(x)}\rho(\mathbf{r})\mathrm{d}\mathbf{r} = \int_{-\infty}^{\infty} e^{-2\pi\imath(k-l)q(x)} q'(x)\mathrm{d}x, \tag{3.66}$$

where $q'(x)$ is the derivative of (3.65),

$$q'(x) = \frac{1}{N}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(x, y, z)\mathrm{d}y\,\mathrm{d}z > 0, \tag{3.67}$$

thus allowing (3.66) to be re–written as

$$\int \phi_k^*(\mathbf{r})\phi_l(\mathbf{r})\mathrm{d}\mathbf{r} = \int_{-\infty}^{\infty} e^{-2\pi\imath(k-l)q(x)} q'(x)\mathrm{d}x = \int_0^1 e^{-2\pi\imath(k-l)q}\mathrm{d}q = \delta_{kl} \tag{3.68}$$

by substitution of x with $q(x)$. A Slater determinant can be constructed from the spatial orbitals ϕ_k and appropriate spin functions $\eta(\sigma_k)$ as discussed in subsection 1.2.1, the density of which is evaluated as

$$\rho_{\mathrm{SD}}(\mathbf{r}) = \sum_{k=1}^{N} |\phi_k(\mathbf{r})|^2 = \sum_{k=1}^{N} \frac{\rho(\mathbf{r})}{N} = \rho(\mathbf{r}), \tag{3.69}$$

therefore demonstrating that each density $\rho \in \mathcal{L}_N^1$ can be obtained from an N–electron Slater determinant.

Whilst the set of densities obtained from antisymmetric wavefunctions can be characterised, this does not ensure that such wavefunctions are admissible; this also requires the expectation value of the Hamiltonian to be finite. To identify such densities, it is first necessary to define the position–dependent inner product between two N–electron wavefunctions Ψ_1 & Ψ_2,

$$\langle \Psi_1 | \Psi_2 \rangle_1 (\mathbf{r}) = \int_{\mathbb{R}^3_\Sigma} \cdots \int_{\mathbb{R}^3_\Sigma} \Psi_1^*(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_N)\Psi_2(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_N)\, \mathrm{d}\sigma_1\, \mathrm{d}\boldsymbol{\tau}_2 \ldots \mathrm{d}\boldsymbol{\tau}_N, \tag{3.70}$$

in terms of which a density $\rho \in \mathcal{I}_N$ associated with wavefunction $\Psi_\rho \in \mathcal{W}_N$ can be expressed as

$$\rho(\mathbf{r}) = N \langle \Psi_\rho | \Psi_\rho \rangle_1 (\mathbf{r}). \tag{3.71}$$

Differentiation of this expression for the density with respect to the Cartesian direction α and subsequent application of the Cauchy–Schwarz inequality gives the expression

$$\begin{aligned} |\nabla_\alpha \rho(\mathbf{r})| &= N |\langle \nabla_\alpha \Psi_\rho | \Psi_\rho \rangle_1 (\mathbf{r}) + \langle \Psi_\rho | \nabla_\alpha \Psi_\rho \rangle_1 (\mathbf{r})| \\ &\leqslant 2N \left[\langle \Psi_\rho | \Psi_\rho \rangle_1 (\mathbf{r}) \langle \nabla_\alpha \Psi_\rho | \nabla_\alpha \Psi_\rho \rangle_1 (\mathbf{r}) \right]^{\frac{1}{2}} \\ &\leqslant 2 \left[N\rho(\mathbf{r}) \langle \nabla_\alpha \Psi_\rho | \nabla_\alpha \Psi_\rho \rangle_1 (\mathbf{r}) \right]^{\frac{1}{2}}, \end{aligned} \tag{3.72}$$

which is re–arranged to yield the inequality

$$\frac{|\nabla\rho(\mathbf{r})|^2}{8\rho(\mathbf{r})} \leqslant \frac{N}{2} \langle \nabla\Psi_\rho | \nabla\Psi_\rho \rangle_1 (\mathbf{r}). \tag{3.73}$$

This can then be integrated on both sides over all space as

$$\begin{aligned} \frac{1}{8} \int \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} \mathrm{d}\mathbf{r} &\leqslant \frac{N}{2} \int \langle \nabla\Psi_\rho | \nabla\Psi_\rho \rangle_1 (\mathbf{r}) \mathrm{d}\mathbf{r} \\ &\leqslant \int \cdots \int |\nabla_1 \Psi_\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_N)|^2 \, \mathrm{d}\sigma_1 \, \mathrm{d}\boldsymbol{\tau}_2 \ldots \mathrm{d}\boldsymbol{\tau}_N \\ &\leqslant \left\langle \Psi_\rho | \hat{T} | \Psi_\rho \right\rangle, \end{aligned} \tag{3.74}$$

thus yielding a lower bound for the kinetic energy of an N–electron wavefunction $\Psi_\rho \in \mathcal{W}_N$ in terms of its corresponding N–representable density $\rho \in \mathcal{I}_N$. The expression for this lower bound is better known as the *von Weizäcker kinetic energy functional*[75].

Definition 3.4: The von Weizäcker Kinetic Energy

For an N–representable density $\rho \in \mathcal{I}_N$, the von Weizäcker kinetic energy functional $T_{\mathrm{W}}(\rho)$ is a lower bound to the kinetic energy of all N–electron wavefunctions that yield the same density,

$$T_{\mathrm{W}}(\rho) \leqslant \inf_{\Psi \to \rho} \left\langle \Psi_\rho | \hat{T} | \Psi_\rho \right\rangle < \infty \ , \ \rho \in \mathcal{I}_N \tag{3.75}$$

where $T_{\mathrm{W}}(\rho)$ has the functional form

$$T_{\mathrm{W}}(\rho) = \frac{1}{2} \int |\nabla \rho^{\frac{1}{2}}(\mathbf{r})|^2 \mathrm{d}\mathbf{r} = \frac{1}{8} \int \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{3.76}$$

[v] The function $q(x) \to 0$ as $x \to -\infty$ and $q(x) \to 1$ as $x \to \infty$, whilst $q'(x) > 0$ hence it is a strictly increasing function.

Whilst this shows that the kinetic energy computed from an N–representable density $\rho \in \mathcal{I}_N$ is finite, it does not necessarily prove that the kinetic energy of a Slater determinant wavefunction Φ_ρ yielding the density $\rho \in \mathcal{L}^1_N$ is finite. Instead, this can be shown using a method similar to that used previously to demonstrate that each $\rho \in \mathcal{L}^1_N$ can be obtained from a Slater determinant. The kinetic energy of a Slater determinant Φ_ρ constructed from N orbitals of form (3.64) can be written as

$$T_{\mathrm{SD}}(\rho) = \frac{1}{2}\sum_{k=1}^{N}\langle\nabla\phi_k|\nabla\phi_k\rangle = \frac{1}{2}\sum_{k=1}^{N}\int|\nabla\phi_k(\mathbf{r})|^2\mathrm{d}\mathbf{r}, \tag{3.77}$$

using the notation of the position–dependent inner product (3.70). The derivative of an orbital (3.64) along a given Cartesian direction α is given by

$$\nabla_\alpha\phi_k(\mathbf{r}) = \frac{1}{\sqrt{N}}\left(\nabla_\alpha\rho^{\frac{1}{2}}(\mathbf{r}) + 2\pi\imath\delta_{\alpha x}(k-1)q'(x)\rho^{\frac{1}{2}}(\mathbf{r})\right)e^{2\pi\imath(k-1)q(x)}, \tag{3.78}$$

which may be substituted into the expression for the kinetic energy of the Slater determinant (3.77) to give

$$\begin{aligned} T_{\mathrm{SD}}(\rho) &= \frac{1}{2}\sum_{k=1}^{N}\frac{1}{N}\int\left(|\nabla\rho^{\frac{1}{2}}(\mathbf{r})|^2 + 4\pi^2(k-1)^2q'^2(x)\rho(\mathbf{r})\right)\mathrm{d}\mathbf{r} \\ &= \frac{1}{2}\int|\nabla\rho^{\frac{1}{2}}(\mathbf{r})|^2\mathrm{d}\mathbf{r} + \frac{2\pi^2}{N}\sum_{k=1}^{N}(k-1)^2\int q'^2(x)\rho(\mathbf{r})\mathrm{d}\mathbf{r} \\ &= T_{\mathrm{W}}(\rho) + \frac{\pi^2}{3}N(N-1)(2N-1)\int_{-\infty}^{\infty}q'^3(x)\mathrm{d}x. \end{aligned} \tag{3.79}$$

Using the Cauchy–Schwarz inequality, it can be shown that

$$\int_{-\infty}^{\infty}q'^3(x)\mathrm{d}x \leqslant \int_{-\infty}^{\infty}\frac{q''^2(x)}{q'(x)}\mathrm{d}x \tag{3.80}$$

where the second derivative of the function $q(x)$ is obtained by differentiating (3.67). The square of $q''(x)$ may also be re–arranged using the Cauchy–Schwarz inequality as

$$q''^2(x) = \left[\frac{1}{N}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial\rho(x,y,z)}{\partial x}\mathrm{d}y\,\mathrm{d}z\right]^2 \leqslant \frac{4}{N}q(x)\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial\rho^{\frac{1}{2}}(x,y,z)}{\partial x}\right)^2\mathrm{d}y\,\mathrm{d}z\right]. \tag{3.81}$$

Dividing both sides through by $q(x)$ and integrating with respect to x over all space gives the result

$$\int_{-\infty}^{\infty}\frac{q''^2(x)}{q'(x)}\mathrm{d}x \leqslant \frac{4}{N}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial\rho^{\frac{1}{2}}(x,y,z)}{\partial x}\right)^2\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \leqslant \frac{4}{N}\int|\nabla\rho^{\frac{1}{2}}(\mathbf{r})|^2\mathrm{d}\mathbf{r} \leqslant \frac{8}{N}T_{\mathrm{W}}(\rho). \tag{3.82}$$

This result may be substituted back into the inequality (3.80), which in–turn is substituted into (3.79) to give the inequality

$$T_{\mathrm{SD}}(\rho) \leqslant \left(1 + \frac{8\pi^2}{3}(N-1)(2N-1)\right)T_{\mathrm{W}}(\rho). \tag{3.83}$$

For $N \geqslant 1$, this can be approximated by the leading–order term as

$$T_{\mathrm{SD}}(\rho) = \left\langle\Phi_\rho|\hat{T}|\Phi_\rho\right\rangle \leqslant \frac{16\pi^2N^2}{3}T_{\mathrm{W}}(\rho), \tag{3.84}$$

thus implying that any density $\rho \in \mathcal{L}^1_N$ for which $T_{\mathrm{W}}(\rho) < \infty$ can be obtained from a Slater determinant wavefunction $\Phi_\rho \in \mathcal{W}_N$ and is therefore N–representable.

Definition 3.5: The Set Of N–Representable Densities

For a density $\rho \in \mathcal{L}^1_N$ yielded by both a general wavefunction Ψ_ρ and a Slater determinant Φ_ρ, the von Weizäcker kinetic energy functional $T_{\mathrm{W}}(\rho)$ is bound by the inequality

$$\frac{3}{16\pi^2N^2}\left\langle\Phi_\rho|\hat{T}|\Phi_\rho\right\rangle \leqslant T_{\mathrm{W}}(\rho) \leqslant \left\langle\Psi_\rho|\hat{T}|\Psi_\rho\right\rangle, \tag{3.85}$$

hence only densities in $\mathcal{L}^1_N$ for which $T_{\mathrm{W}}(\rho) < \infty$ belong to the set of N–representable densities,

$$\mathcal{I}_N = \left\{\rho \in \mathrm{L}^1(\mathbb{R}^3) \;\middle|\; \rho \geqslant 0\,,\ \int\rho(\mathbf{r})\mathrm{d}\mathbf{r} = N\,,\ T_{\mathrm{W}}(\rho) < \infty\right\}. \tag{3.86}$$

3.3.5 Levy–Lieb Constrained Search Functional

Having characterised the space of Coulomb potentials and N–representable densities, the generalised Rayleigh–Ritz minimisation in (3.51) can be applied to search only over these domains by reformulating it as a two–fold nested optimisation; an outer optimisation over all $\rho \in \mathcal{I}_N$ and an inner optimisation over all $\Psi \in \mathcal{W}_N$ that yield the density ρ,

$$\begin{aligned}\mathcal{E}_\lambda(v) = \inf_{\Psi\in\mathcal{W}_N} \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) \right| \Psi \right\rangle &= \inf_{\rho\in\mathcal{I}_N} \inf_{\Psi\to\rho} \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) \right| \Psi \right\rangle \\ &= \inf_{\rho\in\mathcal{I}_N} \left\{ \inf_{\Psi\to\rho} \left\langle \Psi \left| \hat{T} + \hat{W}_\lambda \right| \Psi \right\rangle + (v|\rho) \right\} \quad \forall v \in \mathcal{V}_C.\end{aligned} \tag{3.87}$$

With this approach, the contribution to the energy from the potential v is not required for the inner optimisation over Ψ as the value of $(v|\rho)$ will be equal for all $\Psi \to \rho$. Thus in analogy to the Hohenberg–Kohn functional (3.32), the *Levy–Lieb constrained search functional* $F_{\text{LL}} : \mathcal{I}_N \mapsto \mathbb{R}$ is defined as

$$F_{\text{LL}}(\rho) = \inf_{\Psi\to\rho} \left\langle \Psi \left| \hat{T} + \hat{W}_\lambda \right| \Psi \right\rangle \quad \forall \rho \in \mathcal{I}_N. \tag{3.88}$$

This functional represents the maximum of the kinetic + two–electron energy that can be obtained for a density ρ. It was shown in Ref. 67 that a minimising wavefunction exists for all $\rho \in \mathcal{I}_N$ hence the infimum in (3.88) can be replaced by a minimum over $\Psi \to \rho$, $\Psi \in \mathcal{W}_N$ – thus also showing that $0 < F_{\text{LL}}(\rho) < \infty$ for all $\rho \in \mathcal{I}_N$.

Theorem 3.4: The Levy–Lieb Constrained Search

The ground–state energy of a system in an external Coulomb potential can be written as a Hohenberg–Kohn variation principle over all N–representable densities,

$$\mathcal{E}_\lambda(v) = \inf_{\rho\in\mathcal{I}_N} \{F_{\text{LL}}(\rho) + (v|\rho)\} \quad \forall v \in \mathcal{V}_C, \tag{3.89}$$

where $F_{\text{LL}}(\rho)$ is the Levy–Lieb constrained search functional $F_{\text{LL}} : \mathcal{I}_N \mapsto \mathbb{R}$,

$$F_{\text{LL}}(\rho) = \min_{\Psi\to\rho} \left\langle \Psi \left| \hat{T} + \hat{W}_\lambda \right| \Psi \right\rangle > 0 \quad \forall \rho \in \mathcal{I}_N, \tag{3.90}$$

for which a minimising density always exists and which is finite and positive for all $\rho \in \mathcal{I}_N$.

The Levy–Lieb constrained search differs from the Hohenberg–Kohn variation principle in that it works over the explicitly known sets of densities and potentials, $\mathcal{I}_N$ & $\mathcal{V}_C$ respectively, whereas the Hohenberg–Kohn variation principle is defined for the unknown sets $\mathcal{A}_N$ & $\mathcal{V}_N$ of the densities and potentials respectively. As a result, the optimisation (3.89) is a search for an infimum of $\mathcal{E}_\lambda(v)$ rather than a minimum since the $v \in \mathcal{V}_C$ only ensures a finite energy and not necessarily that a ground–state wavefunction exists.

It can be demonstrated that the set of v–representable densities $\mathcal{A}_N$ is a proper subset of $\mathcal{I}_N$ by considering the ground–state of a Boron atom, which has three degenerate ground–state wavefunctions & densities, each of which are v–representable,

$$\Psi_\alpha \in \mathcal{W}_N \ , \ \rho_\alpha \in \mathcal{A}_N \ , \ \alpha = x, y, z. \tag{3.91}$$

An average of these densities may be taken and shown to satisfy the criteria for N–representability,

$$\bar{\rho} = \frac{1}{3}(\rho_x + \rho_y + \rho_z) \in \mathcal{I}_N \quad \text{since} \quad \bar{\rho}(\mathbf{r}) \geqslant 0 \ , \ \int \bar{\rho}(\mathbf{r})\mathrm{d}\mathbf{r} = N \ , \ T_{\text{W}}(\bar{\rho}) < \infty. \tag{3.92}$$

Assuming that $\bar{\rho}$ comes from an eigenstate of the same atom – $\bar{\Psi} = c_x\Psi_x + c_y\Psi_y + c_z\Psi_z$, it would follow that $N\langle\bar{\Psi}|\bar{\Psi}\rangle = \bar{\rho}$. However, expanding this in terms of individual components requires $c_\alpha^* c_\beta = \delta_{\alpha\beta}/3 \ \forall \alpha, \beta = x, y, z$, which is not satisfied by any set of c_α – hence $\bar{\rho}$ itself cannot be the ground–state density of the Boron atom. Assuming then that $\bar{\rho}$ is the ground–state density of some potential $\bar{v} \neq v + c$, where v is the potential yielding any of the ground–state densities ρ_α, this implies that

$$\mathcal{E}_\lambda(v) = F_{\text{LL}}(\rho_\alpha) + (v|\rho_\alpha) \quad \& \quad \mathcal{E}_\lambda(v) < F_{\text{LL}}(\bar{\rho}) + (v|\bar{\rho}) \quad \Longrightarrow \ F_{\text{LL}}(\bar{\rho}) > \frac{1}{3}\sum_\alpha F_{\text{LL}}(\rho_\alpha). \tag{3.93}$$

If the ground–state energy of $\bar{\rho}$ is given by $\mathcal{E}_\lambda(\bar{v}) = F_{\text{LL}}(\bar{\rho}) + (\bar{v}|\bar{\rho})$, this implies that

$$\mathcal{E}_\lambda(\bar{v}) = F_{\text{LL}}(\bar{\rho}) + (\bar{v}|\bar{\rho}) > \frac{1}{3}\sum_\alpha \left(F_{\text{LL}}(\rho_\alpha) + (\bar{v}|\rho_\alpha)\right) = \frac{1}{3}\sum_\alpha \langle\Psi_\alpha|\hat{\mathcal{H}}_\lambda(\bar{v})|\Psi_\alpha\rangle, \tag{3.94}$$

which directly contradicts the Rayleigh–Ritz variation principle, thus $\bar{\rho}$ is not the ground–state density of any potential but *is* N–representable. The set $\mathcal{A}_N$ is therefore a proper subset of $\mathcal{I}_N$: $\mathcal{A}_N \subsetneq \mathcal{I}_N$. Meanwhile, it can be shown that if $\rho \in \mathcal{A}_N$, the minimising wavefunction in (3.90) will be a ground–state wavefunction yielding ρ and Levy–Lieb functional will become equivalent to the Hohenberg–Kohn functional. It can be said therefore that the Hohenberg–Kohn functional is the restriction of the Levy–Lieb functional to the set of v–representable densities,

$$F_{\mathrm{LL}}|_{\mathcal{A}_N} = F_{\mathrm{HK}}, \tag{3.95}$$

and that the Levy–Lieb functional generalises the Hohenberg–Kohn functional from unknown domain $\mathcal{A}_N$ to the known domain $\mathcal{I}_N$.

3.3.6 Banach Spaces Of Densities & Potentials

In order to further develop this rigorous approach to DFT, it is necessary to be able to characterise the smallest Banach spaces for the density & potential that contain the sets of N–representable densities & Coulomb potentials respectively.

For the set of N–representable densities $\mathcal{I}_N$, the definition (3.86) already requires that $\rho \in \mathrm{L}^1(\mathbb{R}^3)$ for any $\rho \in \mathcal{I}_N$. The requirement for a finite von–Weizäcker kinetic energy can be re–cast in similar terms by considering its definition (3.76),

$$T_{\mathrm{W}}(\rho) = \frac{1}{2}\int |\nabla\rho^{\frac{1}{2}}(\mathbf{r})|^2 \mathrm{d}\mathbf{r} = \frac{1}{2}\||\nabla\rho^{\frac{1}{2}}|\|_2^2 < \infty \quad \Longrightarrow \quad |\nabla\rho^{\frac{1}{2}}| \in \mathrm{L}^2(\mathbb{R}^3). \tag{3.96}$$

This requirement for $|\nabla\rho^{\frac{1}{2}}|$ can be translated into one pertaining to ρ directly by first noting that, given the definition of the first–order Sobolev space in (3.56) & (3.57), the square–root of the density must belong to the first–order Sobolev space,

$$\|\rho^{\frac{1}{2}}\|_{\mathcal{H}^1_1} = \left(\rho^{\frac{1}{2}}, \rho^{\frac{1}{2}}\right)^{\frac{1}{2}}_{\mathcal{H}^1_1} = \left(\int |\rho^{\frac{1}{2}}(\mathbf{r})|^2 \mathrm{d}\mathbf{r} + \int |\nabla\rho^{\frac{1}{2}}(\mathbf{r})|^2 \mathrm{d}\mathbf{r}\right)^{\frac{1}{2}} < \infty \quad \Longrightarrow \quad \rho^{\frac{1}{2}} \in \mathcal{H}^1_1, \tag{3.97}$$

thus allowing the *Sobolev inequality* $\||\nabla f|\|_2^2 \geqslant \frac{1}{36}\|f\|_6^2$, $f \in \mathcal{H}^1_1$ to be applied as

$$\||\nabla\rho^{\frac{1}{2}}|\|_2^2 \geqslant \frac{1}{36}\|\rho^{\frac{1}{2}}\|_6^2 = \frac{1}{32}\left(\int |\rho^{\frac{1}{2}}(\mathbf{r})|^6 \mathrm{d}\mathbf{r}\right)^{\frac{2}{6}} = \frac{1}{32}\left(\int |\rho(\mathbf{r})|^3 \mathrm{d}\mathbf{r}\right)^{\frac{1}{3}} = \frac{1}{36}\|\rho\|_3. \tag{3.98}$$

It follows that a finite $T_{\mathrm{W}}(\rho)$ implies a finite $\|\rho\|_3$ for a given ρ, as $T_{\mathrm{W}}(\rho) = \frac{1}{2}\||\nabla\rho^{\frac{1}{2}}|\|_2^2 \geqslant \frac{1}{72}\|\rho\|_3$, hence for all such densities it is deduced that $\rho \in \mathrm{L}^3(\mathbb{R}^3)$, in addition to the requirement $\rho \in \mathrm{L}^1(\mathbb{R}^3)$. The set of densities which satisfy these requirements must be both a subset of L^1 and of L^3; the intersections and sums of the L^p spaces can be represented by the common notation

$$\mathrm{L}^{(\mathrm{p},\mathrm{q})} = \begin{cases} \mathrm{L}^{\mathrm{p}} \cap \mathrm{L}^{\mathrm{q}} & p \geqslant q \\ \mathrm{L}^{\mathrm{p}} + \mathrm{L}^{\mathrm{q}} & p < q \end{cases}, \tag{3.99}$$

where the vector space $\mathrm{L}^{(p,q)}$ is a Banach space with the norm

$$\|f\|_{(p,q)} = \begin{cases} \max\left(\|f\|_p, \|f\|_p\right) & p \geqslant q \\ \inf\limits_{f=f_p+f_q}\left(\|f_p\|_p + \|f_q\|_q\right) & p < q \end{cases}. \tag{3.100}$$

The *Banach space of densities* $\mathcal{X}$ can be defined using (3.99) as $\mathrm{L}^{(3,1)}$; the intersection between the finite T_{W} condition $\rho \in \mathrm{L}^3$ and the normalisation condition $\rho \in \mathrm{L}^1$,

$$\mathcal{X} = \mathrm{L}^3 \cap \mathrm{L}^1, \tag{3.101}$$

for which the set of N–representable densities is a subset $\mathcal{I}_N \subsetneq \mathcal{X}$.

Having defined the Banach space of densities $\mathcal{X}$, the Banach space of potentials may be derived from the requirement that the interaction term $(v|\rho)$ must be finite for all $\rho \in \mathcal{X}$. For this to be the case, the interaction term must be represented by a bounded linear operator in the dual space of $\mathcal{X}$,

$$T_v(\rho) = (v|\rho) \quad T_v \in \mathcal{X}^*. \tag{3.102}$$

The dual space of $\mathcal{X}^* = \left(\mathrm{L}^3 \cap \mathrm{L}^1\right)^*$ can be determined by generalising the relation between L^p and its dual described previously and applying it to $\mathrm{L}^{(p,q)}$; if $1 \leqslant p, q < \infty$ & $1 < \bar{p}, \bar{q} \leqslant \infty$, with $p^{-1} + \bar{p}^{-1} = 1$ & $q^{-1} + \bar{q}^{-1} = 1$, then $\mathrm{L}^{(p,q)*} = \mathrm{L}^{(\bar{p},\bar{q})}$ with the special case that

$$\left(\mathrm{L}^p \cap \mathrm{L}^q\right)^* = \mathrm{L}^{\bar{p}} + \mathrm{L}^{\bar{q}}. \tag{3.103}$$

The *Banach space of potentials* is thus given by

$$\mathcal{X}^* = \mathrm{L}^{(3,1)*} = \left(\mathrm{L}^3 \cap \mathrm{L}^1\right)^* = \mathrm{L}^{\frac{1}{1-1/3}} + \mathrm{L}^{\frac{1}{1-1/1}} = \mathrm{L}^{\frac{3}{2}} + \mathrm{L}^{\infty}. \tag{3.104}$$

It follows that the energy as defined with the Levy–Lieb constrained–search functional in (3.89) is finite for all $\rho \in \mathcal{X}$ and $v \in \mathcal{X}^*$, of which the N–representable densities & Coulomb potentials are respective subsets, [vi]

$$\begin{aligned} \mathcal{I}_{\mathrm{N}} &= \left\{ \rho \in \mathrm{L}^3 \cap \mathrm{L}^1 \;\middle|\; \rho \geqslant 0 \,,\; \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} = N \,,\; T_{\mathrm{W}}(\rho) < \infty \right\} \subset \mathcal{X}, \\ \mathcal{V}_{\mathrm{C}} &= \left\{ v \in \mathrm{L}^{\frac{3}{2}} + \mathrm{L}^{\infty} \;\middle|\; v(\mathbf{r}_i) = -\sum_{I}^{N_{\mathrm{C}}} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \,,\; N_{\mathrm{C}} \in \mathbb{N} \,,\; Z_I \in \mathbb{R} \right\} \subset \mathcal{X}^*. \end{aligned} \tag{3.105}$$

Definition 3.6: Banach Spaces Of The Density & Potential

The sets of densities ρ and external potentials v that yield a finite expectation value for the energy are characterised as the dual Banach spaces

$$\rho \in \mathcal{X} = \mathrm{L}^3 \cap \mathrm{L}^1 \qquad\qquad v \in \mathcal{X}^* = \mathrm{L}^{\frac{3}{2}} + \mathrm{L}^{\infty} \tag{3.106}$$

of which the set of N–representable densities $\mathcal{I}_{\mathrm{N}}$ & Coulomb potentials $\mathcal{V}_{\mathrm{C}}$ are respective subsets.

The development of the Hohenberg–Kohn variational principle discussed here gives a much more useful form of the universal density functional, the Levy–Lieb functional, sets of ρ and v for which over which it may be evaluated are well characterised. However the form of the Levy–Lieb functional is such that it may have more than one solution; it is not *convex* in the density. Further developments by Lieb[67], resulting in a functional that is convex and well defined, will be discussed in detail in the next section as it is pivotal to the work presented in later chapters.

3.4 Lieb's Convex–Conjugate Functional

In this section the development of DFT beyond the Levy–Lieb constrained search is discussed, leading to the definition of the Lieb functional[67]; an important development as the Lieb functional is finite over well–characterised spaces of the density & potential and always yields a unique solution. To explain this, it is first necessary to present an overview of some additional mathematical concepts from *convex analysis*, a definitive study of which is presented by Rockafellar in Ref. 76.

3.4.1 Convex Sets

This section begins by expanding on the theory set out in subsections 3.3.1 & 3.3.2 to introduce the concepts of open & closed sets. In a metric space, equipped with the metric d, an *open ball* with centre $\mathbf{x} \in \mathcal{X}$ and radius $r > 0$ can be defined by

$$B_r(\mathbf{x}) = \left\{ \mathbf{y} \in \mathcal{X} \;\middle|\; d(\mathbf{x}, \mathbf{y}) < r \right\}. \tag{3.107}$$

Given this definition, the set $\mathcal{A} \subset \mathcal{X}$ is defined as being *open* in $\mathcal{X}$ if, for all $\mathbf{x} \in \mathcal{A}$ there exists an $r > 0$ such that $B_r(x) \subset \mathcal{A}$. Conversely, the set $\mathcal{A} \subset \mathcal{X}$ can be identified as being *closed* in $\mathcal{X}$ if the complement $\mathcal{X} \setminus \mathcal{A}$ is open. With each subset $\mathcal{A}$ of the normed space $\mathcal{X}$ can be associated the following characteristics:

(i) The *interior* $\mathrm{int}(\mathcal{A}) = \left\{ \mathbf{x} \in \mathcal{X} \;\middle|\; B_r(\mathbf{x}) \in \mathcal{A} \text{ for some } r > 0 \right\}$, i.e. $\mathrm{int}(\mathcal{A})$ is the largest open set contained in $\mathcal{A}$,

(ii) The *closure* $\mathrm{cl}(\mathcal{A}) = \left\{ \mathbf{x} \in \mathcal{X} \;\middle|\; B_r(\mathbf{x}) \cap \mathcal{A} \text{ for all } r > 0 \right\}$, i.e. $\mathrm{cl}(\mathcal{A})$ is the smallest closed set containing $\mathcal{A}$,

[vi] It can be shown that the set of Coulomb potentials $\mathcal{V}_{\mathrm{C}}$ for which the Levy–Lieb constrained search is defined is a proper subset of the Banach space of potentials by decomposing the Coulomb operator $v_{\mathrm{c}}(\mathbf{r}) = -|\mathbf{r}|^{-1}$ into the sum of a short–range and a long–range component, $v_{\mathrm{c,sr}}(\mathbf{r}) = -(e^{-|\mathbf{r}|})|\mathbf{r}|^{-1}$ and $v_{\mathrm{c,lr}}(\mathbf{r}) = -(1 - e^{-|\mathbf{r}|})|\mathbf{r}|^{-1}$ respectively. It can be shown that $v_{\mathrm{c,sr}} \in \mathrm{L}^p$, $1 \leqslant p < 3$ and $v_{\mathrm{c,lr}} \in \mathrm{L}^{\infty}$ respectively, hence $v_{\mathrm{c}} \in (\mathrm{L}^p + \mathrm{L}^{\infty})$, $1 \leqslant p < 3$. The set of Coulomb potentials $\mathcal{V}_{\mathrm{C}} \subsetneq (\mathrm{L}^p + \mathrm{L}^{\infty})$, $1 \leqslant p < 3$ is therefore contained in the dual of the sets of densities $\mathrm{L}^{\bar{p}} \cap \mathrm{L}^1$, $\frac{3}{2} < \bar{p} \leqslant +\infty$, in which the Banach space of densities is included.

(iii) The *boundary* $\mathrm{bd}(\mathcal{A}) = \mathrm{cl}(\mathcal{A}) \setminus \mathrm{int}(\mathcal{A})$,

(iv) The *exterior* $\mathrm{ext}(\mathcal{A}) = \mathcal{X} \setminus \mathrm{cl}(\mathcal{A})$.

Finally, a set $\mathcal{A} \subset \mathcal{X}$ can be said to be *dense* in the set $\mathcal{B} \subset \mathcal{X}$ if $\mathcal{B} \subset \mathrm{cl}(\mathcal{A})$, meaning that each open set in $\mathcal{B}$ contains some element in $\mathcal{A}$.

The subset $\mathcal{A}$ of the vector space $\mathcal{V}$ is a *convex set* if, for each $\mathbf{x}, \mathbf{y} \in \mathcal{A}$ the relation

$$\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{A} \qquad 0 \leqslant \alpha \leqslant 1 \tag{3.108}$$

This can be illustrated with the two–dimensional shapes shown in Figure 3.2; the shape on the left represents a convex set as the line segment joining any two points within its area is contained entirely within the shape. This is not the case for the shape on the right, which hence represents a nonconvex set. It can be shown from the basic definition

Figure 3.2: A schematic illustrating the difference between convex and nonconvex sets.

(3.108) that all linear subspaces of the vector space $\mathcal{V}$ are convex and that, if $\mathcal{A}_1$ & $\mathcal{A}_2$ are convex sets, then their combinations $\mathcal{A}_1 \cap \mathcal{A}_2$ & $\mathcal{A}_1 + \mathcal{A}_2$ must also be convex.

A *convex combination* of elements $\mathbf{x}_i$ of the vector space $\mathcal{V}$ is defined as a finite linear combination of the elements with non–negative coefficients that sum to unity,

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i \qquad \sum_i \alpha_i = 1 \qquad \alpha_i \geqslant 0, \tag{3.109}$$

in which $\mathbf{y}$ represents an average of the elements $\mathbf{x}_i$ weighted by coefficients α_i. Clearly, if there are only two elements then it may be written in the form of (3.108) as $\mathbf{y} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ where $0 \leqslant \alpha \leqslant 1$, hence it follows that the set $\mathcal{A} \subset \mathcal{V}$ is only convex if it contains all of its convex combinations.

The *convex hull* $\mathrm{co}(\mathcal{S})$ of the set $\mathcal{S} \subset \mathcal{V}$ is the smallest convex set in $\mathcal{V}$ which contains $\mathcal{S}$, i.e. it is the intersection of all convex sets that contain $\mathcal{S}$, consisting of all convex combinations of the elements in $\mathcal{S}$.

3.4.2 Convex Functions

A function $f : \mathcal{X} \mapsto \mathbb{R}$ can be described as being *convex* if the interpolation characteristic inequality is satisfied,

$$\alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \geqslant f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \qquad \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \qquad 0 < \alpha < 1, \tag{3.110}$$

meaning that linear interpolation along a convex function always yields an estimate that is greater than or equal to the value of the function at that point; a characteristic demonstrated in Figure 3.3. If the function $f : \mathcal{X} \mapsto \mathbb{R}$ is convex, then by definition its additive inverse $-f : \mathcal{X} \mapsto \mathbb{R}$ is *concave*. For *strictly* convex functions, the interpolant is always greater than the function, hence the $\geqslant$ operator in (3.110) is replaced by $>$ for a strictly convex function f. An important corollary to this is that, for a convex function every local minimiser is also a global minimiser. This can be seen by considering a global minimiser $\mathbf{x}_0$ of the convex function f and an arbitrary point $\mathbf{x}_1 \neq \mathbf{x}_0$ in the domain of f, for which it is clear that $f(\mathbf{x}_1) > f(\mathbf{x}_0)$. It follows that, for $0 < \alpha < 1$, the convexity of f implies

$$f(\mathbf{x}_1) = \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_1) > \alpha f(\mathbf{x}_0) + (1-\alpha) f(\mathbf{x}_1) \geqslant f(\alpha\mathbf{x}_0 + (1-\alpha)\mathbf{x}_1), \tag{3.111}$$

which may be re–arranged to give $f(\mathbf{x}_1) > f(\mathbf{x}_1 + \alpha(\mathbf{x}_0 - \mathbf{x}_1))$, implying that an infinitesimal movement from $\mathbf{x}_1$ towards $\mathbf{x}_0$ is accompanied by a decrease in the value of the function, hence $\mathbf{x}_1$ cannot be a local minimiser if $f(\mathbf{x}_1) > f(\mathbf{x}_0)$. The minimisers of a convex function constitute a convex set, however if the function is strictly convex then that set is a singleton and the minimiser is unique.

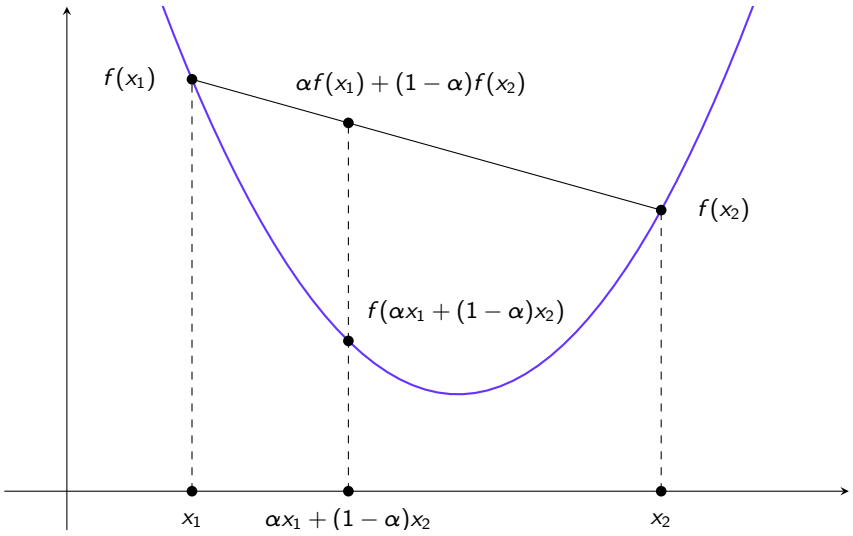

Figure 3.3: A convex function with a linear interpolant between two points, showing that the value of the function in the domain between x_1 and x_2 is always less than or equal to that of the linear interpolant.

Extended Functions

If a real–valued convex function f is defined on the convex subset $\mathcal{A} \subset \mathcal{X}$ as $f : \mathcal{A} \mapsto \mathbb{R}$, a corresponding *extended function*[vii] may be defined over the entire set $\mathcal{X}$ as

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \mathbf{x} \in \mathcal{A} \\ +\infty & \mathbf{x} \in \mathcal{X} \setminus \mathcal{A} \end{cases}. \tag{3.112}$$

The *effective domain* of this extended function is given by

$$\mathrm{dom}(\tilde{f}) = \{\mathbf{x} \in \mathcal{X} \mid \tilde{f}(\mathbf{x}) < +\infty\} \tag{3.113}$$

thus completing a more general description of the function f over $\mathcal{X}$. The effective domain of a convex function can be shown to always be a convex set; if $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{dom}(f)$, where f is convex, by definition $f(\mathbf{x}_1), f(\mathbf{x}_2) < +\infty$. Substituting these values into (3.110) with $0 < \alpha < 1$ yields the result $f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) < +\infty$, hence $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \mathrm{dom}(f)$. A convex function that is finite on $\mathrm{dom}(f)$ but $+\infty$ everywhere outside of $\mathrm{dom}(f)$ is termed a *proper convex function*.

Epigraph & Hypograph

It is useful here to introduce three additional properties which can be used to characterise a function $f : \mathcal{X} \mapsto \tilde{\mathbb{R}}$:

(i) The *epigraph* $\mathrm{epi}(f) = \{(\mathbf{x}, \alpha) \in \mathcal{X} \times \mathbb{R} \mid f(\mathbf{x}) \leqslant \alpha\}$,

(ii) The *graph* $\mathrm{gr}(f) = \{(\mathbf{x}, \alpha) \in \mathcal{X} \times \mathbb{R} \mid f(\mathbf{x}) = \alpha\}$,

(iii) The *hypograph* $\mathrm{hyp}(f) = \{(\mathbf{x}, \alpha) \in \mathcal{X} \times \mathbb{R} \mid f(\mathbf{x}) \geqslant \alpha\}$.

The epigraph and hypograph of a function f on the real axis are illustrated in Figure 3.4. It can be readily deduced that a function $f : \mathcal{X} \mapsto \tilde{\mathbb{R}}$ is convex if and only if its epigraph $\mathrm{epi}(f) \subset \mathcal{X} \times \mathbb{R}$ is a convex set. Given the definition of a convex function in these terms and the definition of the convex hull of a set in the previous subsection, it is possible to define for a function f its convex hull $\mathrm{co}(f)$ as the function with epigraph $\mathrm{co}(\mathrm{epi}(f))$. The convex hull $\mathrm{co}(f)$ is sometimes described as the *largest convex minorant* to f. This can be seen for a nonconvex function on the real axis in Figure 3.5.

Continuity & Semi–Continuity

A function $f : \mathcal{X} \mapsto \mathbb{R}$ can be formally defined as *continuous* at a point $\mathbf{x} \in \mathcal{X}$ if, for each $k, K \in \mathbb{R}$ where $k < f(\mathbf{x}) < K$, there exists a *neighbourhood* $\mathcal{U}$ of $\mathbf{x}$ such that $k < f(\mathcal{U}) < K$.[viii]

[vii] Such functions are referred to as 'extended' as they have values in the set of *extended real numbers* $\tilde{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$.

[viii] A neighbourhood of a point $\mathbf{x} \in \mathcal{X}$ is a subset $\mathcal{U}$ of $\mathcal{X}$ which includes an open set $\mathcal{V}$ containing $\mathbf{x}$, i.e. $\mathbf{x} \in \mathcal{V} \subset \mathcal{U} \subset \mathcal{X}$. In practice, this can be a set of points similar to $\mathbf{x}$ but with incrementally different values.

Figure 3.4: An illustration of the epigraph and hypograph of a 2D function, represented by the shaded areas in the left and right–hand plots respectively.

Figure 3.5: An illustration of a nonconvex function with its epigraph (left) and the convex hull of the function & corresponding epigraph (right).

A convex function may not necessarily be continuous; if a function is not continuous, it may still be possible to characterise it as *lower semi–continuous*. A function $f : \mathcal{X} \mapsto \tilde{\mathbb{R}}$ is lower semi–continuous at a point $\mathbf{x} \in \mathcal{X}$ if, for each $k \in \mathbb{R}$ where $k < f(\mathbf{x})$, there exists a neighbourhood $\mathcal{U}$ of $\mathbf{x}$ such that $k < f(\mathcal{U})$. In practice, this means that if a function is lower semi–continuous at $\mathbf{x}$, then the value of f for elements of $\mathcal{X}$ close to $\mathbf{x}$ are close to *or greater than* $f(\mathbf{x})$. Equally, a function f may be described as *upper semi–continuous* at point $\mathbf{x}$ if $-f$ is lower semi–continuous at $\mathbf{x}$. These concepts are illustrated for a simple 2D function in Figure 3.6. One of the defining characteristics of a

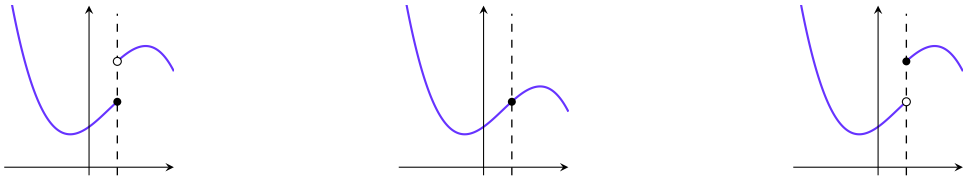

Figure 3.6: An illustration of the concepts of continuity (center), lower semi–continuity (left) and upper semi–continuity (right) as applied to functions. The filled circles indicate that a value exists for the line segment at the given point, whereas an unfilled circle indicates that no value exists for the line segment at the given point.

lower semi–continuous function is that its epigraph comprises a closed set. It follows that the *lower semi–continuous hull* $\bar{f}$ of the function f is that which has an epigraph equal to the closure of the epigraph of f; $\text{epi}(\bar{f}) = \text{cl}(\text{epi}(f))$, in other words it is the largest lower semi–continuous minorant to f. Taking this definition along with that for the convex hull, it is possible to define a lower semi–continuous convex hull - the *closed convex hull* of the function f as $\bar{\text{co}}(f) = \bar{\text{co}}(f)$, which may be interpreted as a lower semi–continuous convex *regularisation* of f.

Conjugate Functions

A function may be defined in terms of *supporting lines*, each of which intersect the function at one point and nowhere exceeds the value of the function, as demonstrated in Figure 3.7. The slope of a supporting line at $(\mathbf{x}, f(\mathbf{x}))$ is defined as the *subgradient* of f at $\mathbf{x}$; if the subgradient is unique at $\mathbf{x}$ then it is simply the derivative of f at that point. Any point on f at which the subgradient is zero can be identified as the global minimum of f.

The graph of a convex function $f : \mathcal{X} \mapsto \mathbb{R}$ has at least one supporting line for all $\mathbf{x} \in \mathcal{X}$, each of which takes the form of a *continuous affine function*

$$h(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x}) - g(\mathbf{y}) \tag{3.114}$$

where $\mathbf{y}$ is the slope of the supporting line at $\mathbf{x}$ and $-g(\mathbf{y})$ is its intersection with the function axis. At a given $\mathbf{x}$, $f(\mathbf{x})$

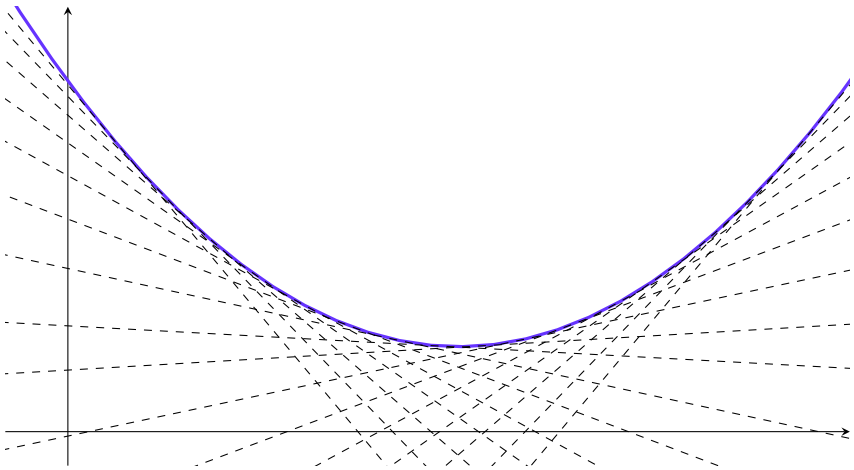

Figure 3.7: A convex function with a set of supporting lines shown.

is equal to the largest supporting line, $f(\mathbf{x}) = \max_{\mathbf{y}} h(\mathbf{x}, \mathbf{y})$, hence a convex function f can always be written as the *pointwise supremum* of supporting lines,

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{X}^*} \{(\mathbf{y}, \mathbf{x}) - g(\mathbf{y})\} \qquad \mathbf{x} \in \mathcal{X}. \tag{3.115}$$

If the function is a *closed convex function* - a proper convex function that is lower semi–continuous thus having a closed convex epigraph, it can be expressed as the pointwise supremum of a set of *closed affine* functions,

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{A}} \{(\mathbf{y}, \mathbf{x}) - g(\mathbf{y})\} \qquad \mathbf{x} \in \mathcal{X} \quad \mathcal{A} \subset \mathcal{X}^* \tag{3.116}$$

The set of all proper closed convex functions on $\mathcal{X}$ can be denoted $\Gamma_0(\mathcal{X})$, which is generalised to extended value functions as $\Gamma(\mathcal{X}) = \Gamma_0(\mathcal{X}) \cup \{+\infty, -\infty\}$.

It can be seen from (3.115) that the expression for the function f in terms of its supporting lines can be re–arranged to yield an expression for the auxiliary function g as

$$\begin{aligned} f(\mathbf{x}) = \sup_{\mathbf{y}} \{(\mathbf{y}, \mathbf{x}) - g(\mathbf{y})\} \quad &\implies \quad f(\mathbf{x}) \geqslant (\mathbf{y}, \mathbf{x}) - g(\mathbf{y}) \\ &\implies \quad g(\mathbf{y}) \geqslant (\mathbf{y}, \mathbf{x}) - f(\mathbf{x}) \quad \implies \quad g(\mathbf{y}) = \sup_{\mathbf{x}} \{(\mathbf{y}, \mathbf{x}) - f(\mathbf{x})\}. \end{aligned} \tag{3.117}$$

The function g can be identified as the *conjugate function* to f; the transformation between a function and its conjugate is known as a *Legendre–Fenchel transform*.

Definition 3.7: The Legendre–Fenchel Transform

For any function $f : \mathcal{X} \mapsto \mathbb{R}$, a conjugate f^* & biconjugate f^{**} can be defined by application of the Legendre–Fenchel transform as

$$\textit{Conjugate}: \qquad f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathcal{X}} \{(\mathbf{y}, \mathbf{x}) - f(\mathbf{x})\} \tag{3.118}$$

$$\textit{Biconjugate}: \qquad f^{**}(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{X}^*} \{(\mathbf{y}, \mathbf{x}) - f^*(\mathbf{y})\} \tag{3.119}$$

in which $\mathcal{X}$ & $\mathcal{X}^*$ are dual vector spaces.

An important characteristic of the conjugate f^* & biconjugate f^{**} functions are that they are always closed convex, given that they are effectively defined by the pointwise supremum of supporting lines in (3.118) & (3.119) respectively. It can be shown using *Fenchel's inequality* $f(\mathbf{x}) + f^*(\mathbf{x}) \geqslant (\mathbf{y}, \mathbf{x})$ that the biconjugate function is in fact the *lower semi–continuous hull* - or the *convex envelope* - to f since $f(\mathbf{x}) \geqslant f^{**}(\mathbf{x})$; the corollary to this is the *Fenchel–Moreau Biconjugate Theorem*.

Definition 3.8: Fenchel–Moreau Biconjugate Theorem

A function $f : \mathcal{X} \mapsto \mathbb{R}$ is equal to its biconjugate f^{**} if and only if f is a closed convex function;

$$f^{**}(\mathbf{x}) = f(\mathbf{x}) \text{ if and only if } f \in \Gamma(\mathcal{X}) \tag{3.120}$$

Conjugation thus establishes a bijective mapping between the sets of closed convex functions $\Gamma(\mathcal{X})$ & $\Gamma^*(\mathcal{X}^*)$ over the dual spaces of $\mathcal{X}$ & $\mathcal{X}^*$ respectively. A pair of functions $f \in \Gamma(\mathcal{X})$ & $g \in \Gamma^*(\mathcal{X}^*)$ that are each others conjugates are said to be dual functions, in which f can be fully reconstructed from g and *vice versa*.

It is important to note that, unless f is closed convex, it cannot be fully recovered from its conjugate f^*, only its convex envelope.

For convenience, it is also possible to define a *concave conjugate* in a similar way to that given in Definition 3.7 for the convex conjugate. The concave conjugate is denoted

$$f^\circ(\mathbf{y}) = -f^*(-\mathbf{y}) \tag{3.121}$$

which results in the accompanying *skew–conjugate* relations

$$f^\circ(\mathbf{y}) = \inf_{\mathbf{x}\in\mathcal{X}} \{f(\mathbf{x}) + (\mathbf{y}, \mathbf{x})\} \tag{3.122}$$

$$f^{**}(\mathbf{x}) = \sup_{\mathbf{y}\in\mathcal{X}^*} \{f^\circ(\mathbf{y}) - (\mathbf{y}, \mathbf{x})\} \tag{3.123}$$

in which the duality between the closed convex functions on $\mathcal{X}$ and the closed concave functions on $\mathcal{X}^*$ forms the bijective mapping $\Gamma(\mathcal{X}) \Longleftrightarrow -\Gamma^*(\mathcal{X}^*)$.

Now that the essential concepts of convex sets, convex functions and conjugate functions have been established, the application of this to DFT to further generalise the Levy–Lieb functional will be considered in the following subsections.

3.4.3 Ensemble State Densities

In the discussion thus far, it has been assumed that densities pertain to *pure–state* wavefunctions of a system, meaning those described by a single normalised N–electron wavefunction. However, for many systems such a wavefunction is a poor representation of their actual electronic state; for these systems, an *ensemble* wavefunction is more appropriate.

An *ensemble density matrix* γ can be constructed as simply a convex combination of pure–state density matrices[ix]

$$\gamma = \sum_i \lambda_i \left|\Psi_i\right\rangle \left\langle\Psi_i\right| \quad , \quad \lambda_i \geqslant 0 \quad , \quad \sum_i \lambda_i = 1 \quad , \quad \Psi_i \in \mathcal{W}_N. \tag{3.124}$$

The ensemble density matrix effectively represents the probability distribution for $\{\Psi_i\}$, where λ_i is the probability that the system occupies a given N–electron state $\Psi_i \in \mathcal{W}_N$. This generalisation provides the means by which systems with degenerate states, such as that described for the Boron atom in subsection 3.3.5, may be properly represented. The ground–state energy of an ensemble state is obtained by generalisation of the Rayleigh–Ritz variation principle to *canonical ensembles*.

Definition 3.9: Canonical–Ensemble Rayleigh–Ritz Variation Principle

For every $v \in \mathcal{X}^*$, the canonical–ensemble ground–state energy is defined as

$$\mathcal{E}_\lambda(v) = \inf_{\gamma\in\mathcal{K}_N} \operatorname{tr} \gamma \hat{\mathcal{H}}_\lambda(v) \tag{3.125}$$

where $\mathcal{K}_N$ is the set all admissible ensemble density matrices,

$$\mathcal{K}_N = \left\{ \gamma = \sum_i \lambda_i \left|\Psi_i\right\rangle \left\langle\Psi_i\right| \;\middle|\; \lambda_i \geqslant 0 \,,\; \sum_i \lambda_i = 1 \,,\; \Psi_i \in \mathcal{W}_N \right\}. \tag{3.126}$$

[ix] The density matrix operator for a pure–state wavefunction Ψ is given by $\hat{\rho} = \left|\Psi\right\rangle \left\langle\Psi\right|$, hence the expectation value of an observable $\hat{O}$ in a system with pure–state density matrix ρ can be written as $\operatorname{tr} \rho\hat{O}$.

Although the ground-state energy can always be defined as the greatest lower bound in either (3.7) or (3.125), the formulation in terms of ensembles is more flexible, allowing for mixed-state solutions. This extra flexibility is important to establish correspondence between the optimisers in the Rayleigh–Ritz variation principle commonly used in *ab initio* theory and the Hohenberg–Kohn variation principle used in DFT[77].

Given the expression (3.125), the Levy–Lieb constrained search can be generalised to ensemble density matrices by analogy to (3.89) to yield

$$\mathcal{E}_\lambda(v) = \inf_{\rho\in\mathcal{I}_N}\left\{\inf_{\gamma\to\rho}\operatorname{tr}\gamma\hat{\mathcal{H}}_\lambda(0) + (v|\rho)\right\} \qquad v\in\mathcal{X}^*, \tag{3.127}$$

for which a generalised form of the Levy–Lieb constrained search functional is defined,

$$F_{\mathrm{DM}}(\rho) = \inf_{\gamma\to\rho}\operatorname{tr}\gamma\hat{\mathcal{H}}_\lambda(0), \tag{3.128}$$

otherwise known as the *Lieb density–matrix constrained–search functional*. An important characteristic that distinguishes F_{DM} from the pure–state form F_{LL} is that, whilst the Levy–Lieb constrained search functional is not convex in ρ, shown in (3.93), the Lieb density–matrix constrained–search functional *is* convex in ρ. This can be seen by considering two densities $\rho_1,\rho_2\in\mathcal{I}_N$ with minimising density matrices $\gamma_1,\gamma_2\in\mathcal{K}_N$ respectively; for the convex combination $\bar{\rho}=\alpha\rho_1+(1-\alpha)\rho_2\in\mathcal{I}_N$ with $0<\alpha<1$, F_{DM} is evaluated as

$$\begin{aligned} F_{\mathrm{DM}}(\bar{\rho}) = \operatorname{tr}\gamma_{\bar{\rho}}\hat{\mathcal{H}}_\lambda(0) &\leqslant \operatorname{tr}\left(\alpha\gamma_1+(1-\alpha)\gamma_2\right)\hat{\mathcal{H}}_\lambda(0) \\ &= \alpha\operatorname{tr}\gamma_1\hat{\mathcal{H}}_\lambda(0) + (1-\alpha)\operatorname{tr}\gamma_2\hat{\mathcal{H}}_\lambda(0) \\ &= \alpha F_{\mathrm{DM}}(\rho_1) + (1-\alpha)F_{\mathrm{DM}}(\rho_2). \end{aligned} \tag{3.129}$$

It follows therefore that F_{DM} is a lower–bound to F_{LL}, $F_{\mathrm{DM}}\leqslant F_{\mathrm{LL}}$, becoming an equality whenever ρ is pure–state representable. Importantly, $F_{\mathrm{LL}}(\rho)$ gives the same ground–state energy $\mathcal{E}_\lambda(v)$ as $F_{\mathrm{DM}}(\rho)$ in the Hohenberg–Kohn variation principle for any given potential $v\in\mathcal{X}^*$, the only difference being that the minimising densities for $F_{\mathrm{LL}}(\rho)$ are always pure states.

3.4.4 The Lieb Functional

The final generalisation of the Hohenberg–Kohn functional to be discussed here is that derived by Lieb[67] using the theory of convex conjugate functions.

It is readily shown that, by virtue of the linearity of the Hamiltonian in the potential – described in subsection 3.2.1 – and application of the Rayleigh–Ritz variation principle (3.7) that the ground–state energy for a given external potential $v\in\mathcal{X}^*$ is concave in the potential. Given first the identity

$$\hat{\mathcal{H}}_\lambda(\alpha v_1+(1-\alpha)v_2) = \alpha\hat{\mathcal{H}}_\lambda(v_1)+(1-\alpha)\hat{\mathcal{H}}_\lambda(v_2) \qquad v_1,v_2\in\mathcal{X}^* \quad 0<\alpha<1, \tag{3.130}$$

the application of the Rayleigh–Ritz variation principle to both sides yields the result

$$\begin{aligned} \mathcal{E}_\lambda(\alpha v_1+(1-\alpha)v_2) &= \inf_\Psi\left\{\alpha\langle\Psi|\hat{\mathcal{H}}_\lambda(v_1)|\Psi\rangle + (1-\alpha)\langle\Psi|\hat{\mathcal{H}}_\lambda(v_2)|\Psi\rangle\right\} \\ &\geqslant \alpha\inf_\Psi\langle\Psi|\hat{\mathcal{H}}_\lambda(v_1)|\Psi\rangle + (1-\alpha)\inf_\Psi\langle\Psi|\hat{\mathcal{H}}_\lambda(v_2)|\Psi\rangle \\ &\geqslant \alpha\mathcal{E}_\lambda(v_1)+(1-\alpha)\mathcal{E}_\lambda(v_2) \end{aligned} \tag{3.131}$$

demonstrating that the ground–state energy is concave in v. It can further be shown that the ground–state energy as defined is continuous on $\mathcal{X}^*$[67], hence it can be characterised as a *closed concave* function on the Banach space $\mathcal{X}^*$. Accordingly, there exists a corresponding conjugate function that is closed–convex on the dual space to $\mathcal{X}^*$; this is the *Lieb functional* $F_{\mathrm{L}}:\mathcal{X}\mapsto\mathbb{R}$ and the pair of skew–conjugate relations (3.122) & (3.123) represent the Hohenberg–Kohn & Lieb variation principles respectively.

Theorem 3.5: Lieb's Convex–Conjugate Functional

For $v \in \mathcal{X}^*$ & $\rho \in \mathcal{X}$, the ground–state energy $\mathcal{E}_\lambda$ and the Lieb functional F_L are defined by the following variation principles

$$\textbf{Hohenberg–Kohn Variation Principle:} \qquad \mathcal{E}_\lambda(v) = \inf_{\rho \in \mathcal{X}} \{F_\mathrm{L}(\rho) + (v|\rho)\} \tag{3.132}$$

$$\textbf{Lieb Variation Principle:} \qquad F_\mathrm{L}(\rho) = \sup_{v \in \mathcal{X}^*} \{\mathcal{E}_\lambda(v) - (v|\rho)\} \tag{3.133}$$

in which $\mathcal{E}_\lambda$ and F_L are closed–concave and closed–convex functions on the dual spaces of $\mathcal{X}^*$ and $\mathcal{X}$ respectively, where $\mathcal{X}^{**} \equiv \mathcal{X}$ since it is a reflexive Banach space.

As a result of this construction, the Lieb functional is lower semi–continuous and always convex; it is in fact the closed convex hull to all *admissible* density functionals – that is, those that yield the correct energy. By definition, the minima of the Lieb functional are therefore known to be global minima.

The Lieb functional is identical to the Hohenberg–Kohn and Levy–Lieb functionals for pure, non–degenerate ground–state densities, but only the Lieb and Levy–Lieb functionals are defined for other densities. The Lieb functional can be identified as the convex envelope to the Levy–Lieb functional, meaning that it will always give a unique solution whereas this cannot be ensured for the Levy–Lieb functional. It is because of these properties that the Lieb functional the basis for the computational procedures used in this work, a detailed discussion of which is presented in Chapter 5.

3.5 Kohn–Sham Theory

In the evolution of density functional theory from Hohenberg & Kohn's first proposal to Lieb's formulation described thus far, mathematical analysis has been used to properly characterise the sets of admissible densities and potentials and generalise the form of the universal density functional such that it is finite and has only global minimisers for all admissible densities.

However, there remains a problem with the original Hohenberg–Kohn theory unaddressed by these developments; the explicit functional form of the universal density functional is unknown and it must be approximated. Without a reliable way of approximating the universal density functional, the scope for application of DFT would remain very limited.

A solution to this problem was first set out by Kohn & Sham in Ref. 65, where a self–consistent field approach with similarities to Hartree–Fock theory was proposed, which has since become the basis for almost all practical applications of DFT. The Kohn–Sham formulation of DFT (KS DFT) will be described briefly in the present section.

3.5.1 A Model Non–Interacting System

In the previous section, the Levy–Lieb functional was defined by (3.88) for a general system with two–electron interaction scaled by coefficient λ and has the form of an expectation value of the kinetic energy + two–electron operators, $\hat{T} + \hat{W}_\lambda$. However, as seen for Hartree–Fock theory, the inclusion of the two–electron term $\hat{W}_\lambda$ precludes an analytical solution to the electronic Schrödinger Equation for many–electron systems and thus prevents exact analytical forms of the universal density functional from being obtained for a many–electron system.

Considering instead a theoretical model system for which $\lambda = 0$, the Levy–Lieb functional would only be an expectation value over the kinetic energy operator since the two–electron operator vanishes at zero interaction strength, in other words the universal density functional would simply be the *noninteracting kinetic energy functional* $T_\mathrm{s} : \mathcal{I}_N \mapsto \mathbb{R}$,

$$F_\mathrm{LL}^{\lambda=0}(\rho) = \inf_{\Psi_0 \to \rho} \left\langle \Psi_0 \middle| \hat{T} \middle| \Psi_0 \right\rangle = T_\mathrm{s}(\rho) \qquad \forall \rho \in \mathcal{I}_N. \tag{3.134}$$

For such a system, the exact wavefunction takes the form of a single Slater determinant and may be fully represented in terms of one–electron orbitals. It therefore follows that re–introducing one–electron spinorbitals ϕ described in subsection 1.2.1 would allow the non–interacting kinetic energy functional to be evaluated exactly as

$$T_\mathrm{s}(\rho) = \min_{\sum_i |\phi_i|^2 \to \rho} \sum_i \left\langle \phi_i \middle| -\frac{1}{2}\nabla_i^2 \middle| \phi_i \right\rangle, \tag{3.135}$$

where the density at can be evaluated from the spinorbitals as

$$\rho(\mathbf{r}) = \sum_{\sigma} \sum_{i} |\phi_i(\{\mathbf{r}, \sigma\})|^2. \tag{3.136}$$

For this theoretical model system, $T_s(\rho)$ is equal to the exact kinetic energy functional. However, in a physical system for which $\lambda = 1$, the noninteracting kinetic energy functional would not be equal to the exact kinetic energy functional $T(\rho) = \langle \Psi_1 | \hat{T} | \Psi_1 \rangle$, $\Psi_1 \to \rho$; it can be shown instead that it is a lower bound to the exact kinetic energy since

$$T_s(\rho) = \inf_{\Psi_0 \to \rho} \left\langle \Psi_0 \left| \hat{T} \right| \Psi_0 \right\rangle \leqslant \left\langle \Psi_1 \left| \hat{T} \right| \Psi_1 \right\rangle. \tag{3.137}$$

Given this inequality and assuming that the non–interacting kinetic energy functional constitutes a fair approximation to the exact kinetic energy functional, $T_s(\rho)$ may be considered an explicit approximation to a part of the universal density functional - that pertaining to the kinetic energy. This approach forms part of the Kohn–Sham decomposition of the universal density functional, discussed in the ensuing subsection.

3.5.2 The Kohn–Sham Equations

The concept of the non–interacting model system provides the basis for the formulation of DFT presented by Kohn & Sham in 1965, in which an *exact* description of a system may be obtained by solving the one–electron Schrödinger Equation[65]. The key to this theory is the assumption that for an arbitrary N–electron system with ground–density $\rho(\mathbf{r})$, a non–interacting N–electron can be constructed with with the same density $\rho(\mathbf{r})$.

Based on this premise, Kohn & Sham introduced the following decomposition of the universal density functional, where for the present discussion $\lambda = 1$ since it is the energy of the physical system that the model seeks to recover

$$\begin{aligned} F(\rho) &= \inf_{\Psi \to \rho} \left\langle \Psi \left| \hat{T} + \hat{W} \right| \Psi \right\rangle = T(\rho) + W(\rho) \\ &= T_s(\rho) + J(\rho) + (T(\rho) - T_s(\rho)) + (W(\rho) - J(\rho)) \\ &= T_s(\rho) + J(\rho) + E_{xc}(\rho). \end{aligned} \tag{3.138}$$

Here $T_s(\rho)$ is the non–interacting kinetic energy functional defined in (3.134), $J(\rho)$ represents the classical Coulomb part of the electron interaction,

$$J(\rho) = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, \mathrm{d}\mathbf{r}_1 \, \mathrm{d}\mathbf{r}_2, \tag{3.139}$$

and $E_{xc}(\rho)$ is the *exchange–correlation energy functional*. As indicated in (3.138), the exchange–correlation energy term can be expressed as the sum of two components

$$E_{xc}(\rho) = T_c(\rho) + \mathcal{W}_{xc}(\rho), \tag{3.140}$$

which in turn are defined as

$$\textit{kinetic energy correlation correction:} \qquad T_c(\rho) = T(\rho) - T_s(\rho) \geqslant 0, \tag{3.141}$$

$$\textit{two–electron exchange \& correlation:} \qquad \mathcal{W}_{xc}(\rho) = W(\rho) - J(\rho) \leqslant 0. \tag{3.142}$$

From this identity, it can be seen that the addition of an exchange–correlation term to $T_s(\rho) + J(\rho)$, which are known and can be evaluated *exactly*, gives the exact density functional $F(\rho)$. This reduces the problem of approximating the functional form of $F(\rho)$ to that of approximating $E_{xc}(\rho)$, which is only a relatively small component of $F(\rho)$.

The variational expression for the ground–state energy of an N–electron system in Kohn–Sham DFT can thus be written as

$$\mathcal{E}(v) = \min_{\rho} \{F(\rho) + (v|\rho)\} = \min_{\rho} \{T_s(\rho) + J(\rho) + E_{xc}(\rho) + (v|\rho)\}. \tag{3.143}$$

In the KS scheme, (3.143) is evaluated using one–electron orbitals as in the Hartree–Fock method, from which the density is evaluated. These are subject to the constraints that the spinorbitals are orthonormal and that the sum of their occupations is equal to the number of electrons N.

The formula for the ground–state energy (3.143) further implies that $\mathcal{E}(v)$ can be calculated by finding the stationary value of $F(\rho) + (v|\rho)$ with respect to the density; in other words, taking the functional derivative with respect to ρ and finding its zero–values. This can be written as the *Euler equation* for the Hohenberg–Kohn functional,

$$\frac{\delta}{\delta \rho(\mathbf{r})} \left\{ F(\rho) + \int v(\mathbf{r}) \rho(\mathbf{r}) \mathrm{d}\mathbf{r} - \mu \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} \right\} = 0, \tag{3.144}$$

where the term μ is a Lagrange multiplier constraining the density such that the number of electrons remains constant. This may be re–arranged as

$$\frac{\delta F(\rho)}{\delta \rho(\mathbf{r})} + v(\mathbf{r}) = \mu, \tag{3.145}$$

where it is assumed that $F(\rho)$ is differentiable. In reality however, $F(\rho)$ is known to be everywhere discontinuous and non–differentiable[78]; its convexity & subdifferentiablity are sufficient that its minimum can be characterised. This can be addressed by Moreau–Yosida regularisation, in which the energy functional is modified by addition of an infinitesimal regularisation parameter such that it is strictly concave and its convex conjugate $F(\rho)$ is continuous & differentiable[79].

For the non–interacting system described in subsection 3.5.1, $F(\rho)$ is simply the non–interacting kinetic energy functional. The Euler equation (3.145) for the non–interacting system at the stationary point can be written as

$$\frac{\delta T_\mathrm{s}(\rho)}{\delta \rho(\mathbf{r})} + v_\mathrm{s}(\mathbf{r}) = \mu, \tag{3.146}$$

where $v_\mathrm{s}(\mathbf{r})$ is the *Kohn–Sham potential* which ensures that the density $\rho(\mathbf{r})$ is equal to the exact ground–state density. For the system at physical interaction strength, the Euler equation (3.145) can be written with the definition of $F(\rho)$ substituted in from (3.138) to give

$$\frac{\delta T_\mathrm{s}(\rho)}{\delta \rho(\mathbf{r})} + \frac{\delta J(\rho)}{\delta \rho(\mathbf{r})} + \frac{\delta E_\mathrm{xc}(\rho)}{\delta \rho(\mathbf{r})} + v_\mathrm{ext}(\mathbf{r}) = \mu, \tag{3.147}$$

where $v_\mathrm{ext}(\mathbf{r})$ is the potential that the electrons experience by virtue of the charges of the atomic nuclei in the system. Defining the functional derivative of the Coulomb energy with respect to the density as the Coulomb potential, $\delta J(\rho)/\delta\rho(\mathbf{r}) = v_\mathrm{J}(\mathbf{r})$ and that of the exchange–correlation energy with respect to the density as the exchange–correlation potential $\delta E_\mathrm{xc}(\rho)/\delta\rho(\mathbf{r}) = v_\mathrm{xc}(\mathbf{r})$, (3.146) & (3.147) may be solved simultaneously to yield the definition of the Kohn–Sham potential,

$$v_\mathrm{s}(\mathbf{r}) = v_\mathrm{ext}(\mathbf{r}) + v_\mathrm{J}(\mathbf{r}) + v_\mathrm{xc}(\mathbf{r}). \tag{3.148}$$

The corollary to this is a set of one–electron equations in the molecular orbitals ϕ_i of a Slater determinant Φ, similar to the canonical Hartree–Fock equations (1.93), that can be solved self–consistently to give the energy (3.143); these are the *Kohn–Sham equations*

$$\left[-\frac{1}{2}\nabla^2 + v_\mathrm{s}(\mathbf{r})\right]\phi_i(\{\mathbf{r},\sigma\}) = \varepsilon_i\phi_i(\{\mathbf{r},\sigma\}) \qquad \varepsilon_1 \leqslant \varepsilon_2 \leqslant \ldots \tag{3.149}$$

Here, ε_i are the energies of the Kohn–Sham orbitals ϕ_i, which define the spatial density according to (3.136), and v_s is the *Kohn–Sham potential* defined in (3.148).

It is important to note here the similarities and differences between the Kohn–Sham equations (3.149) and the HF equations (1.93). Both are mean–field approximations that rely on the self–consistent evaluation of a set of one–electron equations for a system represented by a Slater–determinant wavefunction. However, in the Hartree–Fock method electron correlation is completely neglected, and the resulting solution for both the total energy and the energies of the molecular orbitals are approximate.

In Kohn–Sham theory, electron correlation is explicitly treated in the exchange–correlation term; an exact form of the exchange–correlation density functional would give the exact density as the solution to the Kohn–Sham equations. The use of orbitals is essential to the Kohn–Sham scheme as it allows the exact evaluation of the non–interacting kinetic energy (that is a reasonable approximation to the true kinetic energy), however they do not necessarily share the same physical meaning as the Hartree–Fock orbitals. The solution to the Kohn–Sham equations is the *density* that the orbitals yield, rather than the orbitals themselves (although the physical meaning of the Kohn–Sham orbitals is the subject of discussion, see for example Ref. 80).

In principle, the Kohn–Sham scheme is an exact one, however in practice it is an approximate one. This is because the analytical form of the exchange–correlation term is unknown, and in density functional calculations it is approximated by an *exchange–correlation functional* (*xc*–functional). The number of these functionals has grown since the introduction of the Kohn–Sham scheme such that it is now possible to select a functional (or combinations of functionals) that is best–suited to the system being modelled, the level of accuracy required and the intended application (i.e. for the calculation of particular properties).

Before discussing the presently available *xc*–functionals, it is worth noting the main feature of Kohn–Sham theory that led it to become the most widely used model in quantum chemical calculations. Hohenberg & Kohn had previously

shown that a many–electron system could be described exactly by its electron density. However, their universal functional has an unknown analytical form, making it difficult to develop approximations that can be systematically improved.

The Kohn–Sham scheme is built upon the Hohenberg–Kohn theorems, however by posing the problem in terms of an antisymmetric set of one–electron orbitals, Kohn & Sham could then resolve the Hohenberg–Kohn functional into a sum of exactly calculable terms – non–interacting kinetic energy and Coulomb repulsion – and one unknown term; the *xc*–energy. The terms that make the largest contributions to the total energy can now be evaluated analytically, leaving an unknown term much smaller in magnitude. Although this does compromise the efficiency of the pure–density scheme of Hohenberg & Kohn, as solutions must be obtained for a manifold of one–electron orbitals, the Kohn–Sham approach is a much more practical one. The level of approximation required is small by comparison to the total energy functional, providing greater control of the overall calculation and isolating errors to the *xc* term, for which many approximations have been developed over the past 50 years[81].

a) The Hohenberg–Kohn Scheme b) The Kohn–Sham Scheme

3.5.3 Exchange–Correlation Functionals

The exchange–correlation term in the Kohn–Sham scheme is the only part that cannot be analytically evaluated and the focus of method–development work in DFT is on improving the approximations for this. It is convenient to resolve this term further into separate exchange and correlation functionals as

$$E_{xc}(\rho) = E_x(\rho) + E_c(\rho) \tag{3.150}$$

such that the two components may be treated with individual approximations, rather than one combined term. It has been observed that by far the larger contribution to the *xc* energy is from the exchange term; an example of this can be seen in the exchange and correlation energies of some closed–shell atoms given in Table 3.3. Therefore it is

Atom	$\mathcal{E}$	E_x	E_c
He	−2.902704	−1.024084	−0.041249
Be	−14.665701	−2.673557	−0.094528
Ne	−128.918017	−12.075856	−0.376824
Ar	−527.427595	−30.168319	−0.620232

Table 3.3: The exchange and correlation energies of four atoms, calculated by taking the difference of CCSD(T) calculations and non–interacting WY calculations (described in Chapter 5) for each atom, in the *aug-cc-pCVQZ* basis for He, Be & Ne, and the *aug-cc-pCVTZ* basis for Ar.

essential in any DFT calculation that a robust approximation to E_x is applied, as it makes a substantial contribution to the total energy. Although the correlation energy is much smaller in magnitude, without an accurate approximation to this term, the result of a DFT calculation will be qualitative at best; it is important to note that in Hartree–Fock theory, exchange is described *exactly*, but the neglect of correlation is why predictions made by Hartree–Fock largely fall short of chemical accuracy.

This may also suggest that it is unnecessary to model exchange with a density functional, because the calculation of exact–exchange is well established, further reducing the unknown component of the total energy to correlation only. However, there is no equivalent form by which correlation can be computed; the only definitive description of correlation in the one–electron spinorbital model is that obtained by full configuration–interaction as described in Section 2.2.

In Hartree–Fock theory, the exact exchange energy may be written in terms of the one–electron spinorbitals as

$$E_x^{\mathrm{HF}} = -\frac{1}{2}\sum_{ij}\int\int \frac{\phi_i(\{\mathbf{r}_1,\sigma_i\})\phi_j(\{\mathbf{r}_2,\sigma_j\})\phi_j(\{\mathbf{r}_1,\sigma_j\})\phi_i(\{\mathbf{r}_2,\sigma_i\})}{|\mathbf{r}_1-\mathbf{r}_2|}\delta_{\sigma_i\sigma_j}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2. \tag{3.151}$$

This may be reduced into an expression for exact exchange energy per unit charge & unit volume,

$$w_{\mathrm{x}}^{\mathrm{HF}}(\mathbf{r}_1) = -\frac{1}{2\rho(\mathbf{r}_1)} \sum_{ij} \int \frac{\phi_i(\{\mathbf{r}_1, \sigma_i\})\phi_j(\{\mathbf{r}_2, \sigma_j\})\phi_j(\{\mathbf{r}_1, \sigma_j\})\phi_i(\{\mathbf{r}_2, \sigma_i\})}{|\mathbf{r}_1 - \mathbf{r}_2|} \delta_{\sigma_i \sigma_j} \mathrm{d}\mathbf{r}_2 \tag{3.152}$$

which to compute requires an integral over two coordinates rather than one; Hartree–Fock exchange is a *non–local* quantity rather than a local one. Therefore it is very difficult for simple density functional approximations to reliably model exact exchange, since they are they are typically based on local quantities such as the density and its gradient at a single point in space.

Some hybrid functionals employ this form of orbital–based exchange either fully of partially. However, it is difficult to construct correlation functionals that work well with non–local exchange of the form (3.152). The reasons for this are explained by Buijse *et. al.* in Refs. 82,83.

In the following paragraphs, some of the observations from which density functional approximations (DFAs) are constructed will be discussed, followed by a brief outline of some of the most commonly used types of exchange and correlation functionals. The work presented later in Chapter 6 draws on this discussion, as some of the newer ideas in correlation approximations are investigated.

3.5.3.1 Properties Of Exchange–Correlation Functionals

Although the exact form of the *xc*–functional is unknown, there are several constraints that must be satisfied by candidate functionals, and these have played an important part in guiding the development of *xc* approximations. The first of these is simply

$$E_{\mathrm{x}}(\rho) < 0 \qquad \& \qquad E_{\mathrm{c}}(\rho) \leqslant 0. \tag{3.153}$$

These follow from the observations that exchange is present in every electronic system, correlation is absent in one–electron systems, and that every exchange and correlation interaction lowers the overall energy of the system. The second constraint is closely related to this, which is that the exact exchange energy of a one–electron system cancels the Coulomb self–repulsion term,

$$E_{\mathrm{xc}}(\rho_{N=1}) = E_{\mathrm{x}}(\rho_{N=1}) = -J(\rho_{N=1}). \tag{3.154}$$

A further constraint is derived from the exchange energy of of a uniform electron gas, which is known analytically. This is that, in the limit of slowly–varying density, the exchange energy of an N–electron system must approach that of an equivalent N–electron uniform electron gas (UEG),

$$\lim_{\nabla\rho(\mathbf{r})\to 0} E_{\mathrm{x}}(\rho) = E_{\mathrm{x}}^{\mathrm{UEG}}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}}(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.155}$$

The fourth known constraint was derived by Lieb & Oxford[84] and it provides a bound on the exact exchange–correlation energy of a system, given by

$$E_{\mathrm{x}}(\rho) \geqslant E_{\mathrm{xc}}(\rho) \geqslant -1.68 \int \rho^{\frac{4}{3}}(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.156}$$

The final set of constraints to be given here are all derived from the *scaling relations* of Levy[85] which provide explicit relations for the density with respect to uniform scaling of the coordinates,

$$\rho_\lambda(\mathbf{r}) = \lambda^3 \rho(\lambda\mathbf{r}). \tag{3.157}$$

From this basic principle, the following properties of exchange and correlation functionals have been derived

$$E_{\mathrm{x}}(\rho_\lambda) = \lambda E_{\mathrm{x}}(\rho) \qquad \lim_{\lambda\to\infty} E_{\mathrm{c}}(\rho_\lambda) > -\infty \qquad \lim_{|\mathbf{r}|\to\infty} v_{\mathrm{xc}}(\mathbf{r}) = -\frac{1}{|\mathbf{r}|} \qquad \lim_{|\mathbf{r}|\to\infty} w_{\mathrm{xc}}(\mathbf{r}) = -\frac{1}{2|\mathbf{r}|}, \tag{3.158}$$

where $w_{\mathrm{xc}}(\mathbf{r})$ is the exchange–correlation energy per electron & unit volume, from which the total exchange–correlation energy is recovered as

$$E_{\mathrm{xc}}(\rho) = \int w_{\mathrm{xc}}(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.159}$$

3.5.3.2 The Local Density Approximation

In the original paper by Kohn & Sham, it was proposed that exchange–correlation effects could be modelled with reasonable accuracy for most systems with a functional that depends on only the density at the point in space at which it is being evaluated. This type of functional is a *local density approximation* (LDA). One of the earliest LDAs for exchange was *Slater's X_α method*[86], in which the exact exchange component of HF theory is approximated with Dirac's exchange functional[10] for the UEG, scaled by an adjustable parameter α. The energy as a functional of the density for the X_α method is evaluated as

$$\mathcal{E}(\rho) = T_\mathrm{s}(\rho) + J(\rho) + (v|\rho) - 2\alpha\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\int \rho^{\frac{4}{3}}(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.160}$$

This model for exchange was adopted by Kohn & Sham, who found that a value of $\alpha = \frac{2}{3}$ was required to avoid overestimation of exchange effects in their LDA. The principal merit of this approximation comes from the exact nature of the Dirac exchange functional for the UEG, which is often a fair approximation for the electronic structure of metallic systems.

There is no such analytical expression for the correlation energy of a UEG other than for the high and low–density limits; in these definitions, the density is usually expressed in terms of the *Wigner–Seitz radius* r_s, which is theoretically the radius of a charge density sphere containing the charge of one electron,

$$r_\mathrm{s}(\mathbf{r}) = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{\frac{1}{3}}. \tag{3.161}$$

With this definition, the high[87] and low[88] density limits of the correlation energy density are given by

$$\begin{aligned} \text{High density:} \quad & w_\mathrm{c}(r_s) = A\ln(r_s) + B + r_s\left(C\ln(r_s) + D\right), \\ \text{Low density:} \quad & w_\mathrm{c}(r_s) = \frac{1}{2}\left(\frac{g_0}{r_s} + \frac{g_1}{r_s^{3/2}} + \ldots\right), \end{aligned} \tag{3.162}$$

in which A, B, C, D and $g_0, g_1, \ldots$ are derived coefficients of each term in the two expansions respectively. Other than for these limits however, approximating real systems by either of these extremes does not lead to accurate calculation of the correlation energy. Instead, these are combined with data collected from numerical studies of small systems, for example the *Quantum Monte–Carlo* (QMC) simulations of Ceperley & Alder[89], and the data parametrised as an analytical functional. An example is the LDA functional of Perdew & Wang (PW91)[90],

$$w_\mathrm{c}(r_s) = -2A\left(1 + \alpha_1 r_s\right)\ln\left[1 + \frac{1}{2A\left(\beta_1 r_s^{1/2} + \beta_2 r_s + \beta_3 r_s^{3/2} + \beta_4 r_s^2\right)}\right], \tag{3.163}$$

where the constants A, α_1 and β_{1-4} are derived from the parametrisation of Ceperley & Alder's QMC data.

3.5.3.3 The Density Gradient Expansion

The LDAs were a reasonable initial approximation for exchange and correlation in DFT; there are some systems that are adequately modelled by LDA functionals only, usually metals, which have a highly uniform valence electron density in their ground state. However, for systems such as molecules, chemical accuracy cannot be reached with LDAs alone because the electron density has significant inhomogeneity, particularly in valence regions most important to their chemical properties. In an attempt to address this, *xc*–functionals were developed that depended not only on the density, but on its gradient at a given point in space. This set of functionals became known as the *generalised gradient approximations* (GGAs), and unlike the LDAs, there are a large number of different GGAs that have been developed for use in DFT calculations.

The general form taken by the GGA functionals is that of an LDA functional, modified by a correction term that depends on the density gradient,

$$E_\mathrm{xc}^\mathrm{GGA}(\rho) = \int w_\mathrm{xc}^\mathrm{LDA}(\mathbf{r})F_\mathrm{xc}(\rho, s)\rho(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{3.164}$$

where $F_\mathrm{xc}(\rho, s)$ is an *enhancement factor* and $s(\mathbf{r})$ is the *reduced density gradient* - a dimensionless form of the density gradient, given by

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}. \tag{3.165}$$

It is in the form of the enhancement factor that individual GGAs differ; the precise form of an enhancement factor is dependent on the empirical data and exact conditions that guided its construction, and to a certain extent the particular applications the functional was intended for. An example of one of the most widely used exchange functionals is the Becke–88 GGA[91], in which the enhancement factor has the functional form

$$F_{\mathrm{x}}(\rho, s) = 1 + 4\beta \frac{|s|^2}{1 + 6\beta s \sinh^{-1}(s)}, \tag{3.166}$$

in which β is a constant with value 0.0042, determined by fitting to data for Hartree–Fock exchange over a series of noble–gas atoms. An example of a well–established correlation GGA functional would be the PBE correlation functional of Perdew, Burke & Ernzerhof[92]. The form of the PBE functional has a large number of terms and for brevity will be omitted here, however the underlying structure is that of a gradient correction to the LDA correlation functional – in this case the PW91 functional.

A further group of functionals can be identified that extend the GGA description by the inclusion of additional terms, with the aim of more accurately describing the inhomogeneity of an electronic system. Such additional terms occasionally contain the *Laplacian* of the density $\nabla^2\rho(\mathbf{r})$, which is its second–order spatial derivative, however most often the additional terms depend on the (positive definite) *kinetic energy density*,

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_i \sum_\sigma \left|\nabla\phi_i\left(\{\mathbf{r}, \sigma\}\right)\right|^2. \tag{3.167}$$

Density functional approximations that contain τ and/or $\nabla^2\rho$ constitute a class of functionals known as the *meta generalised–gradient approximations*, or *meta-GGAs*. Two notable examples of meta-GGAs are the Becke–Roussel exchange functional[93] and the LYP correlation functional[94]. Collectively, the GGAs and meta-GGAs and have been highly successful because they predict exchange and correlation energies more accurately than LDAs, but without being significantly more expensive to compute. It could be argued that it was only with the development of robust GGAs that DFT became a tool that could be used to make reliable predictions for chemistry, with accuracies often rivalling those of post-Hartree–Fock methods but at only a fraction of the cost.

3.5.3.4 The Hybrid Density Functionals

One of the most important developments contributing to the success of the GGAs was the B88 exchange functional (3.166), which was unique amongst exchange functionals at the time in that it produced exchange energy densities that exhibited the correct asymptotic decay. As briefly mentioned earlier in subsection 3.5.3, it had been found that a greater overall accuracy was achieved when correlation density functionals were combined with exchange density functionals, rather than with exact non–local exchange. However, in 1993 Becke showed that an accuracy exceeding that of the GGAs and meta-GGAs may be obtained by downscaling the contribution to the total *xc*–energy from the exchange functional and making–up the difference with exact exchange[95]. Such combinations of *xc*–functional and exact exchange became known as the *hybrid functionals*, and have the general form

$$E_{\mathrm{xc}} = \int \left[a w_{\mathrm{x}}^{\mathrm{HF}}(\mathbf{r}) + (1-a) w_{\mathrm{x}}^{\mathrm{DFT}}(\mathbf{r}) + w_{\mathrm{c}}(\mathbf{r})\right] \rho(\mathbf{r}) \mathrm{d}\mathbf{r}. \tag{3.168}$$

Here, a is the parameter that determines the amount of exact exchange to be included in the hybrid functional and is often determined with the aid of experimental data. These hybrids built on the success of the GGAs, representing a particular improvement in the accuracy of thermochemical predictions; perhaps the most widely used DFT functional of them all, B3LYP[91,94,95], is a hybrid functional with form

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = E_{\mathrm{x}}^{\mathrm{LDA}} + a_0\left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LDA}}\right) + a_{\mathrm{x}}\left(E_{\mathrm{x}}^{B88} - E_{\mathrm{x}}^{\mathrm{LDA}}\right) + E_{\mathrm{c}}^{\mathrm{VWN}} + a_{\mathrm{c}}\left(E_{\mathrm{c}}^{\mathrm{LYP}} - E_{\mathrm{c}}^{\mathrm{VWN}}\right), \tag{3.169}$$

where $E_{\mathrm{c}}^{\mathrm{VWN}}$ is the LDA correlation functional of Vosko, Wilk & Nusair[96] and the three mixing parameters are $a_0 = 0.20$, $a_{\mathrm{x}} = 0.72$ and $a_{\mathrm{c}} = 0.81$.

With the advancement of Kohn–Sham density functional approximations, DFT has become by far the most commonly used model in quantum chemical simulations, combining an accuracy that can rival that of the post-Hartree–Fock correlation methods discussed in Chapter 2, but with impressive scalability thus greatly extending the applicability of quantum chemical tools in general. However, in the absence of an exact correlation functional, there remains a large number of applications for which even the most advanced functionals fall short of chemical accuracy, highlighting the continuing need for improvement.

There also remains the need for greater understanding of correlation in general, so that a more systematic approach may be taken to the its treatment within Kohn–Sham DFT, which is the main focus of the present work. A theoretical

approach which allows exchange & correlation in DFT to be examined in much closer detail, and which underpins much of the work presented in later chapters, is the adiabatic connection method – which is discussed in the following section.

3.6 The Adiabatic Connection

The main area of approximation in Kohn–Sham DFT is that of the exchange–correlation functional, which approximates the difference between the non–interacting Kohn–Sham system and the equivalent system at physical interaction strength. It follows that examining how the behaviour of the non–interacting system evolves when electron–electron interaction is gradually re–introduced could provide further insight into the relationship between a physical system and its Kohn–Sham non–interacting equivalent.

This was first proposed in 1975 by Langreth & Perdew[97–100], who suggested that new insight into the nature of electron interaction may be gained by modelling an electronic system as the interaction between electrons is smoothly increased from the non–interacting state to that of the physically interacting system. Over this range, the system evolves through a family of eigenstates that form what is referred to as the *adiabatic connection* (AC). Most frequently, a linear path between these two limits is considered[100], where the Coulomb operator is simply scaled linearly by the the value of a coupling–constant coefficient λ. However, generalised ACs[101] have been explored along non–linear paths[102,103]; in the present work, only the linear path is considered but the generalisation to non–linear paths is straightforward.

The Hamiltonian at coupling–constant strength λ is given by (3.1); for the linear AC the λ–interacting Hamiltonian may be written more explicitly as

$$\hat{\mathcal{H}}_\lambda = \hat{T} + \lambda\hat{W} + \hat{V}^\lambda \qquad \lambda \in \{0,1\}, \tag{3.170}$$

where $\hat{T}$ is the kinetic energy operator, $\hat{W}$ is the two–electron interaction operator and $\hat{V}^\lambda$ is the potential that binds the density at a given λ such that the density ρ_λ is constrained to equal that of the physical system ρ_1 for all $\lambda \in \{0,1\}$.

For a linear AC, the λ–interacting universal density functional can be defined using the constrained–search formalism as

$$F_\lambda(\rho) = \min_{\Psi_\lambda \to \rho} \left\langle \Psi_\lambda \left| \hat{T} + \lambda\hat{W} \right| \Psi_\lambda \right\rangle, \tag{3.171}$$

where the minimisation is over all admissible wavefunctions yielding ρ. The λ–interacting functional can be related to its non–interacting counterpart via

$$F_\lambda(\rho) = F_0(\rho) + \int_0^\lambda \frac{\partial F_\nu(\rho)}{\partial \nu}\,\mathrm{d}\nu, \tag{3.172}$$

where the derivative term can be simplified by application of the *Hellmann-Feynman Theorem*[104,105] to (3.171), yielding[x]

$$\frac{\partial F_\lambda(\rho)}{\partial \lambda} = \left\langle \Psi_\lambda \left| \frac{\partial\left(\hat{T} + \lambda\hat{W}\right)}{\partial\lambda} \right| \Psi_\lambda \right\rangle = \left\langle \Psi_\lambda \left| \hat{W} \right| \Psi_\lambda \right\rangle = W_\lambda(\rho). \tag{3.173}$$

Furthermore, as discussed in subsection 3.5.1, the constrained–search functional (3.171) at $\lambda = 0$ simply reduces to the non–interacting kinetic energy

$$F_0(\rho) = \min_{\Psi_0 \to \rho} \left\langle \Psi_0 \left| \hat{T} \right| \Psi_0 \right\rangle = T_\mathrm{s}(\rho), \tag{3.174}$$

thus the λ–interacting functional (3.172) may be re–written as

$$F_\lambda(\rho) = T_\mathrm{s}(\rho) + \int_0^\lambda W_\nu(\rho)\,\mathrm{d}\nu. \tag{3.175}$$

By comparison with the Kohn–Sham decomposition (3.138) and the definition of the classical Coulomb term (3.139), noting that the density is fixed for all λ, it can be deduced that

$$W_0(\rho) = J(\rho) + E_\mathrm{x}(\rho), \tag{3.176}$$

[x] The Hellmann-Feynman theorem states that the non–degenerate eigenvalues of a Hermitian operator in a parameter–dependent eigensystem, i.e. $\hat{\mathcal{H}}_\lambda\Psi_\lambda = \mathcal{E}_\lambda\Psi_\lambda$, vary with respect to the parameter as

$$\frac{\partial \mathcal{E}_\lambda}{\partial\lambda} = \left\langle \Psi_\lambda \left| \frac{\partial \hat{\mathcal{H}}_\lambda}{\partial\lambda} \right| \Psi_\lambda \right\rangle$$

since no correlation energy is present at $\lambda = 0$ hence $W_0(\rho) - J(\rho)$ yields only the exchange energy. Defining the *adiabatic connection integrand* as

$$\mathcal{W}_\lambda(\rho) = W_\lambda(\rho) - J(\rho), \tag{3.177}$$

this may be substituted back into (3.175) to give

$$\begin{aligned} F_\lambda(\rho) &= T_\mathrm{s}(\rho) + \int_0^\lambda J(\rho)\,\mathrm{d}\nu + \int_0^\lambda \mathcal{W}_\nu(\rho)\,\mathrm{d}\nu \\ &= T_\mathrm{s}(\rho) + \lambda J(\rho) + \lambda E_\mathrm{x}(\rho) + \int_0^\lambda \mathcal{W}_{\mathrm{c},\nu}(\rho)\,\mathrm{d}\nu \\ &= T_\mathrm{s}(\rho) + \lambda J(\rho) + \lambda E_\mathrm{x}(\rho) + E_{\mathrm{c},\lambda}(\rho), \end{aligned} \tag{3.178}$$

where $\mathcal{W}_{\mathrm{c},\lambda}$ is simply the correlation component of the AC integrand,

$$\mathcal{W}_{\mathrm{c},\lambda}(\rho) = \mathcal{W}_\lambda(\rho) - \mathcal{W}_0(\rho). \tag{3.179}$$

This therefore provides the AC expression for the λ–interacting correlation energy in DFT,[xi]

$$E_{\mathrm{c},\lambda}(\rho) = \int_0^\lambda \mathcal{W}_{\mathrm{c},\nu}(\rho)\,\mathrm{d}\nu. \tag{3.180}$$

As a brief adjunct, it is noted here that the AC approach can be applied to ensemble states of the form discussed in subsection 3.4.3. For this, the λ–dependent universal density functional is written in the constrained–search [66,67,107] form for canonical ensembles,

$$F_\lambda(\rho) = \min_{\gamma\to\rho} \mathrm{tr}\,\gamma\hat{\mathcal{H}}_\lambda(0) \tag{3.181}$$

where the minimisation is over all density matrices γ associated with the interacting electron density ρ. This functional is convex in ρ, concave in λ and non-negative for $\lambda \geqslant 0$. This results in an expression for the correlation energy equivalent to (3.180) given by

$$E_{\mathrm{c},\lambda}(\rho) = \int_0^\lambda \{\mathrm{tr}\,\gamma_\nu^\rho \hat{W} - \mathrm{tr}\,\gamma_0^\rho \hat{W}\}\,\mathrm{d}\nu, \tag{3.182}$$

where γ_λ^ρ is the minimising ensemble state at interaction strength λ.

The adiabatic connection model provides the central framework on which much of the work presented in the forthcoming chapters is based. Quantitatively, it provides an identity for the exact Kohn–Sham correlation energy whilst qualitatively the features of the adiabatic connection in many cases provide an intuitive means of characterising patterns of correlation energy in systems and provides a guide as to how it may be most appropriately modelled. This theme will be developed throughout the forthcoming chapters, in which computation of the adiabatic connection and subsequently modelling of it are considered.

[xi] For a more in–depth review of the adiabatic connection formalism, see Ref. 106

4 Variation Principles In Wavefunction & Density Functional Theories

4.1 Introduction

Variation principles lie at the heart of many quantum-chemical theories, giving practical prescriptions for how to obtain the best electronic energy, wavefunction or electron density via optimisation. They may also offer insight into the connections between traditional *ab initio* wavefunction based approaches and density-functional theory (DFT). In this work, we examine a new variation principle, proposed by Gidopoulos in Ref. 108 for the determination of the non-interacting system of relevance to Kohn–Sham theory.

The variation principle proposed by Gidopoulos consists of minimising the left-hand side of the inequality

$$\langle \Psi | \hat{\mathcal{H}}_0(v) | \Psi \rangle - \mathcal{E}_0(v) \geqslant 0, \tag{4.1}$$

with respect to variations of the potential v, for a fixed electronic wavefunction Ψ corresponding to a system of interest – typically the physical ground-state wavefunction for the system. The energy $\mathcal{E}_0(v)$ in (4.1) is the ground-state energy of a non-interacting system, associated with the non-interacting Hamiltonian $\hat{\mathcal{H}}_0(v) = \hat{T} + \sum_i v(\mathbf{r}_i)$, where $\hat{T}$ is the kinetic energy operator.

As discussed in Ref. 108, the minimisation of the left-hand side of (4.1) yields the Kohn–Sham non-interacting potential v_s associated with a non-interacting system that has the same density as that of the chosen input wavefunction Ψ. The same variation principle was also described earlier by Davidson in a different context [109] and used as a tool to understand the analytic properties of the first–order reduced density matrix associated with Ψ. This complements its use in Ref. 108, where it provides a tool for the optimisation of the potential v. Here we refer to (4.1) as the *Gidopoulos–Davidson variation principle*.

At first glance, the Gidopoulos–Davidson variation principle appears to be markedly different from alternative approaches for determining the Kohn–Sham system corresponding to a reference wavefunction or density. For example, in Levy's constrained-search approach to DFT, [66,107] a constraint on the electron density is explicitly applied to determine the Kohn–Sham system. More closely related is the Lieb variation principle, which for a non-interacting system corresponds to maximising the left-hand side of the inequality [67]

$$\mathcal{E}_0(v) - (v|\rho) \leqslant T_s(\rho) \tag{4.2}$$

with respect to variations of the potential v for a given input electron density ρ, using the notation $(v|\rho) = \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}$. Both the Gidopoulos–Davidson and Lieb variation principles involve an unconstrained optimisation over v, yielding the Kohn–Sham potential v_s as their optimiser. Furthermore, their functional derivatives are identical up to a sign [67,108].

These observations motivate us to explore the connection between the Lieb and Gidopoulos–Davidson variation principles in (4.1) & (4.2), respectively. We begin by reviewing standard variation principles in Section 4.2. In Section 4.3, we highlight the connections between the Gidopoulos–Davidson and Lieb variation principles, including extensions to general interaction strengths and to mixed states.

Having established the close connection between these alternative variation principles, we present some results from numerical implementation in a common framework in Section 4.4, highlighting the equivalent information they yield both in the non-interacting limit and for arbitrary interaction strengths. In Section 4.5, we make some concluding remarks and discuss possible directions for future work.

4.2 Hohenberg–Kohn & Lieb Variation Principles

At a given interaction strength λ, the ground–state energy of an N–electron eigenfunction Ψ of the Hamiltonian $\hat{\mathcal{H}}_\lambda(v)$ can be defined in the context of wavefunction theory by varying the wavefunction Ψ according to the *Rayleigh–Ritz variation principle*,

$$\mathcal{E}_\lambda(v) = \inf_{\Psi \in \mathcal{W}_N} \left\langle \Psi \left| \hat{\mathcal{H}}_\lambda(v) \right| \Psi \right\rangle, \tag{4.3}$$

where $\mathcal{W}_N$ is the set of all L^2-normalised, antisymmetric N-electron wavefunctions with a finite kinetic energy,

$$\mathcal{W}_N = \left\{ \Psi \;\middle|\; \langle\Psi|\Psi\rangle = 1 \;;\; \textstyle\sum_i \langle\nabla_i\Psi|\nabla_i\Psi\rangle < \infty \right\}. \tag{4.4}$$

The Rayleigh–Ritz variation principle is well defined for all potentials v belonging to the vector space $\mathcal{X}^* = L^{3/2} + L^\infty$, which includes all Coulomb potentials.[67]

Being concave and continuous in v, the ground–state energy of (4.3) may be expressed in terms of the *Hohenberg–Kohn variation principle*

$$\mathcal{E}_\lambda(v) = \inf_{\rho\in\mathcal{X}} \{F_\lambda(\rho) + (v|\rho)\}, \tag{4.5}$$

where Lieb's universal density functional F_λ is obtained from the ground–state energy by the *Lieb variation principle*[67]

$$F_\lambda(\rho) = \sup_{v\in\mathcal{X}^*} \{\mathcal{E}_\lambda(v) - (v|\rho)\}. \tag{4.6}$$

The functionals $\mathcal{E}_\lambda$ and F_λ are a conjugate pair, related by mutual Legendre–Fenchel transforms. The vector spaces of admissible densities and potentials are the Banach spaces $\mathcal{X} = L^3 \cap L^1$ and $\mathcal{X}^* = L^{3/2} + L^\infty$ respectively, encompassing all N–representable densities $\rho \in \mathcal{X}$ and all Coulomb potentials $v \in \mathcal{X}^*$ with which the density has a finite interaction energy.

4.3 Gidopoulos–Davidson Variation Principle

The variation principle of Gidopoulos in Ref. 108 allows for the determination of the non-interacting system of relevance to Kohn–Sham theory and may be written in the form

$$D_0(\Psi) = \inf_{v\in\mathcal{X}^*} \left\{ \langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle - \mathcal{E}_0(v) \right\}, \tag{4.7}$$

where $\Psi \in \mathcal{W}_N$ is an electronic wavefunction corresponding to the physical system of interest; typically the physical ground-state wave function of $\hat{\mathcal{H}}_1(v)$ for some $v \in \mathcal{X}^*$. The energy $\mathcal{E}_0(v)$ is the ground-state energy of the non–interacting system, defined according to (4.3). Note that $D_0(\Psi)$ is well defined since $\langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle - \mathcal{E}_0(v) \geqslant 0$ for each $\Psi \in \mathcal{W}_N$ by the Rayleigh–Ritz variation principle.

4.3.1 Relationship To Lieb Variation Principle

The Gidopoulos–Davidson variation principle is related in a simple manner to the non-interacting Lieb variation principle

$$F_0(\rho) = \sup_{v\in\mathcal{X}^*} \{\mathcal{E}_0(v) - (v|\rho)\}. \tag{4.8}$$

To see the relation, we decompose the non-interacting expectation value $\langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle$ in the manner

$$\langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle = T(\Psi) + (v|\rho_\Psi), \tag{4.9}$$

where $T(\Psi) = \langle\Psi|\hat{T}|\Psi\rangle$ and ρ_Ψ are the kinetic energy and density yielded by Ψ, respectively. A comparison of the functionals in (4.7) & (4.8) then gives,

$$\begin{aligned} D_0(\Psi) &= T(\Psi) - F_0(\rho_\Psi) \\ &= T(\Psi) - \bar{F}_0(\rho_\Psi), \end{aligned} \tag{4.10}$$

where, in the final step, the Lieb functional is replaced with the Levy constrained search functional, noting that ρ_Ψ is pure–state representable. We conclude that *the Gidopoulos–Davidson functional of a given system is simply the total kinetic energy of this system minus the non-interacting Lieb functional*.

Since the non-interacting Lieb functional is the non-interacting Kohn–Sham kinetic energy,

$$F_0(\rho) = T_s(\rho) \tag{4.11}$$

we find that the Gidopoulos–Davidson functional is the Kohn–Sham kinetic-energy correlation energy,

$$D_0(\Psi) = T(\Psi) - T_s(\rho_\Psi). \tag{4.12}$$

Introducing the constrained-search formalism,[66,67,107] we obtain

$$D_0(\Psi) = \langle\Psi|\hat{T}|\Psi\rangle - \inf_{\Phi\to\rho_\Psi} \langle\Phi|\hat{T}|\Phi\rangle \tag{4.13}$$

where Φ is a single Slater determinant describing the non-interacting Kohn–Sham system.

The relationship of the Gidopoulos–Davidson functional to the correlation kinetic energy is well known[108]. Here we see that the non-interacting Gidopoulos–Davidson and Lieb variation principles yield the same Kohn–Sham system from different directions. The Lieb variation principle minimises the value of the non-interacting kinetic energy T_s, subject to a density constraint, whilst the Gidopoulos-Davidson variation principle maximises the correlation kinetic energy $T_c = T - T_s$ subject to a similar density constraint. Following the discussion in Section 4.2, we observe that the potential in the Gidopolous–Davidson variation principle of (4.7) may be viewed as the Lagrange multiplier for the density constraint in (4.13).

4.3.2 Objective Functions

Being related in such a simple manner, the optimisations of the Gidopoulos–Davidson and Lieb functional are also related in a simple way. Expressing the functionals in terms of their objective functions, we find

$$D_0(\Psi) = \inf_{v\in\mathcal{X}^*} G_0(v,\Psi) \tag{4.14}$$

$$F_0(\rho) = \sup_{v\in\mathcal{X}^*} L_0(v,\rho), \tag{4.15}$$

where

$$G_0(v,\Psi) = \langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle - \mathcal{E}_0(v) \tag{4.16}$$

$$L_0(v,\rho) = \mathcal{E}_0(v) - (v|\rho). \tag{4.17}$$

Hence we obtain, in agreement with (4.10),

$$G_0(v,\Psi) = T(\Psi) - L_0(v,\rho_\Psi). \tag{4.18}$$

The functional $L_0(v,\rho)$ is concave in v and affine in ρ, whereas $G_0(v,\Psi)$ is convex in v. After a generalisation to mixed states, G_0 becomes convex also in the second variable; see subsection 4.3.4.

4.3.3 Functional Derivatives Of Objective Functions

To determine the stationary points of the Gidopoulos–Davidson and Lieb variation principles, we note that ground-state energy $\mathcal{E}_0(v)$ is differentiable with functional derivative ρ_v if v supports a ground state with a unique density ρ_v. For a given $\Psi\in\mathcal{W}_N$, the expectation value $\langle\Psi|\hat{\mathcal{H}}_0(v)|\Psi\rangle$ is always differentiable with respect to v, with functional derivative ρ_Ψ. Hence, assuming differentiability of $\mathcal{E}_0$ at v, we have

$$\frac{\delta G_0(v,\Psi)}{\delta v(\mathbf{r})} = \rho_\Psi(\mathbf{r}) - \rho_v(\mathbf{r}), \tag{4.19}$$

$$\frac{\delta L_0(v,\rho)}{\delta v(\mathbf{r})} = \rho_v(\mathbf{r}) - \rho(\mathbf{r}). \tag{4.20}$$

When $\rho = \rho_\Psi$, the functional derivatives are identical except for the sign difference.

The second derivatives of G_0 and L_0 with respect to the potential v may also be readily evaluated. They are equal to (minus and plus) one half the non-interacting static density response function of the system,

$$\begin{aligned}\frac{\delta^2 G_0(v,\Psi)}{\delta v(\mathbf{r})\delta v(\mathbf{r}')} &= -\frac{\delta^2 L_0(v,\rho)}{\delta v(\mathbf{r})\delta v(\mathbf{r}')} = -\frac{1}{2}\chi_0(\mathbf{r},\mathbf{r}') \\ &= -\frac{1}{2}\sum_{ia}\frac{\varphi_i(\mathbf{r})\varphi_i^*(\mathbf{r}')\varphi_a(\mathbf{r}')\varphi_a^*(\mathbf{r})}{\varepsilon_i-\varepsilon_a} + \text{c.c.},\end{aligned} \tag{4.21}$$

where the indices i and a denote occupied and virtual orbitals respectively, whose orbital energies are ε_i and ε_a.

In Ref. 108 focus is placed on the optimisation of G_0 with respect to v. In passing, we note that the non-interacting Hamiltonian readily separates into one-electron contributions $\hat{\mathcal{H}}_0(v) = \sum_i \hat{h}_i(v)$ with $\hat{h}_i(v) = -\frac{1}{2}\nabla_i^2 + v(\mathbf{r}_i)$ and that

the orbitals entering (4.21) are the eigenfunctions of this one-electron Hamiltonian. The non-interacting ground–state energy is the sum of the occupied orbital energies, $\mathcal{E}_0(v) = \sum_i \varepsilon_i$. We also remark that, although $\pm\frac{1}{2}\chi_0(\mathbf{r},\mathbf{r}')$ is positive/negative semi-definite, this does not imply that G_0/L_0 are convex/concave in v since the derivatives in (4.21) are not defined for all potentials.

Throughout this discussion we have assumed the differentiability of $L_0(v,\rho)$ and $G_0(v,\Psi)$. The functional $L_0(v,\rho)$ is not straightforwardly differentiable as discussed by Lammert[78], however this issue can be avoided by using a regularised form as discussed in Ref. 79. Since the derivative of $G_0(v,\Psi)$ amounts to taking the derivative of $-L_0(v,\rho_\Psi)$ (see (4.18)), the same regularisation techniques can be applied to this functional.

In the Gidopoulos–Davidson and Lieb variation principles, the potential plays the role of a Lagrange multiplier. To see this, we rewrite the Lieb variation principle as a minimax problem

$$\begin{aligned} F_\lambda(\rho) &= \sup_{v\in\mathcal{X}^*}\inf_{\Psi\in\mathcal{W}_N}\left\{\langle\Psi|\hat{\mathcal{H}}_\lambda(v)|\Psi\rangle - (v|\rho)\right\} \\ &= \sup_{v\in\mathcal{X}^*}\inf_{\Psi\in\mathcal{W}_N}\left\{\langle\Psi|\hat{\mathcal{H}}_\lambda(0)|\Psi\rangle - (v|\rho-\rho_\Psi)\right\}. \end{aligned} \tag{4.22}$$

Since the search over v terminates when the stationary condition $\rho(\mathbf{r}) - \rho_\Psi(\mathbf{r}) = 0$ is satisfied, the Lieb variation principle represents a minimisation of the expectation value $\langle\Psi|\hat{\mathcal{H}}_\lambda(0)|\Psi\rangle$ subject to the constraint $\Psi\to\rho$. The potential plays precisely the same role in the Gidopoulos–Davidson variation principle, which becomes clear in light of the relationship between the two functionals given by (4.10).

4.3.4 Generalisation To Ensembles

It is well known that the Lieb functional $F_0(\rho)$ is convex in ρ[67] (see Section 3.4). It follows that $-F_0(\rho)$ is concave in ρ but not that the Gidopoulos–Davidson functional $D_0(\Psi) = T(\Psi) - F_0(\rho_\Psi)$ is concave in Ψ since $T(\Psi)$ is not.

Generalising the Gidopoulos–Davidson functional for pure states $\Psi\in\mathcal{W}_N$ to canonical ensembles $\gamma\in\mathcal{K}_N$, we obtain the concave functional

$$\begin{aligned} \mathcal{D}_0(\gamma) &= \inf_{v\in\mathcal{X}^*}\left\{\operatorname{tr}\gamma\hat{\mathcal{H}}_0(v) - \mathcal{E}_0(v)\right\} \\ &= \mathcal{T}(\gamma) - \sup_{v\in\mathcal{X}^*}\left\{\mathcal{E}_0(v) - (v|\rho_\Psi)\right\}, \end{aligned} \tag{4.23}$$

where $\mathcal{T}(\gamma) = \operatorname{tr}\gamma\hat{T}$. To show concavity, we select $\gamma_1,\gamma_2\in\mathcal{K}_N$ and obtain for each $0<\alpha<1$ the inequality

$$\begin{aligned} \mathcal{D}_0(\alpha\gamma_1 + (1-\alpha)\gamma_2) &= \alpha\operatorname{tr}\gamma_1\hat{T} + (1-\alpha)\operatorname{tr}\gamma_2\hat{T} - F_0(\alpha\rho_1 + (1-\alpha)\rho_2) \\ &\leqslant \alpha\operatorname{tr}\gamma_1\hat{T}(1-\alpha)\operatorname{tr}\gamma_2\hat{T} - \alpha F_0(\rho_1) - (1-\alpha)F_0(\rho_2) \\ &= \alpha\mathcal{D}_0(\gamma_1) + (1-\alpha)\mathcal{D}_0(\gamma_2), \end{aligned} \tag{4.24}$$

where in the second step we have used the convexity of the Lieb functional.

Since Ψ occurs quadratically in $D_0(\Psi)$, a similar proof is precluded for the pure state Gidopoulos–Davidson functional, which is indeed not concave. Note that, for pure states $\gamma_\Psi = |\Psi\rangle\langle\Psi|$, the ensemble Gidopoulos–Davidson functional reduces to the original functional: $\mathcal{D}_0(\gamma_\Psi) = D_0(\Psi)$.

4.3.5 Generalisation To Arbitrary Interaction Strengths

The Gidopoulos–Davidson functional may be extended to interacting systems in the manner

$$D_\lambda(\Psi) = \inf_{v\in\mathcal{X}^*}\left\{\langle\Psi|\hat{\mathcal{H}}_\lambda(v)|\Psi\rangle - \mathcal{E}_\lambda(v)\right\}, \tag{4.25}$$

which is related to the Lieb functional via

$$D_\lambda(\Psi) = \langle\Psi|\hat{T} + \hat{W}_\lambda|\Psi\rangle - F_\lambda(\rho_\Psi) \tag{4.26}$$

and can re-expressed in the constrained-search form as

$$D_\lambda(\Psi) = \langle\Psi|\hat{T} + \hat{W}_\lambda|\Psi\rangle - \inf_{\Phi\to\rho_\Psi}\langle\Phi|\hat{T} + \hat{W}_\lambda|\Phi\rangle. \tag{4.27}$$

The first derivative of the objective functional, $G_\lambda(v,\Psi) = \langle\Psi|\hat{\mathcal{H}}_\lambda(v)|\Psi\rangle - \mathcal{E}_\lambda(v)$, is again a simple density difference,

$$\frac{\delta G_\lambda(v,\Psi)}{\delta v(\mathbf{r})} = \rho_\Psi(\mathbf{r}) - \rho_v(\mathbf{r}), \tag{4.28}$$

and its second derivative can be expressed in terms of the λ–interacting density response function

$$\frac{\delta^2 G_\lambda(v,\Psi)}{\delta v(\mathbf{r})\delta v(\mathbf{r}')} = -\frac{1}{2}\chi_\lambda(\mathbf{r},\mathbf{r}'). \tag{4.29}$$

To perform practical optimisations using (4.25), we therefore require knowledge not only of the kinetic energy associated with the input wavefunction Ψ but also the λ-interacting two-electron interaction energy, $W_\lambda(\Psi) = \langle\Psi|\hat{W}_\lambda|\Psi\rangle$. In practice, these quantities can be computed from the one- and two-particle reduced density matrices respectively.

To make practical use of these expressions, approaches for the calculation of the λ–interacting wave functions yielding a chosen electron density are required, see e.g. Refs. 110–112. The constraint that the density is fixed for all λ may be easily enforced by supplying fixed arguments ρ and Ψ to (4.6) & (4.25) for all λ. We now discuss our implementation of the (generalised) Gidopoulos–Davidson variation principle, exploring the close connections to the generalised Lieb functional numerically.

4.4 Results

From the discussion in the Section 4.3, it is evident that the Gidopoulos–Davidson and Lieb optimisations are sufficiently closely related that they may be implemented in a common computational framework. We first discuss some details of our implementation; we then demonstrate the equivalence of the two approaches by performing numerical optimisations for a set of small atomic and molecular systems.

4.4.1 Computational Details

The variation principle given in (4.25) allows a value to be obtained for the generalised Gidopoulos–Davidson functional by evaluating its infimum with respect to v. If the density yielded by the reference wave function ρ_Ψ is v–representable, the infimum becomes a minimum. To vary v such that an optimisation over the potential may be carried out in a practical computational scheme, the potential is modelled using the basis-set expansion of Wu and Yang [113,114]

$$v_\lambda(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + (1-\lambda)v_{\text{ref}}(\mathbf{r}) + \sum_t b_t g_t(\mathbf{r}). \tag{4.30}$$

Here $v_{\text{ext}}(\mathbf{r})$ is the external potential due to interaction of the electrons with the atomic nuclei, $v_{\text{ref}}(\mathbf{r})$ is a fixed reference potential chosen to ensure that $v_\lambda(\mathbf{r})$ has the correct asymptotic behaviour and $\{g_t\}$ are a set of Gaussian basis functions with coefficients $\{b_t\}$. The reference potential employed in the present work is the Fermi–Amaldi potential [115]. With this choice of potential expansion, the derivatives corresponding to (4.28) & (4.29) may be readily implemented as described in Refs. 111,112,114, allowing the objective functional to be effectively optimised with respect to the set of coefficients $\{b_t\}$. This optimisation procedure is discussed in greater detail in Section 5.2.

An un-contracted form of the Gaussian basis set aug-cc-pVTZ [116,117] in the spherical–harmonic basis is used for both the orbital expansion and for the potential expansion in (4.30) for all systems. An approximate Newton method is employed to accelerate convergence of the optimisation [118], in which the Hessian is regularised using a truncated singular value decomposition with a threshold of 10^{-6} a.u. In all calculations, the convergence threshold was set to 10^{-6} a.u. on the L^2 norm of the objective functional gradient.

To obtain a reasonably accurate approximation to the Kohn–Sham system, the input quantities for each functional $F_\lambda(\rho_\Psi)$ and $D_\lambda(\Psi)$ were determined at the coupled-cluster singles-and-doubles (CCSD) level of theory. All calculations were carried out with the QUEST rapid development platform [63]; an electronic-structure code developed in Python and exploiting just-in-time compilation using the NUMBA package [119,120].

4.4.2 Kohn–Sham Non-Interacting System

In Table 4.1, the optimised values of the non-interacting Lieb functional $F_0(\rho_\Psi)$ and Gidopoulos–Davidson functional $D_0(\Psi)$ are presented for a series closed–shell atoms and for the hydrogen molecule at several bond lengths. Additionally, Kohn–Sham energy components are presented, including the internuclear repulsion energy E_{nn}, the non-interacting

Species	F_0	D_0	E_{nn}	T_s	E_{ne}	E_J	E_x	E_c	T	W	$\mathcal{E}_1(v)$
He	2.8611	0.0355	0.0000	2.8611	−6.7455	2.0464	−1.0232	−0.0756	2.8967	0.9477	−2.9011
Be	14.5835	0.0661	0.0000	14.5835	−33.6945	7.2122	−2.6725	−0.1534	14.6496	4.3863	−14.6586
Ne	128.5050	0.2720	0.0000	128.5050	−310.9007	65.9350	−12.0691	−0.6071	128.7765	53.2589	−128.8654
H_2 ($R = 0.7$)	1.7263	0.0324	1.4286	1.7263	−4.8614	1.6508	−0.8254	−0.0700	1.7588	0.7554	−0.9187
H_2 ($R = 1.4$)	1.1390	0.0328	0.7143	1.1390	−3.6469	1.3215	−0.6607	−0.0729	1.1718	0.5879	−1.1729
H_2 ($R = 3.0$)	0.8279	0.0418	0.3333	0.8279	−2.6181	0.9539	−0.4769	−0.1184	0.8697	0.3586	−1.0564
H_2 ($R = 5.0$)	0.9520	0.0224	0.2000	0.9520	−2.3809	0.8193	−0.4097	−0.2063	0.9744	0.2033	−1.0033
H_2 ($R = 7.0$)	0.9919	0.0052	0.1429	0.9918	−2.2826	0.7669	−0.3835	−0.2406	0.9971	0.1429	−0.9998
H_2 ($R = 10.0$)	0.9981	0.0005	0.1000	0.9982	−2.1983	0.7245	−0.3623	−0.2623	0.9986	0.1000	−0.9997

Table 4.1: Optimised functional values and energy components calculated in the aug-cc-pVTZ basis using the variation principles of (4.7) & (4.8). All quantities are in atomic units.

kinetic energy T_s, the electron–nuclear attraction E_{ne}, the Coulomb energy E_J, the exchange energy E_x, and the correlation energy E_c. These components have the same definition when computed from $F_0(\rho_\Psi)$ and $D_0(\Psi)$. For comparison, the total kinetic energy T and total electron–electron interaction energy W are included, along with the total interacting ground-state energy $\mathcal{E}_1(v)$.

The consistency of the optimisations was verified by comparing the optimised values of $F_0(\rho_\Psi)$ and $D_0(\Psi)$ presented in Table 4.1 with the energetic components T_s and T_c respectively. The value of T_s was determined from the Kohn–Sham orbitals obtained at $\lambda = 0$ and the value of T_c was obtained by subtraction of T_s from T, where the latter was determined directly from the $\lambda = 1$ calculation.

The H_2 molecule provides a simple prototypical system with which the variation between dynamic and static correlation may be explored. At equilibrium geometry, the electron densities of the two hydrogen atoms overlap substantially, thus binding the molecule and leading to both kinetic and potential contributions to the correlation energy. As the interatomic bond is extended, the system approaches that of two isolated hydrogen atoms, with no kinetic correlation energy; see Table 4.1, where the value of the Gidopoulos–Davidson functional D_0 decreases as the interatomic bond length R increases, becoming just 0.0005 a.u. at $R = 10.0$ a.u.

4.4.3 General Interaction Strengths

In Figure 4.1, results of optimisations pertaining to the generalised Lieb and Gidopoulos–Davidson functionals, according to (4.6) & (4.25) respectively, are presented for interaction strength λ in the range 0 to 1.

In the upper panel, the Lieb functional $F_\lambda(\rho_\Psi)$ is shown as a function of λ for the H_2 molecule with bond length $R = 0.7$, 1.4, 3.0, 5.0, 7.0 and 10.0 a.u. The variation of $F_\lambda(\rho_\Psi)$ in λ is broadly linear, indicating that $T_{c,\lambda} = T_1 - T_\lambda$ is relatively small and reflecting the dominance of the Coulomb and exchange energies in the two–electron energy W, both of which are linear in λ. The slope of $F_\lambda(\rho_\Psi)$ in λ becomes progressively smaller as the bond length is increased. This behaviour reflects the fact that the H_2 molecule dissociates into two one-electron fragments with $\lambda E_J + \lambda E_x + E_{c,\lambda} \to 0$ as $R \to \infty$ (static correlation energy cancelling the Coulomb and exchange energy).

In the lower panel of Figure 4.1, the Gidopoulos–Davidson functional $D_\lambda = T_1 - T_\lambda + \lambda(W_1 - W_\lambda)$ is also plotted as a function of interaction strength. This functional adopts the value of T_c at $\lambda = 0$ and decreases with increasing λ to become 0 at $\lambda = 1$. In contrast to the Lieb functional, this small correlation contribution to the energy reveals the higher-order dependence of the correlation energy on λ at increasingly extended bond lengths. As the bond length R increases and the system approaches one of independent atoms, the value of the Gidopoulos–Davidson functional is smaller at $\lambda = 0$, reflecting a decrease in T_c. However, it also exhibits more pronounced curvature, indicating higher-order dependence on λ as static correlation becomes more significant.

4.4.4 Constructing The Adiabatic Connection

As described in Section 3.6, the adiabatic connection (AC) comprises a link between the non-interacting Kohn–Sham auxiliary system and the physically interacting system through variation in interaction strength, modulated by coupling–constant λ, with the density equal to the physical density of $\lambda = 1$ for all λ. The AC integrand is expressed in (3.177), from which an exact definition of the correlation energy may be constructed according to (3.180). Given that the exchange energy scales linearly with λ (for the linear–attenuation AC path), the exchange contribution to (3.177) is simply a constant and may be subtracted to give the correlation component of the AC integrand,

$$\mathcal{W}_{c,\lambda}(\rho) = \mathcal{W}_\lambda(\rho) - \mathcal{W}_0(\rho). \tag{4.31}$$

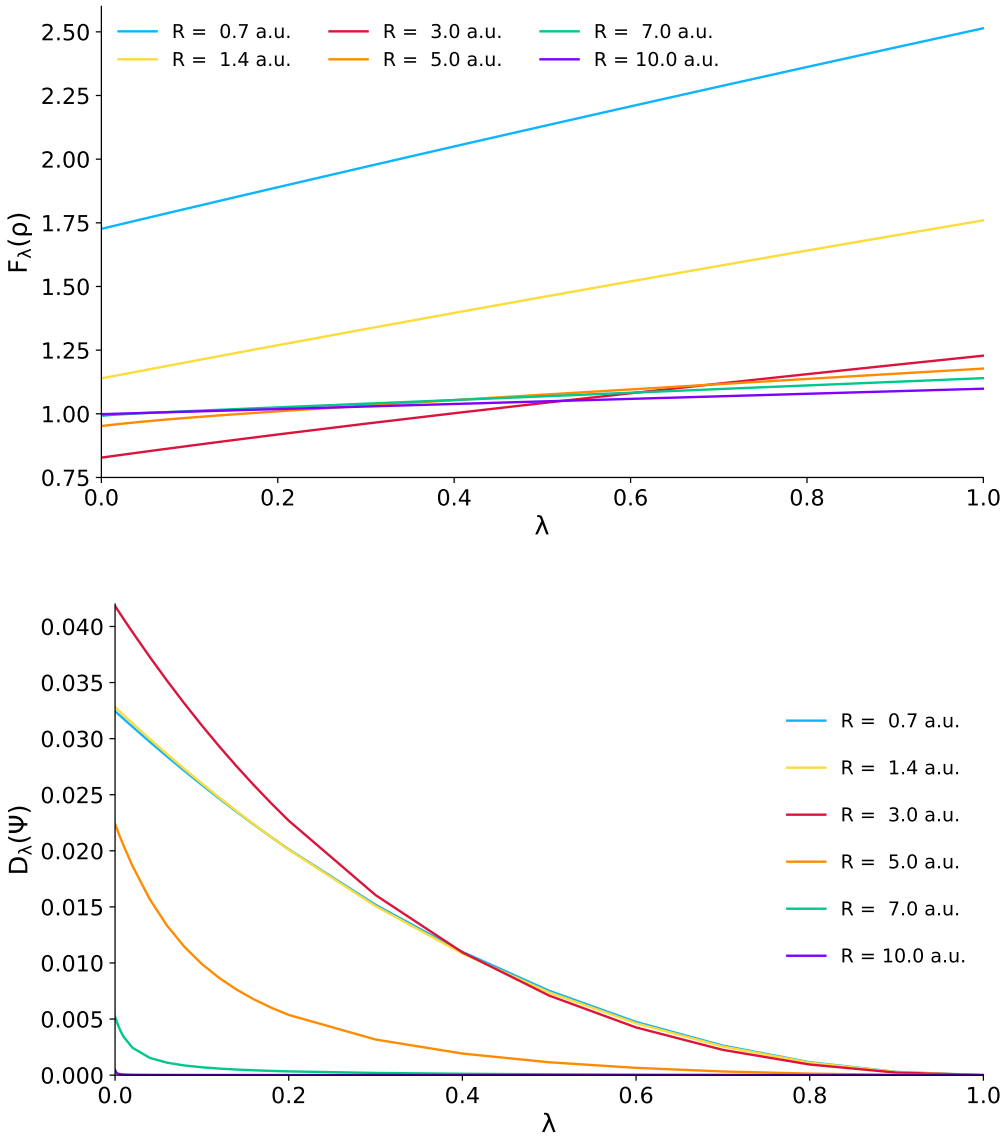

Figure 4.1: F_λ of (4.6), upper panel, and D_λ of (4.25), lower panel (a.u.) as functions of the interaction strength λ for the H_2 molecule at internuclear separations $R = 0.7$, 1.4, 3.0, 5.0, 7.0 and 10.0 a.u.

The Gidopoulos–Davidson variation principle of (4.25) and the Lieb variation principle of (4.6) can both be exploited to calculate this integrand, using the same input ρ_Ψ or Ψ but with a range of different values of λ, to construct the AC using (4.31).

The equivalence of the AC curves constructed from the Lieb and Gidopoulos–Davidson functionals is confirmed numerically for the H_2 molecule at the same geometries considered in Table 4.1, with the AC integrands $\mathcal{W}_{c,\lambda}$ plotted as a function of λ in Figure 4.2. Here, values of $\mathcal{W}_{c,\lambda}$ computed with the Lieb functional (4.6) are represented by solid lines, whilst values obtained from the Gidopoulos–Davidson functional (4.25) are denoted by the point markers. It is evident from Figure 4.2 that the AC curves of these two methods agree to within the convergence of the optimisation procedures.

The correlation energy can be computed from these curves using (3.180) and the numerical values of E_c are presented in Table 4.1. The ratio $|E_c|/T_c$ has been used to assess the relative importance of static correlation[121]. The $|E_c|$ corresponds to the area above each curve in Figure 4.2, whilst T_c corresponds to the area between each curve and a horizontal line defined by its value of $W_1(\Psi)$. As R increases, this ratio grows and the curves approach an L shape characteristic of systems dominated by strong correlation, indicating that the value of T_c is approaching zero, consistent with the effects of hydrogen molecule dissociation discussed in subsection 4.4.3.

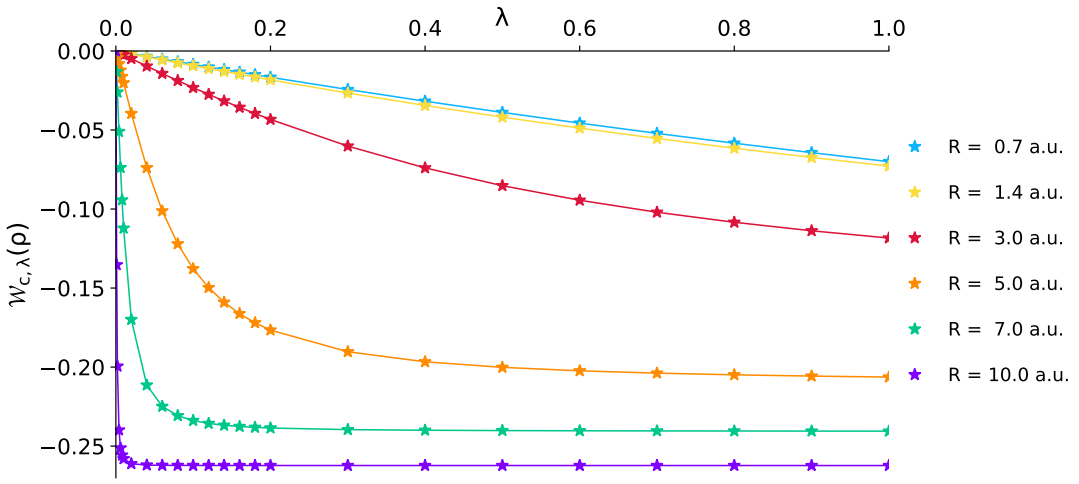

Figure 4.2: The correlation adiabatic connection integrand values (a.u.) of (4.31), calculated using the optimisation of (4.6), lines, and (4.25), point markers, for the H_2 molecule at internuclear separations $R = 0.7$, 1.4, 3.0, 5.0, 7.0 and 10.0 a.u.

4.5 Conclusions

The variation principle proposed in different contexts by Gidopoulos[108] and Davidson[109] has been examined and shown to be closely linked to the Lieb variation principle[67]. For the non-interacting system, the two functionals approach the Kohn–Sham system from different directions. The Lieb functional minimises the non-interacting kinetic energy T_s subject to the constraint that the density is equal to that of the physical system, whereas the Gidopoulos–Davidson functional maximises the kinetic correlation energy T_c under the same density constraint. In both cases, an unconstrained optimisation can be conducted with respect to the potential expansion coefficients in (4.30), allowing a straightforward implementation similar to those described in Refs. 111,112. The external potential is hence a Lagrange multiplier, ensuring that the density constraint is satisfied at the stationary point for each functional.

An extension of the Gidopoulos–Davidson functional to ensembles was also presented, for which the associated functional can be shown to be concave with respect to γ. This contrasts the pure-state functional which is not concave with respect to Ψ. A further extension to treat general electronic interaction strengths λ has also been presented, as has previously been done with the Lieb functional[67,110–112]. Utilising this extension, it was shown that either functional may be used to calculate the adiabatic connection between the Kohn–Sham system of non-interacting electrons and the physically-interacting system, highlighting the fact that the two functionals are essentially equivalent, differing simply by a change of sign and addition of the constant $T(\Psi)$. As such, they are amenable to implementation in a common computational framework.

5 Energy Densities In Atoms & Molecules

5.1 Introduction

A great deal of progress has been made in the development of increasingly accurate Kohn–Sham[65] (KS) density-functional approximations (DFAs) in recent decades. Over this time, there has been a general shift away from simple DFAs, constructed using the purely local information such as the electron density and its derivatives, towards increasingly non-local models involving the KS orbitals directly[81,122–124]. In most cases, this is limited to the use of occupied orbitals only via orbital-dependent exchange or dependence on the kinetic energy density, however some models extend this to include virtual orbitals and their energies to allow a more accurate treatment of correlation.

Historically, the progression from local density approximation (LDA) models to generalised gradient approximation (GGA) models delivered a step change in the accuracy of practical KS calculations. The development of hybrid functionals, that include a proportion of exact (orbital) exchange, has led to further significant improvements in the accuracy of thermochemical predictions. However, although several meta–GGA functionals have become well established[93,125–127], their performance in many other applications only amounts to an incremental refinement over that of standard GGAs.

Questions remain as to how best local functionals may be constructed to give a higher level of accuracy than that typical of GGAs. The meta-GGAs go beyond the GGA level by including additional dependencies; most meta-GGAs utilise the kinetic energy density τ, however a small number instead use the Laplacian of the density $\nabla^2\rho$. In addition to the best choice of variables, it is not clear under which circumstances one might expect substantial improvements by using meta-GGA type constructions.

In an attempt to shed light on the strengths and weaknesses of existing DFAs, a number of groups have developed techniques for the calculation of accurate density-functional quantities using systematically improvable *ab initio* methods. Such approaches include: *ab initio* DFT[128] in which orbital-dependent exchange–correlation functionals are developed and the corresponding KS equations solved by the optimised effective potential (OEP) method[129,130], inversion techniques that deliver KS potentials, orbitals and orbital energies[114,131,132], and techniques to compute the adiabatic connection (AC)[97–100] linking the non–interacting KS system to the physical system[110,111,114].

The success of the KS approach is largely due to the availability and broad applicability of simple, local DFAs. Whilst each of the techniques that employ a combination of *ab initio* and DFT methods have been useful in elucidating the shortcomings of simple DFAs, it is often difficult to use these results to gain direct insight into how to construct / modify simple local DFAs. In this work, we use the method of Lieb optimisation to calculate accurate energy densities for systems along the AC, from which we can derive the coupling constant averaged (λ–averaged) exchange–correlation energy density. This *local* function can be regarded as a target for modelling by simple DFAs.

In Chapter 3, the theoretical framework underlying DFT was discussed in detail, including the definition of the Lieb functional in Section 3.4 and the AC in Section 3.6. These two concepts are of particular importance for the present discussion; Section 5.2 builds on the AC approach to obtain an unambiguous definition for the exchange–correlation energy density, with a well–defined method for its calculation at any point along the AC derived using the Lieb functional. In this analysis, an expression for the λ–averaged energy density is presented and its relevance as a target to be modelled by novel DFAs discussed. Subsequent to this, energy densities computed from accurate *ab initio* densities are presented and analysed for a collection of atoms and small molecules.

In Section 5.3, the exchange and correlation energy densities pertaining to four closed–shell atoms are given, along with the respective computational details. In subsection 5.3.2, the atomic exchange energy densities are compared with the Becke–Roussel exchange energy model[93]; a simple DFA that is defined in the correct gauge to allow such a side–by–side comparison, in subsection 5.3.3, the atomic correlation energy densities are plotted for each value of λ alongside the corresponding λ–averaged energy density, and in subsection 5.3.4 properties of the local adiabatic connection integrand are considered, as this can provide further insight into the departure from exchange–only behaviour at each point in space.

In Section 5.4, the interesting example of correlation in a dissociating H_2 molecule is considered from the perspective of correlation energy densities, using both the global and local adiabatic connections computed from *ab initio* density matrices to examine the relative significance of static and dynamical correlation in H_2 as it approaches dissociation.

The patterns observed in the λ–averaged energy densities are compared with those of local observables typically used in the construction of DFAs in Section 5.5. Energy densities computed for the diatomic molecules HF & N_2 are presented here alongside atomic energy densities to examine the patterns exhibited by the λ–averaged energy density in the presence of a chemical bond and how this contrasts to the picture for atoms.

In Section 5.6, some initial results of the spin–resolved λ–averaged energy densities and local adiabatic connections are presented. Comparisons between the spin–parallel and spin–antiparallel contributions indicate notable differences in their properties, consideration of which could be useful in improving the performance of DFAs. Finally, in Section 5.7 an overview on these findings and the relevance of this data is given, and the direction of further work set–out.

5.2 Energy Densities In DFT

5.2.1 Exchange–Correlation Energy Densities

In Section 3.6, an expression for the exchange–correlation energy (*xc* energy) was derived using the AC approach. Whilst the integrand of the AC in (3.177) is a *global* quantity for the system at interaction strength λ, it can be equivalently written as the spatial integral of a *local* quantity. This is similar to the expression relating the local value of an *xc* functional and the total *xc* energy in (3.159), however here it is more logical to relate this local quantity to the global AC integrand by integration with the density over all space,

$$E_{\mathrm{xc},\lambda}(\rho) = \int_0^\lambda \int \rho(\mathbf{r}) w_{\mathrm{xc},\lambda}(\mathbf{r})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\lambda, \tag{5.1}$$

where $w_{\mathrm{xc},\lambda}(\mathbf{r})$ is the *exchange–correlation energy density*, defined as the *xc* energy per electron, per unit volume at spatial position $\mathbf{r}$.[i] The principal focus of the work presented later in this document pertains to this quantity – elucidating its structure through experiment and, using these observations, attempting to accurately reconstruct it from local quantities such as the density and its derivatives.

It is important to note here that this energy density cannot be *uniquely* defined; an arbitrary number of terms may be added to $w_{\mathrm{xc},\lambda}(\mathbf{r})$ and the same $\mathcal{W}_\lambda$ will be recovered if their spatial integral is zero. Thus an identical global AC could be constructed from incomparable local quantities, hence the only meaningful comparisons that can be made of $w_{\mathrm{xc},\lambda}(\mathbf{r})$ are with other local quantities defined in the *same gauge*. For many *xc*–functionals, the mathematical manipulations used to make the local form simple to evaluate also have the effect of changing the gauge of the functional such that it is no longer in the gauge of its original (physically–derived) model but in an alternative and less useful gauge instead. For further discussion of this point, see Refs. 133,134.

In the context of the AC however, there is a gauge in which it is both convenient and physically meaningful to define the energy density; the gauge of the *exchange–correlation hole*[134,135]. In this gauge, the *xc* energy density is defined as

$$w_{\mathrm{xc},\lambda}(\mathbf{r}_1) = \frac{1}{2}\int \frac{h_{\mathrm{xc},\lambda}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2. \tag{5.2}$$

The *xc* hole $h_{\mathrm{xc}}(\mathbf{r}_1,\mathbf{r}_2)$ may be interpreted physically as the probability of finding an electron at position $\mathbf{r}_2$ given an electron at position $\mathbf{r}_1$, hence it is an expression of conditional probability. It follows that the *xc* hole is itself defined in terms of the density at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ and the *two–electron density* $P_2(\mathbf{r}_1,\mathbf{r}_2)$, which describes for two electrons the probability that one is at $\mathbf{r}_1$ whilst the other is simultaneously at $\mathbf{r}_2$,

$$P_2^\lambda(\mathbf{r}_1,\mathbf{r}_2) = N(N-1)\int\ldots\int |\Psi_\lambda(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\ldots,\boldsymbol{\tau}_N)|^2\,\mathrm{d}\sigma_1\mathrm{d}\sigma_2\mathrm{d}\boldsymbol{\tau}_3\ldots\mathrm{d}\boldsymbol{\tau}_N \qquad \boldsymbol{\tau}_i=\{\mathbf{r}_i,\sigma_i\}, \tag{5.3}$$

with which the *xc*–hole may be written as

$$h_{\mathrm{xc},\lambda}(\mathbf{r}_1,\mathbf{r}_2) = \frac{P_2^\lambda(\mathbf{r}_1,\mathbf{r}_2)}{\rho(\mathbf{r}_1)} - \rho(\mathbf{r}_2). \tag{5.4}$$

As implied in (5.3), the two–electron density is *not* λ–invariant in the same way that the one–electron density is, hence it evolves through the AC as Ψ_λ does. The *xc* energy density may now be written directly as a function of the one and two–electron densities as

$$w_{\mathrm{xc},\lambda}(\mathbf{r}_1) = \frac{1}{2\rho(\mathbf{r}_1)}\int \frac{P_2^\lambda(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 - \frac{1}{2}\int\frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2. \tag{5.5}$$

[i] An equivalent expression may be written for correlation alone by considering a local form of the AC integrand defined in (3.179).

This definition of the energy density has been widely used in the literature, for example Refs. 136–139, and has recently been computed for $\lambda = 0, 1$ and ∞ by Mirtschink *et al.*[134]. The first term of (5.5) for $\lambda = 1$ may also be identified as the potential $v_{\text{cond}}(\mathbf{r})$ discussed in Ref. 140.

The choice of gauge implied by the definition of (5.2) is particularly convenient for several reasons. Firstly, it is that which is defined by the electrostatic potential of the *xc* hole for a given interaction strength; it has physical significance as a manifestation of exchange and correlation effects, thus suggesting that the variation of this energy density with respect to i) spatial position and ii) λ could serve as a somewhat intuitive picture of exchange and correlation effects.

Secondly, many properties of the *xc* energy density in this gauge are well understood, in particular it is known that

$$\lim_{|\mathbf{r}|\to\infty} w_{\text{xc}}(\mathbf{r}) = -\frac{1}{2|\mathbf{r}|}, \tag{5.6}$$

providing a useful asymptotic bound on observed energy densities[141,142]. Thirdly, for $\lambda = 0$ the KS system is described by a single Slater determinant, hence the corresponding energy density reduces to the exchange–only energy density, which is independent of λ and can be written as

$$w_{\text{x}}(\mathbf{r}_1) = w_{\text{xc},0}(\mathbf{r}_1) = \frac{1}{2}\int \frac{h_{\text{x}}(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 \tag{5.7}$$

where $h_{\text{x}}(\mathbf{r}_1, \mathbf{r}_2)$ is the *exchange hole*. As a result, the *xc* energy density may be resolved into exchange and correlation components at any given point along the AC; the λ–dependent correlation energy density may therefore be isolated as

$$w_{\text{c},\lambda}(\mathbf{r}) = w_{\text{xc},\lambda}(\mathbf{r}) - w_{\text{x}}(\mathbf{r}). \tag{5.8}$$

The final and most compelling reason for the choice of gauge in (5.2) is that the only quantity (other than $\rho(\mathbf{r})$) required to compute the energy density is the λ–dependent two–electron density $P_2^{\lambda}(\mathbf{r}_1, \mathbf{r}_2)$, for which well–established methods of accurate calculation are available based on *ab initio* theories; the two–particle density in coupled cluster theory was derived using the method of analytical derivatives in subsection 2.5.2 and implemented in the spin–unrestricted form for the present work[63].

5.2.2 Coupling–Constant Averaged Energy Densities

Whilst the structure of the of the energy densities at each interaction strength λ is of interest, integration of these functions with the density ρ over space yields only the global adiabatic integrand $\mathcal{W}_{\lambda}$ of (3.177). This quantity can then be integrated over λ to obtain the total *xc* energy of the system. Instead, this work focuses on the λ–averaged energy density $\bar{w}_{\text{xc}}(\mathbf{r})$. The use of this quantity reduces the calculation of the *xc* energy to a single spatial integration. Considering the definition that $(\rho_{\lambda} = \rho_1)\ \forall\, \lambda \in \{0, 1\}$, a simple re–arrangement of the AC formula allows the λ–averaged energy to be defined as

$$\bar{w}_{\text{xc}}(\mathbf{r}) = \int_0^1 w_{\text{xc},\lambda}(\mathbf{r})\mathrm{d}\lambda, \tag{5.9}$$

with which the *xc* energy is computed as

$$E_{\text{xc}}(\rho) = \int \bar{w}_{\text{xc}}(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{5.10}$$

This λ–averaged *xc* energy density may be trivially resolved into exchange and correlation components, the form of which can be seen from (5.8) to be

$$\bar{w}_{\text{x}}(\mathbf{r}) = w_{\text{x}}(\mathbf{r}) \qquad \bar{w}_{\text{c}}(\mathbf{r}) = \int_0^1 \{w_{\text{xc},\lambda}(\mathbf{r}) - w_{\text{x}}(\mathbf{r})\}\mathrm{d}\lambda = \bar{w}_{\text{xc}}(\mathbf{r}) - w_{\text{x}}(\mathbf{r}). \tag{5.11}$$

Given the physical interpretation of the λ–averaged energy density, it follows that the function $\bar{w}_{\text{xc}}(\mathbf{r})$ and its components may be considered as targets to be modelled by new and improved *density functional approximations*. The λ–dependent $w_{\text{xc},\lambda}(\mathbf{r})$ defines a *local* AC integrand, as opposed to the *global* quantity $\mathcal{W}_{\lambda}$ in (3.177). Although this integrand must then be evaluated spatially (practically, through numerical quadrature over a large spatial grid), the local approach offers some important advantages. Firstly, as outlined in Ref. 143, modelling of local quantities inherently results in size–consistent AC models in the absence of degeneracy, something that is not easily achieved using non–linear interpolation models of the global AC[112,144–146].

Secondly, the local nature of this energy density is conceptually similar to the energy densities often calculated by conventional DFAs for exchange and correlation; although care must be taken to ensure a consistent gauge in

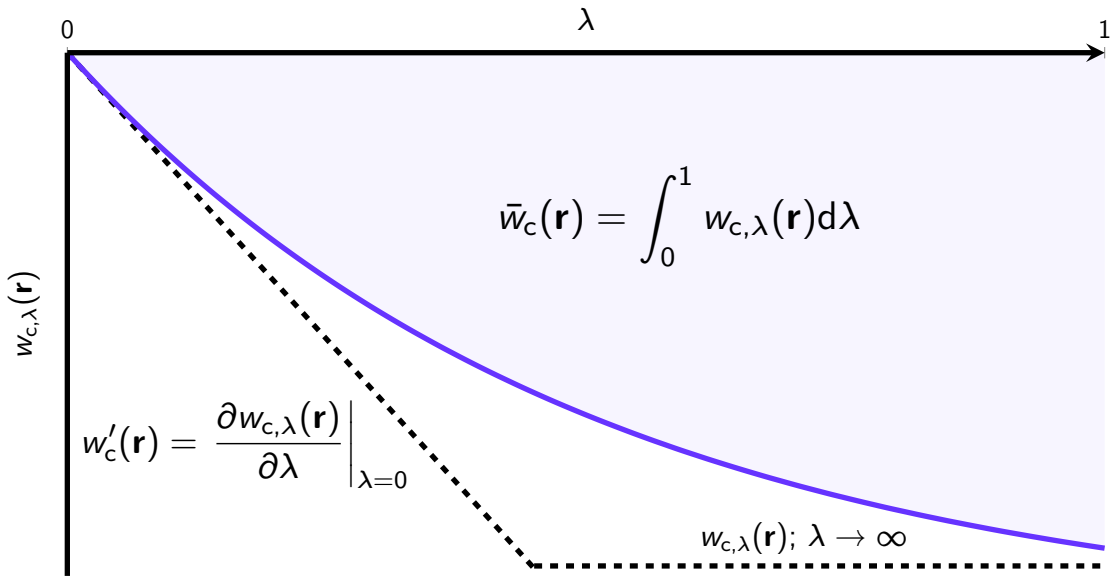

Figure 5.1: A schematic of the local adiabatic connection at a point in space $\mathbf{r}$. The value of the λ–averaged correlation energy density $\bar{w}_c(\mathbf{r})$ is recovered by integration of this function over the electronic interaction strength λ

any comparisons, this information should provide valuable insight to guide to construction of new DFAs through comparison with local quantities of the density and/or KS orbitals. These two principles underpin much of the work discussed in Chapter 6, where various models for the local AC and $\bar{w}_c(\mathbf{r})$ are investigated.

The key quantities provided by the correlation component of the *local AC* that are examined in this work are illustrated in Figure 5.1. The integral of $w_{c,\lambda}(\mathbf{r})$, shown as the shaded area, defines the value of λ–averaged energy density $\bar{w}_c(\mathbf{r})$. However, further information may be elicited through studying the *shape* of this local integrand. Given that the exchange contribution at each point in space is constant in λ, the shape of the local integrand contains information about how the electron–interaction departs from exchange–only behaviour at that point. One possible way of quantifying this behaviour is through considering the gradient of this quantity at $\lambda = 0$, which will be examined in due course.

5.2.3 Computing Energy Densities From *Ab Initio* Theory

It is possible, with *ab initio* methods, to reach a very high accuracy through systematic increases in the level of theory and size of the basis set. Whilst such calculations have a very high computational cost, they are invaluable for the development of new models as they provide insight into the fundamental properties of electronic systems, through near–exact predictions for small model systems. An overview of these *ab initio* methods may be found in Chapter 2.

In this work high quality *ab initio* methods, principally CCSD & CCSD(T) implemented as described in Section 2.4 & subsection 2.5.2 but also the FCI implementation of Ref. 147, are used to calculate near–exact one–particle density matrices (which are subsequently stored). These are then taken as an input for a DFT calculation that optimises the universal density functional $F(\rho)$ with the constraint that the iterating density is equal to the input density. The solution to this *constrained search* may be used to calculate exchange–correlation energy densities according to (5.5) which, assuming that the *ab initio* density is exact and the constrained search is fully converged, are exact.

Although simple in principle, such a calculation in practice requires a carefully planned method and a robust optimisation algorithm to reach a well–converged solution. The following subsections outline the methods developed to do this, concluding with the principal method used in this work – the *Lieb optimisation*.

5.2.3.1 The Wu–Yang Constrained Search

The *Wu–Yang* (WY) constrained search, published by WY in 2003[114], is a method by which the Kohn–Sham orbitals can be re–constructed from an accurate input density, modelled as non–interacting v–representable.

For an input density ρ_1, the KS non–interacting kinetic energy is given by

$$T_s(\rho_1) = \min_{\Phi\to\rho_1} T(\Phi), \tag{5.12}$$

where Φ is an N–electron Slater determinant wavefunction with N orthonormal one–electron spinorbitals $\{\phi_i\}$. Thus the calculation of $T_s(\rho_1)$ is a constrained minimisation over all such Φ that yield density ρ_1. It is noted that, considering the two spin components α & β of the density, the expression for the non–interacting kinetic energy (5.12) is given by

$$T_s(\rho_{1,\alpha},\rho_{1,\beta}) = \min_{\Phi\to\rho_{1,\alpha},\rho_{1,\beta}} T(\Phi) = \frac{1}{2}T_s(2\rho_{1,\alpha}) + \frac{1}{2}T_s(2\rho_{1,\beta}), \tag{5.13}$$

thus becoming the sum of two separate optimisations. For simplicity the present discussion will assume a spin–resolved density, with analysis pertaining to a single spin contribution ρ_σ, constructed from N^σ one–electron spatial orbitals φ^σ, with the number of occupied and unoccupied orbitals denoted $occ(\sigma)$ & $vir(\sigma)$ respectively, as

$$\rho_\sigma(\mathbf{r}) = \sum_i^{occ(\sigma)} \left|\varphi_i^\sigma(\mathbf{r})\right|^2. \tag{5.14}$$

The density constraint may be imposed by introducing a Lagrange multiplier function $v^\sigma(\mathbf{r})$, transforming the problem to the minimisation of the Lagrangian function

$$W_s^\sigma(\Phi, v^\sigma) = \sum_i^{occ(\sigma)} \langle\varphi_i^\sigma|\hat{T}|\varphi_i^\sigma\rangle + \int\{\rho_\sigma(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r})\}v^\sigma(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{5.15}$$

Therefore, to minimise $T(\Phi)$ subject to $\rho_\sigma = \rho_{1,\sigma}$, (5.15) must have reached a stationary value and the orbitals $\{\varphi_i^\sigma\}$ must be normalised, which can be written concisely as

$$\left[\hat{T} + v^\sigma(\mathbf{r})\right]\varphi_i^\sigma = \varepsilon_i^\sigma\varphi_i^\sigma, \tag{5.16}$$

where $\{\varepsilon_i^\sigma\}$ are a set of Lagrange multipliers that impose the normalisation constraint on $\{\varphi_i^\sigma\}$. Limiting the possible solutions Φ to only those in which $\{\varphi_i^\sigma\}$ are eigenfunctions of (5.16)[ii], Φ can be seen to be an explicit function of the N–electron coordinates $\mathbf{r}^N$ and an *implicit* function of $v^\sigma(\mathbf{r})$, while the individual φ_i^σ become implicit functions of $v^\sigma(\mathbf{r})$,

$$\Phi = \Phi(v^\sigma) \qquad\qquad \varphi_i^\sigma = \varphi_i^\sigma(v^\sigma). \tag{5.17}$$

For a fixed $v^\sigma(\mathbf{r})$, the minimisation of $W_s^\sigma(\Phi, v^\sigma)$ with normalised orbitals $\{\varphi_i^\sigma\}$ is simply the problem given by (5.16). Therefore, given the identities (5.17), the minimisation of W_s^σ with respect to the potential can be written simply as

$$W_s^\sigma(\Phi(v^\sigma), v^\sigma) = \min_{\varphi_i^\sigma,\langle\varphi_i^\sigma|\varphi_i^\sigma\rangle=1} W_s^\sigma(\Phi, v^\sigma). \tag{5.18}$$

It was shown by WY that, with this form of W_s^σ, its first functional derivative with respect to v^σ is the difference between the iterating density and the input density,

$$\frac{\delta W_s^\sigma(\Phi(v^\sigma), v^\sigma)}{\delta v^\sigma(\mathbf{r})} = \rho_\sigma(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r}). \tag{5.19}$$

Given that the variation of $W_s^\sigma(\Phi, v^\sigma)$ with respect to $v^\sigma(\mathbf{r})$ is zero at the minimum value of $T(\Phi)$, $T(\Phi)$ is only minimised when the density constraint is satisfied. WY further showed that the *second* functional derivative of W_s^σ with respect to v^σ is given by

$$\frac{\delta^2 W_s^\sigma(\Phi(v^\sigma), v^\sigma)}{\delta v^\sigma(\mathbf{r}')\delta v^\sigma(\mathbf{r})} = \frac{\delta\rho_\sigma(\mathbf{r})}{\delta v^\sigma(\mathbf{r}')} = \sum_i^{occ(\sigma)}\sum_a^{vir(\sigma)} \frac{\langle\varphi_i^\sigma|\delta v^\sigma|\varphi_a^\sigma\rangle\,\langle\varphi_a^\sigma|\delta v^\sigma|\varphi_i^\sigma\rangle}{\varepsilon_i^\sigma - \varepsilon_a^\sigma} + \text{c.c.} \leqslant 0 \tag{5.20}$$

The second functional derivative of $W_s^\sigma(\Phi(v^\sigma), v^\sigma)$ can be shown to be less than or equal to zero for an arbitrary $\delta v^\sigma(\mathbf{r})$, at any $v^\sigma(\mathbf{r})$, hence the functional itself is concave in $v^\sigma(\mathbf{r})$. This leads to the key statement of the WY constrained search, expressing the KS kinetic energy corresponding to input density $\rho_{1,\sigma}$ as the following unconstrained maximisation

$$T_s(\rho_1) = \max_{v^\sigma} W_s^\sigma(\Phi(v^\sigma), v^\sigma) \tag{5.21}$$

[ii]Hence the restriction of ρ to ground–state, non–interacting v–representable densities.

Equally important however is the model of $v^\sigma(\mathbf{r})$ proposed by WY for practical implementation of this scheme. In their model, initially published a year previously in the context of the *optimised effective potential* (OEP) problem[113], the potential is expanded in an *auxiliary basis* as

$$v^\sigma(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{ref}}(\mathbf{r}) + \sum_t b_t^\sigma g_t(\mathbf{r}), \tag{5.22}$$

where $v_{\text{ext}}(\mathbf{r})$ is the standard external potential due to the nuclei, $v_{\text{ref}}(\mathbf{r})$ is a fixed reference potential chosen to ensure correct appropriate asymptotic behaviour of $v^\sigma(\mathbf{r})$ and $\{g_t(\mathbf{r})\}$ are a set of Gaussian basis functions with expansion coefficients $\{b_t^\sigma\}$ to be determined through optimisation. The fixed reference potential suggested by WY was the *Fermi-Amaldi potential*[115], defined as

$$v_{\text{ref}}(\mathbf{r}) = \frac{N-1}{N}\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}', \tag{5.23}$$

which has the advantage of exhibiting the correct long–range behaviour for $v(\mathbf{r})$ and also being a fair approximation to the exact $v(\mathbf{r})$ at more intermediate ranges. This expansion for the potential can be substituted into (5.15) to give the expression in terms of the auxiliary basis set

$$\begin{aligned} W_s^\sigma(\Phi, v^\sigma) = \sum_i^{occ(\sigma)} \langle\varphi_i^\sigma|-\frac{1}{2}\nabla_i^2|\varphi_i^\sigma\rangle + \int\{v_{\text{ext}}(\mathbf{r}) + v_{\text{ref}}(\mathbf{r})\}\{\rho_\sigma(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r})\}\mathrm{d}\mathbf{r} \\ + \int\sum_t b_t^\sigma g_t(\mathbf{r})\{\rho_\sigma(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r})\}\mathrm{d}\mathbf{r}, \end{aligned} \tag{5.24}$$

with first and second derivatives

$$\begin{aligned} \frac{\delta W_s^\sigma(\Phi(v^\sigma), v^\sigma)}{\delta b_t^\sigma} &= \int g_t(\mathbf{r})\{\rho_\sigma(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r})\}\mathrm{d}\mathbf{r} \\ \frac{\delta^2 W_s^\sigma(\Phi(v^\sigma), v^\sigma)}{\delta b_t^\sigma \delta b_u^\sigma} &= \sum_i^{occ(\sigma)}\sum_a^{vir(\sigma)} \frac{\langle\varphi_i^\sigma|g_t|\varphi_a^\sigma\rangle\,\langle\varphi_a^\sigma|g_u|\varphi_i^\sigma\rangle}{\varepsilon_i^\sigma - \varepsilon_a^\sigma} + \text{c.c.}. \end{aligned} \tag{5.25}$$

These equations are the form of the WY optimisation used in practical implementations, as it allows the constrained search to be treated as an unconstrained maximisation of $W_s^\sigma(\Phi(v^\sigma), v^\sigma)$ with respect to the auxiliary basis set coefficients $\{b_t^\sigma\}$.

5.2.3.2 The Lieb Optimisation

The Lieb optimisation extends the WY constrained search such that the DFT energy of a system is optimised for a given input density at an arbitrary position along the AC, i.e. for $0 \leqslant \lambda \leqslant 1$ and using an arbitrary *ab initio* method to calculate the reference density[110–112,114].

As the name would suggest, the mathematical foundations for the Lieb maximisation lie in its use of Lieb's convex–conjugate density functional (3.133) and its well–defined characteristics, described in Section 3.4. It was shown by (3.131) that the variationally–defined energy $\mathcal{E}_\lambda(v)$ is concave in v for a Hamiltonian $\hat{\mathcal{H}}_\lambda(v)$; it can be seen that this is the case for all values of λ since the concavity of $\mathcal{E}_\lambda(v)$ in v arises from the linearity of $\hat{\mathcal{H}}_\lambda(v)$ in v, a result which is independent of λ. Therefore, the λ–dependent Hamiltonian $\hat{\mathcal{H}}_\lambda(v)$ yields ground–state energies $\mathcal{E}_\lambda(v)$, which must be concave in v for all $0 \leqslant \lambda \leqslant 1$.

The λ–dependent Lieb functional $F_\lambda(\rho)$ is written as the convex–conjugate to $\mathcal{E}_\lambda(v)$,

$$F_\lambda(\rho) = \sup_{v\in\mathcal{X}^*}\left\{\mathcal{E}_\lambda(v) - \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}\right\}, \tag{5.26}$$

which can be read–across directly from (3.133). As described in subsection 3.4.2, a further Legendre–Fenchel transform of a conjugate functional does not necessarily recover the original functional. This is only the case if the original function is closed–convex as described in Definition 3.8; the biconjugate functional is the convex envelope otherwise. Here, this means that the concave conjugate to the Lieb functional $F_\lambda(\rho)$, given by

$$\mathcal{E}_\lambda^{**}(v) = \inf_{\rho\in\mathcal{X}}\left\{F_\lambda(\rho) + \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}\right\} \tag{5.27}$$

is only equal to the original energy $\mathcal{E}_\lambda(v)$ in (5.26) if the original energy was concave in v. In practical calculations, $\mathcal{E}_\lambda(v)$ may *not* be concave in v because it may not be fully variationally optimised; for such approximate energies, $\mathcal{E}_\lambda^{**}(v)$ is the concave envelope to $\mathcal{E}_\lambda(v)$.

For a given value of λ, the expression for the Lieb functional (5.26) at the optimising potential v_λ – a stationary point – is given simply by

$$F_\lambda(\rho) = \mathcal{E}_\lambda(v_\lambda) - (v_\lambda|\rho)\,. \tag{5.28}$$

Assuming that $F_\lambda(\rho)$ is differentiable (see subsection 3.5.2), it functional derivative with respect to the density $\rho(\mathbf{r})$ is given by

$$\frac{\delta F_\lambda(\rho)}{\delta\rho(\mathbf{r})} = -v_\lambda(\mathbf{r}). \tag{5.29}$$

The optimising potential may be resolved into separate components by substituting the definition for $F_\lambda(\rho)$ in (3.178) into (5.29), yielding

$$\begin{aligned}
\frac{\delta F_\lambda(\rho)}{\delta\rho(\mathbf{r})} &= \frac{\delta T_\mathrm{s}(\rho)}{\delta\rho(\mathbf{r})} + \lambda\frac{\delta J(\rho)}{\delta\rho(\mathbf{r})} + \lambda\frac{\delta E_\mathrm{x}(\rho)}{\delta\rho(\mathbf{r})} + \frac{\delta E_{\mathrm{c},\lambda}(\rho)}{\delta\rho(\mathbf{r})} \\
&= \frac{\delta T_\mathrm{s}(\rho)}{\delta\rho(\mathbf{r})} + \lambda v_\mathrm{J}(\mathbf{r}) + \lambda v_\mathrm{x}(\mathbf{r}) + v_{\mathrm{c},\lambda}(\mathbf{r}) \\
-v_\lambda(\mathbf{r}) &= -v_\mathrm{s}(\mathbf{r}) + \lambda v_\mathrm{J}(\mathbf{r}) + \lambda v_\mathrm{x}(\mathbf{r}) + v_{\mathrm{c},\lambda}(\mathbf{r}) \\
\implies v_\lambda(\mathbf{r}) &= v_\mathrm{s}(\mathbf{r}) - \lambda v_\mathrm{J}(\mathbf{r}) - \lambda v_\mathrm{x}(\mathbf{r}) - v_{\mathrm{c},\lambda}(\mathbf{r}).
\end{aligned} \tag{5.30}$$

At the non–interacting limit $\lambda = 0$, it can be seen that $v_\lambda(\mathbf{r})$ simply becomes $v_\mathrm{s}(\mathbf{r})$[iii], confirming the stationary condition for the noninteracting system. For the system at physical interaction strength $\lambda = 1$, $v_\lambda(\mathbf{r})$ must be equal to the external potential of the system $v_\mathrm{ext}(\mathbf{r})$, hence

$$v_\mathrm{s}(\mathbf{r}) = v_\mathrm{ext}(\mathbf{r}) + v_\mathrm{J}(\mathbf{r}) + v_\mathrm{x}(\mathbf{r}) + v_{\mathrm{c},1}(\mathbf{r}). \tag{5.31}$$

Substituting this back into (5.30) to eliminate $v_\mathrm{s}(\mathbf{r})$ results in the expression for the effective potential

$$v_\lambda(\mathbf{r}) = v_\mathrm{ext}(\mathbf{r}) + (1-\lambda)v_\mathrm{J}(\mathbf{r}) + (1-\lambda)v_\mathrm{x}(\mathbf{r}) + [v_{\mathrm{c},1}(\mathbf{r}) - v_{\mathrm{c},\lambda}(\mathbf{r})]\,, \tag{5.32}$$

where $v_\lambda(\mathbf{r})$ is the potential that binds the density at λ such that it is equal to that of $\lambda = 1$. For practical calculations, this expression immediately suggests the following modification to the WY expansion of $v^\sigma(\mathbf{r})$ of (5.22),

$$v^\sigma_\lambda(\mathbf{r}) = v_\mathrm{ext}(\mathbf{r}) + (1-\lambda)v_\mathrm{ref}(\mathbf{r}) + \sum_t b^\sigma_t g_t(\mathbf{r}). \tag{5.33}$$

With this, the Lieb functional may be re–written as a function of the expansion coefficients of the auxiliary basis,

$$G_{\lambda,\rho}(\{b^\sigma_t, b^{\sigma'}_t\}) = \mathcal{E}_\lambda(v^\sigma_\lambda, v^{\sigma'}_\lambda) - \int\{v^\sigma_\lambda(\mathbf{r})\rho_\sigma(\mathbf{r}) + v^{\sigma'}_\lambda(\mathbf{r})\rho_{\sigma'}(\mathbf{r})\}\mathrm{d}\mathbf{r}. \tag{5.34}$$

The maximisation of $G_{\lambda,\rho}(\{b^\sigma_t, b^{\sigma'}_t\})$ with respect to $\{b^\sigma_t, b^{\sigma'}_t\}$, for a given electronic structure method that yields $\mathcal{E}_\lambda(v^\sigma_\lambda, v^{\sigma'}_\lambda)$, is the mechanism by which the Lieb optimisation is carried–out in practical calculations.

Convergence of the Lieb optimisation can be accelerated by the use of Newton methods as employed by WY[113,114], with the first and second functional derivatives of the objective functional $G_{\lambda,\rho}(\{b^\sigma_t, b^{\sigma'}_t\})$ with respect to $\{b^\sigma_t, b^{\sigma'}_t\}$ given by

$$\begin{aligned}
\frac{\delta G_{\lambda,\rho}(\{b^\sigma_t, b^{\sigma'}_t\})}{\delta b^\sigma_t} &= \int[\rho_{v,\sigma}(\mathbf{r}) - \rho_{1,\sigma}(\mathbf{r})]g_t(\mathbf{r})\mathrm{d}\mathbf{r} \\
\frac{\delta^2 G_{\lambda,\rho}(\{b^\sigma_t, b^{\sigma'}_t\})}{\delta b^\sigma_t \delta b^\sigma_u} &= \int\int g_t(\mathbf{r})g_u(\mathbf{r}')\frac{\delta\rho_{v,\sigma}(\mathbf{r})}{\delta v^\sigma(\mathbf{r}')}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'.
\end{aligned} \tag{5.35}$$

To use this optimisation with the full Hessian for general λ and a variety of non-variational wavefunctions, care must be taken to ensure that the correct $\delta\rho_{v,\sigma}(\mathbf{r})/\delta v^\sigma(\mathbf{r}')$ is evaluated. This is the approach taken for the results presented in Section 5.3, in which the relaxed Lagrangian formulation of Helgaker and Jørgensen[148] is used to obtain *relaxed* densities for non–variational wavefunctions, with which $\delta\rho_{v,\sigma}(\mathbf{r})/\delta v^\sigma(\mathbf{r}')$ may be calculated using response theory[111,112] – implemented in the DALTON quantum chemistry software package[149,150].

However elsewhere in this work – including the calculations presented in Section 4.4 and those discussed in Sections 5.4, 5.5 & 5.6, the full Hessian is not computed but rather approximated by the non–interacting equivalent in (5.25). This *approximate Newton* method[118] primarily uses the coupled cluster methods described in Chapter 2 to compute the fully spin–unrestricted $\mathcal{E}_\lambda(v^\sigma_\lambda, v^{\sigma'}_\lambda)$ & relaxed densities $\{\rho_{v,\sigma}, \rho_{v,\sigma'}\}$; these methods were implemented for this work in the QUEST rapid development electronic structure package[63].

Once the maximisation of (5.26) is complete, the optimising potential v^σ_λ is such that $\rho_{\lambda,\sigma} = \rho_{1,\sigma}$. In addition, $\mathcal{E}_\lambda(v^\sigma_\lambda, v^{\sigma'}_\lambda)$ and the relaxed λ–interacting one and two particle Lagrangian density matrices are readily obtained. As a

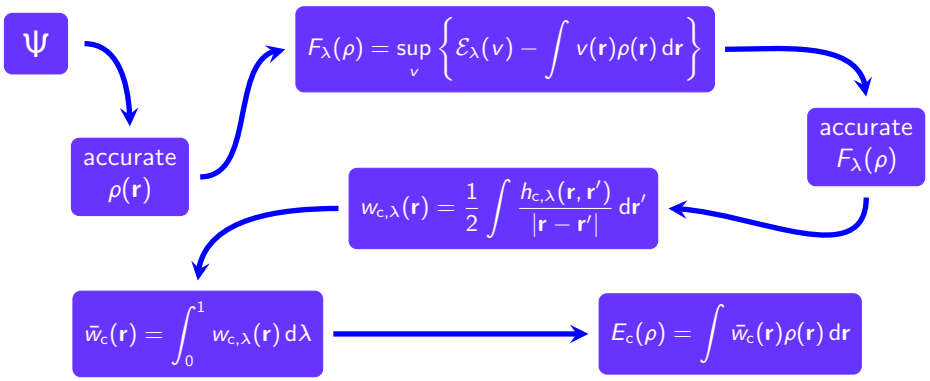

Figure 5.2: A schematic for the calculation of energy densities by Lieb optimisation.

result, the quantities required for the construction of $w_{xc,\lambda}(\mathbf{r})$ in (5.5) at the optimised solution are available and it may be evaluated at arbitrary points in space. A schematic for the Lieb optimisation is given in Figure 5.2. In this work, $w_{xc,\lambda}(\mathbf{r})$ was evaluated on both a standard DFT integration grid – the Lebedev angular and Lindh-Malmqvist-Gagliardi (LMG) radial quadrature[151] implemented in the DALTON quantum chemistry software package[149,150], and also on evenly spaced Cartesian grids to enable a detailed analysis of the spatial variation of $w_{xc,\lambda}(\mathbf{r})$ to be undertaken. In the following sections, the results of such calculations are presented for a selection of atoms and small molecules, for which the energy densities are analysed.

5.3 Energy Densities In Atoms

5.3.1 Computational Details

All of the quantities required for the evaluation of the energy densities using (5.5) and the λ–averaged energy density defined in (5.10) are computed with a development version of the DALTON quantum chemistry program[149,150]. In particular, we have modified the DFT module to allow the evaluation of these energy densities on standard quadrature grids and also on evenly spaced Cartesian grids to allow easier analysis.

Here, we consider the application of this method for calculating energy densities to the closed shell atoms He, Be, Ne and Ar. Accurate electron densities ρ_1 for these systems have been computed at the CCSD(T) level of theory[46] using the *u*-aug-cc-pCVQZ basis set for He, Be & Ne and the *u*-aug-cc-pCVTZ basis for Ar, where the prefix *u* signifies that the basis set has been fully uncontracted. These densities are used as the input for Lieb optimisation calculations at all λ. The Lieb optimisation uses the implementation described in Refs. 111,112, in which the basis set used to expand the external potential v, as in (5.33) is the same as that used to model the molecular orbitals.

In these calculations, the Hessian of $F_\lambda(\rho)$ is numerically regularised via a singular value decomposition with threshold 10^{-7}. The calculations were considered converged if the 2-norm of the gradient of $F_\lambda(\rho)$ fell below 10^{-6}. To obtain an accurate estimate of the λ–averaged energy densities $\bar{w}_{xc}(\mathbf{r})$, calculations were carried–out for the following values of λ:

$$\lambda \to \left\{ \begin{matrix} 0.0000, 0.0005, 0.0010, 0.0015, 0.0020, 0.0040, 0.0060, 0.0080, 0.0100, 0.0500 \\ 0.1000, 0.2000, 0.3000, 0.4000, 0.5000, 0.6000, 0.7000, 0.8000, 0.9000, 1.0000 \end{matrix} \right\}$$

These values were selected to provide sufficient samples over the range $\lambda = 0 \to 1$ for integration over the coupling constant. In addition, a number of densely spaced values of λ close to 0 were included so that the behaviour of the local adiabatic connection in the non–interacting limit can be examined more closely (see subsection 5.3.4).

5.3.2 Exchange Energy Densities

The exchange energy density, w_x, is independent of λ at each point in space. This function is examined for the He, Be, Ne and Ar atoms in Figure 5.3. It is well-documented[152,153] that the exchange-energy density in atomic systems varies radially from the nucleus in a relatively smooth and uncomplicated way; the data presented in Figure 5.3 mirrors those observations.

[iii] Since the correlation energy vanishes at $\lambda = 0$, it follows that $v_{c,0}(\mathbf{r}) = 0$.

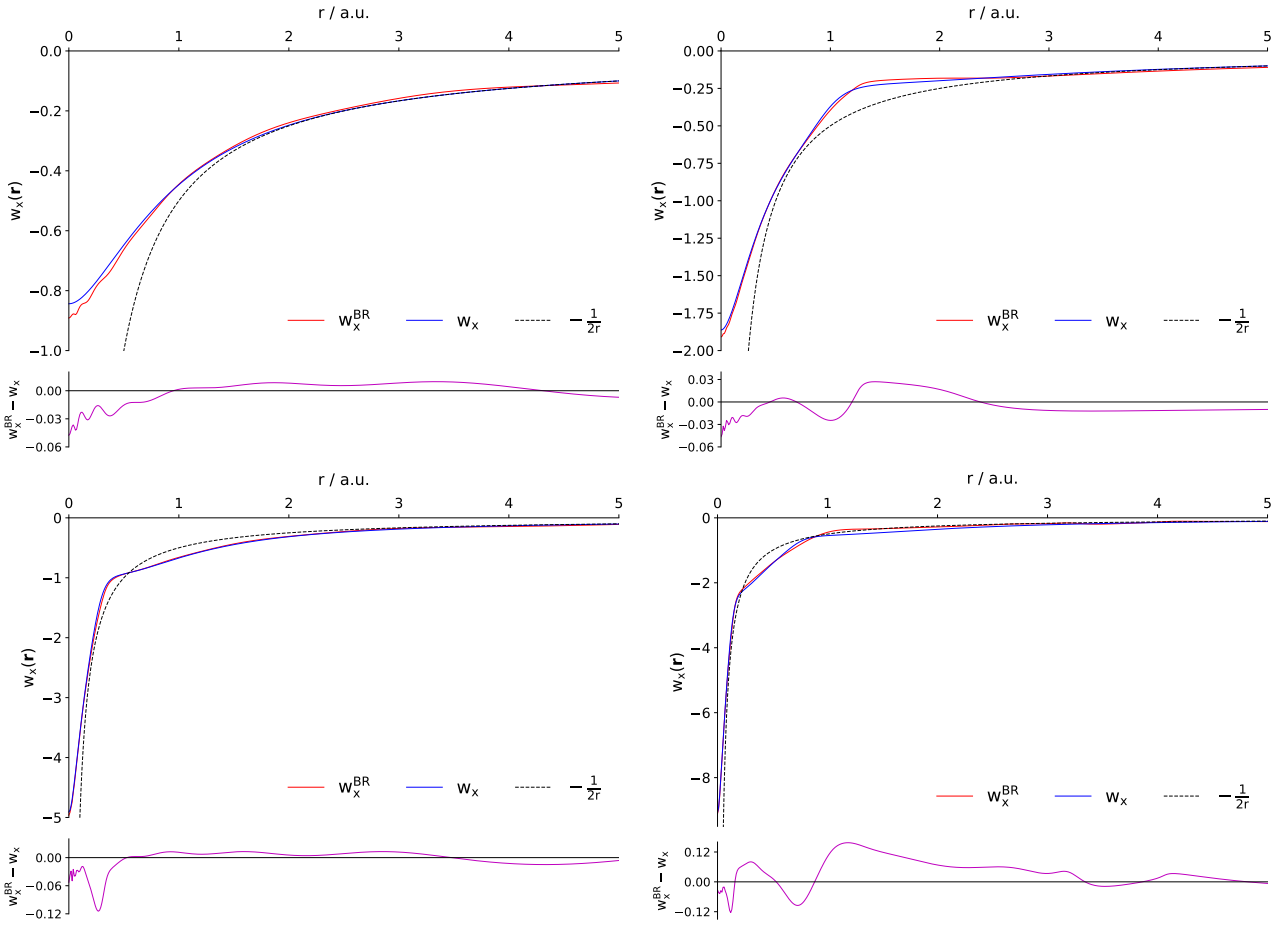

Figure 5.3: The exchange energy density, $w_x(\mathbf{r})$ as a function of the radial distance r from the nucleus (blue). Plotted for Helium (upper left), Beryllium (upper right), Neon (lower left) and Argon (lower right). For comparison the Becke–Roussel exchange energy density is also presented, $w_x^{BR}(\mathbf{r})$ (red), along with the function $-1/2r$ (dashed) that describes the asymptotic energy density. The lower insets in each panel plot the difference $w_x^{BR}(\mathbf{r}) - w_x(\mathbf{r})$.

In the individual cases, the exchange energy density for He shows a smooth monotonic decay with increasing distance from the nucleus, tending towards the asymptotic limit. Those for Be, Ne and Ar exhibit more varied structures, with features in regions typically associated the atomic shell structures. However, it important to consider that the magnitude of the exchange-energy density at each point in the system exceeds that of the correlation energy density by often several orders of magnitude. Therefore it is essential that DFAs model this component accurately independent of the correlation component if error cancellations are to be avoided.

The vast majority of DFAs do not attempt to model the exchange energy density in the gauge of the exchange hole. Instead typical functionals include GGA contributions in which Laplacian terms have been eliminated using integration by parts; their energy densities are thus transformed into a different gauge. As a result they cannot be easily compared with the accurate energy densities of the present work, for further discussion of this point see for example Ref. 133. An alternative approach is taken in the construction of the Becke–Roussel (BR) model for the exchange energy, in which it is instead modelled by the Taylor expansion of the spherically averaged exchange hole to second-order[142],

$$h_{x,\sigma}(\mathbf{r}, s) = -\rho_\sigma(\mathbf{r}) - Q_\sigma(\mathbf{r})s^2 + \ldots, \tag{5.36}$$

which refers to a shell of radius s, centred on reference point $\mathbf{r}$. This exact form is used to parametrise a model exchange hole based on the spherical hydrogenic orbital. The exchange hole curvature for spin σ is given by

$$Q_\sigma(\mathbf{r}) = \frac{1}{6}\left[\nabla^2\rho_\sigma(\mathbf{r}) - 4\tau_\sigma(\mathbf{r}) - \frac{1}{2}\frac{|\nabla\rho_\sigma(\mathbf{r})|^2}{\rho_\sigma(\mathbf{r})}\right], \tag{5.37}$$

where τ_σ is the everywhere positive–definite local kinetic energy density

$$\tau_\sigma(\mathbf{r}) = \frac{1}{2}\sum_i^{\text{occ}} |\nabla\varphi_i^\sigma(\mathbf{r})|^2 . \tag{5.38}$$

The BR exchange hole model takes the form

$$h^{\mathrm{BR}}_{\mathrm{x},\sigma}(\mathbf{r},s) = \frac{a}{16\pi bs}\left[(a|b-s|+1)\exp(-a|b-s|) - (a|b+s|+1)\exp(-a|b+s|)\right], \tag{5.39}$$

where the parameters a and b are determined by equating the values of (5.36) & (5.39) at $s=0$ and equating the curvature $Q_\sigma(\mathbf{r})$ with that of the model in (5.36). These parameters can be obtained by solving the 1D non–linear equation for $x_\sigma = a_\sigma b_\sigma$,

$$\frac{x_\sigma \exp\left(-\frac{2x_\sigma}{3}\right)}{x_\sigma - 2} = \frac{\frac{2}{3}\pi^{2/3}\rho_\sigma^{5/3}(\mathbf{r})}{Q_\sigma(\mathbf{r})} \tag{5.40}$$

at each grid point. Setting

$$b_\sigma^3 = \frac{x_\sigma^3 \exp(-x_\sigma)}{8\pi\rho_\sigma(\mathbf{r})}, \tag{5.41}$$

the exchange energy density at a given reference point $\mathbf{r}$ is then given by

$$w^{\mathrm{BR}}_{\mathrm{x},\sigma}(\mathbf{r}) = -\frac{1}{2b_\sigma}\left[1 - \exp(-x_\sigma) - \frac{1}{2}x_\sigma\exp(-x_\sigma)\right] \tag{5.42}$$

and the total exchange energy is given by

$$E^{\mathrm{BR}}_{\mathrm{x}} = \sum_\sigma \int w^{\mathrm{BR}}_{\mathrm{x},\sigma}(\mathbf{r})\rho_\sigma(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{5.43}$$

For closed shell systems, as considered in these calculations, $w^{\mathrm{BR}}_{\mathrm{x}}(\mathbf{r}) = w^{\mathrm{BR}}_{\mathrm{x},\alpha}(\mathbf{r}) = w^{\mathrm{BR}}_{\mathrm{x},\beta}(\mathbf{r})$.

The exchange energy densities $w_{\mathrm{x}}(\mathbf{r})$ and $w^{\mathrm{BR}}_{\mathrm{x}}(\mathbf{r})$, evaluated on the CCSD(T) electron density, are shown in Figure 5.3 for the atoms He, Be, Ne and Ar. On the scale of $w_{\mathrm{x}}(\mathbf{r})$ the BR exchange energy density is reasonably accurate. By construction the model has the correct $-1/2|\mathbf{r}|$ asymptotic behaviour (shown by the dashed lines) in the energy density and is close to $w_{\mathrm{x}}(\mathbf{r})$ in most regions of space. The difference $w^{\mathrm{BR}}_{\mathrm{x}}(\mathbf{r}) - w_{\mathrm{x}}(\mathbf{r})$ is shown in more detail below the plots for each system. In general the BR model gives an overly–negative energy density near to the nucleus in these systems. Some oscillatory behaviour is also present in this region, arising from the Gaussian basis functions used in this work. For the systems Be, Ne and Ar the largest differences between the BR model and and the accurate energy density occur in regions where the energy density is most rapidly changing, such as features close to inter–shell transitions.

To quantify the quality of the BR model we consider the exchange energies obtained by integration with the CCSD(T) electronic density in Table 5.1. The errors in the exchange energy δE_{x} are also presented. In general the BR model is accurate to within 1–2% for E_{x}. For the He, Be and Ne atoms the overly negative behaviour of $w^{\mathrm{BR}}_{\mathrm{x}}(\mathbf{r})$ close to the nucleus dominates and leads to erroneously negative values of E_{x}. For Ar the behaviour is different, with positive errors in the valence region close to the inter-shell features in the electron density dominating to give a value for E_{x} that is too positive overall.

Quantity	He	Be	Ne	Ar
E_{x}	−1.0241238	−2.6677743	−12.0764264	−30.1732130
$E^{\mathrm{BR}}_{\mathrm{x}}$	−1.0381828	−2.6826734	−12.1702073	−30.0970295
δE_{x}	0.0140590	0.0148991	0.0937809	−0.0761835

Table 5.1: The exchange energies in Hartree obtained from the accurate energy densities computed in this work, E_{x}, and the BR model, $E^{\mathrm{BR}}_{\mathrm{x}}$. All values are evaluated on the CCSD(T) electron density.

5.3.3 Correlation Energy Densities

The λ–dependent correlation energy densities $w_{\mathrm{c},\lambda}(\mathbf{r})$ and the λ–averaged correlation energy density $\bar{w}_{\mathrm{c}}(\mathbf{r})$ are plotted for each of He, Be, Ne and Ar in Figures 5.4, 5.5, 5.6 & 5.7 respectively.

The λ–averaged correlation energy density is computed by numerical integration over λ. It is clear that whilst the magnitude of the correlation energy density changes with λ, its structure is not sensitive to the interaction strength. The structure of the correlation energy densities is clearly more complex than that of the exchange energy density and $\bar{w}_{\mathrm{c}}(\mathbf{r})$ is the target for DFAs to model. In Section 5.5 we discuss to what extent these features may be recognised using

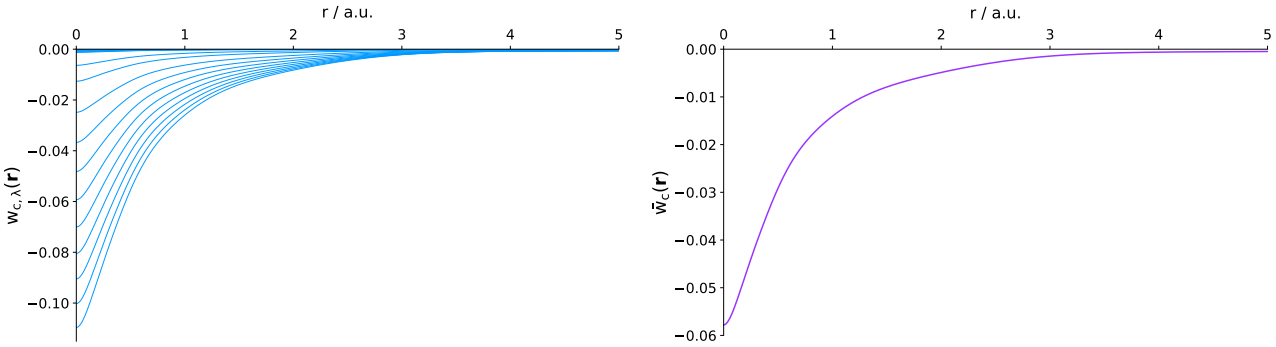

Figure 5.4: The correlation energy density at each value of the coupling–constant, λ, (left) and the λ–averaged correlation energy density (right) for the He atom.

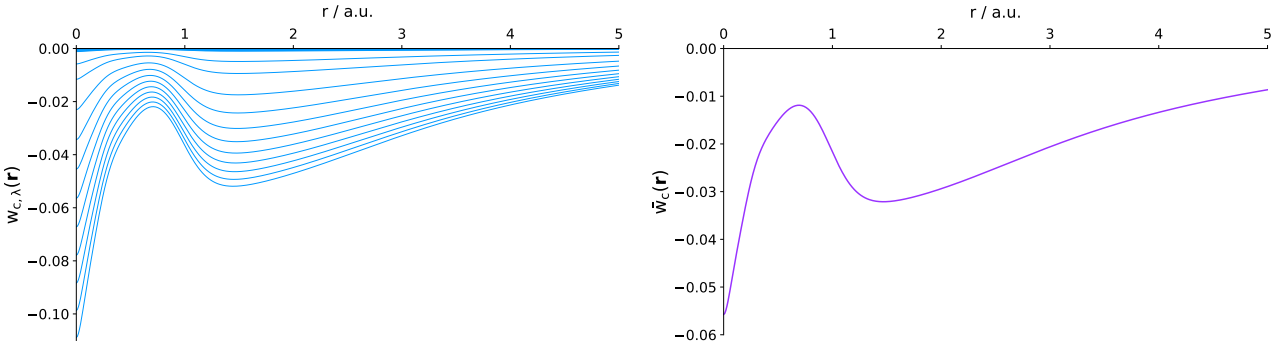

Figure 5.5: The correlation energy density at each value of the coupling–constant, λ, (left) and the λ–averaged correlation energy density (right) for the Be atom.

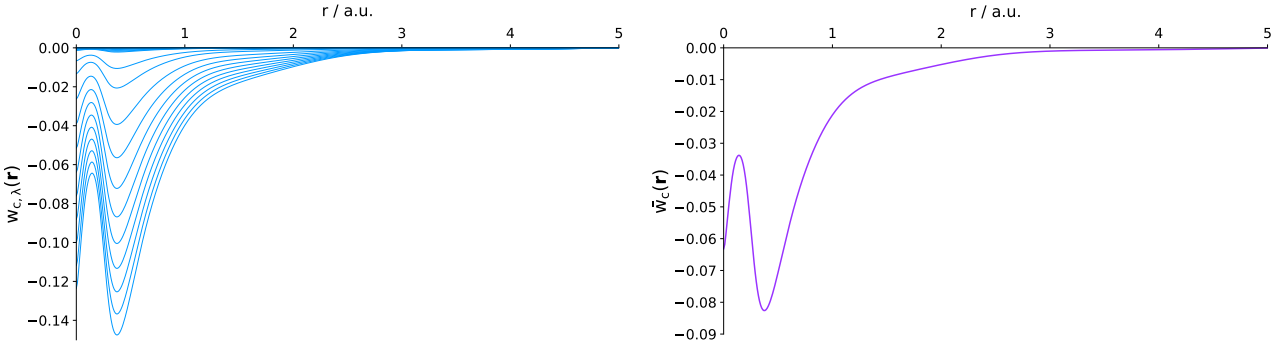

Figure 5.6: The correlation energy density at each value of the coupling–constant, λ, (left) and the λ–averaged correlation energy density (right) for the Ne atom.

quantities typically used in simple local DFAs. The variation of $w_{c,\lambda}(\mathbf{r})$ with λ defines a local adiabatic connection at each point in space and this is discussed in subsection 5.3.4.

The Kohn–Sham energy components corresponding to the CCSD(T) electronic density are given in Table 5.2. The non-interacting kinetic (T_s), Coulomb (E_J) exchange (E_x) and electron-nuclear (E_{ne}), components are computed by performing the Lieb optimisation at $\lambda = 0$ and using the resulting molecular orbitals. Subtracting these from the total CCSD(T) energy (the value of $\mathcal{E}_1(v)$) gives the value of the Kohn–Sham correlation energy E_c. To confirm the accuracy of the calculated λ–averaged correlation energy densities, we have also evaluated the correlation energy by numerical integration of this quantity with the electronic density, denoted in the table as $E_c^{w_c}$. Both estimates of E_c agree to within the accuracy expected, confirming the quality of the calculated λ–averaged correlation energy densities.

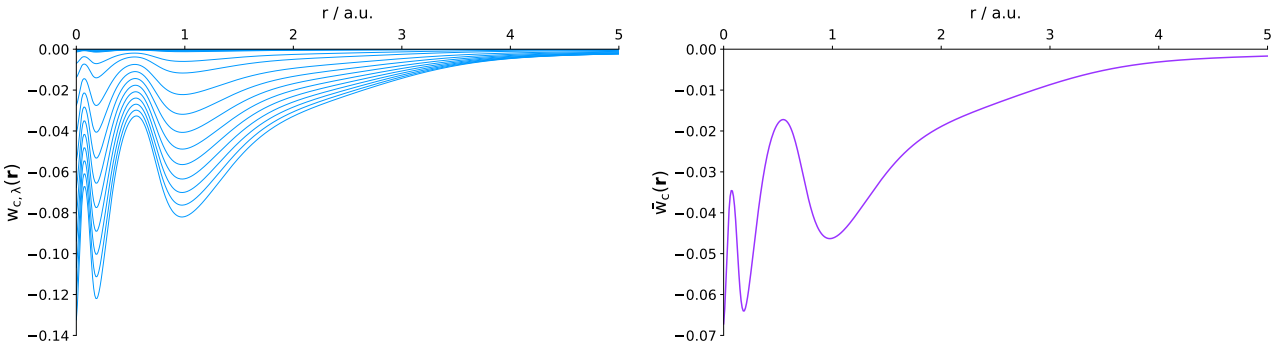

Figure 5.7: The correlation energy density at each value of the coupling–constant, λ, (left) and the λ–averaged correlation energy density (right) for the Ar atom.

System	$\mathcal{E}^{\mathrm{CCSD(T)}}$	T_s	E_J	E_x	E_{ne}	E_c^{ref}	$E_c^{w_c}$
Helium	−2.9027040	2.8649869	2.0481687	−1.0240843	−6.7505261	−0.0412492	−0.0412490
Beryllium	−14.6657007	14.5924154	7.2171170	−2.6735568	−33.7071485	−0.0945277	−0.0945267
Neon	−128.9180171	128.5876144	65.9928278	−12.0758565	−311.0457791	−0.3768238	−0.3768261
Argon	−527.4275953	526.8149932	231.6859657	−30.1683190	−1255.1400029	−0.6202324	−0.6202741

Table 5.2: Kohn–Sham electronic energy contributions in Hartree for He, Be, Ne and Ar. The correlation energy is computed by differences between the $\lambda = 0$ and $\lambda = 1$ energies and by spatial integration of the λ–averaged correlation energy density.

5.3.4 The Local Adiabatic Connection

The challenge for practical DFAs is to model the λ–averaged energy density. As is discussed in Section 5.5, it may be possible to identify some correspondence between features in the structure local observables and in that of the energy density. However, to gain further insight into the nature of the correlation at each point in space, it is necessary consider the *local* AC. As representative examples we consider the Ne atom, where the correlation is expected to be mainly dynamic in nature, and Be where static correlation effects play a role.

For the *global* integrand $\mathcal{W}_\lambda$ defined in (3.177), it is understood that, as the nature of the correlation shifts from predominantly dynamic to more static in character, the curvature of this integrand becomes more and more pronounced. We now consider the behaviour of the local integrands in these systems.

In Figures 5.8 & 5.9, we present $\bar{w}_c(\mathbf{r})$ and the local AC at several points in space for Be and Ne respectively. These points are marked on the $\bar{w}_c(\mathbf{r})$ plots by a coloured symbol, with the corresponding local AC curve plotted with the same marker in the right hand panel.

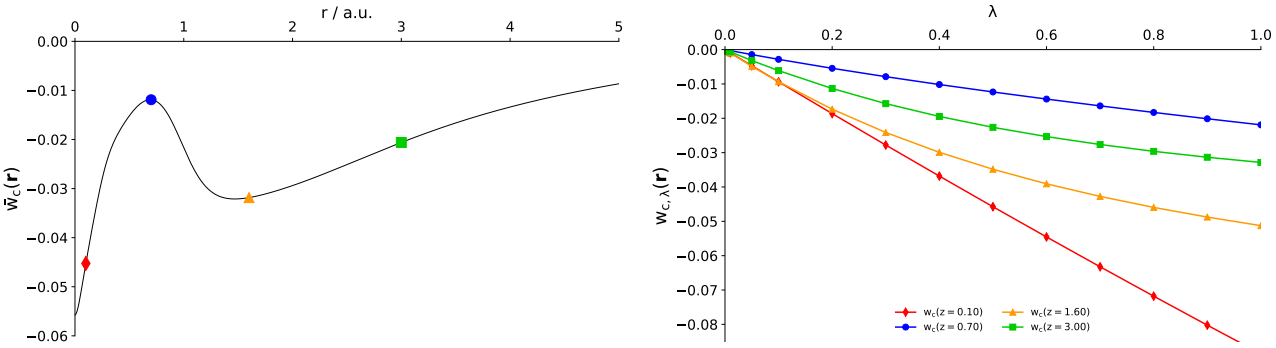

Figure 5.8: The local AC at $r = 0.1, 0.7, 1.6$ and 3.0 a.u. in Beryllium atom. The positions considered are shown by markers on the λ–averaged energy density, $\bar{w}_c(\mathbf{r})$, in the left panel. The corresponding markers and colours are used for the local ACs as a function of interaction strength λ in the right hand panel.

The first feature observed for both species is that the corresponding local ACs do not vary in magnitude (i.e. from bottom to top of the right hand panels of Figures 5.8 & 5.9) sequentially with variation in r. This is because of the oscillatory nature of $\bar{w}_c(\mathbf{r})$ and the fact that the area between each local AC curve and the x–axis in the right–hand panel represents the value of $\bar{w}_c(\mathbf{r})$ in the left hand panel.

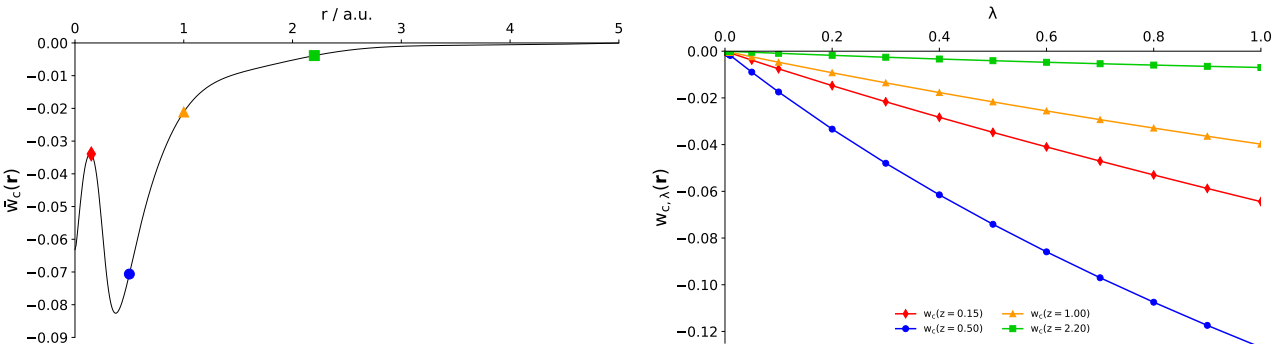

Figure 5.9: The local AC at $r = 0.15, 0.50, 1.00$ and 2.20 a.u. in Neon atom. The positions considered are shown by markers on the λ–averaged energy density, $\bar{w}_c(\mathbf{r})$, in the left panel. The corresponding markers and colours are used for the local ACs as a function of interaction strength λ in the right hand panel.

More interestingly, we see that the local shape of the integrands varies in an intuitive way with r. For both systems the local AC in the core region is strongly linear in λ, indicating that the nature of the correlation is relatively simple. For the Ne atom, the local AC is reasonably linear for all r values, indicating that the correlation is relatively simple throughout the system. As a result, the main challenge for a DFA is then to capture the structure of $\bar{w}_c(\mathbf{r})$ as discussed in the previous section.

For the Be atom, the behaviour of $\bar{w}_c(\mathbf{r})$ is more complex; the first two points in the core and inner valence regions show simple, relatively linear local ACs. However, the remaining two points in the valence and outer valence regions show much more pronounced curvature, indicating a more complex higher order dependence on λ, in line with the expected global behaviour for a system exhibiting static correlation effects. However, in the present example, it can be seen that these more complex effects are essentially localised to the valence and outer valence regions.

This connection between the nature of the correlation and the location in the system should be of significant utility in the construction and testing of DFAs. In particular it suggests the possibility of a topological DFT, in which DFA energy densities are constructed by taking into account both the local λ–variation and the spatial position in the system considered.

As indicated in the schematic of Figure 5.1, one could also consider the initial slope of the local AC as a measure of how rapidly the local model must depart from exchange–only behaviour at each point in space. In Figure 5.10 we present this derivative for each of the atomic systems considered.

In general, the shape of these functions closely resembles the energy density itself. This arises because a larger $\bar{w}_c$ is obtained when the area enclosed by w_c is larger, generally involving a more negative initial–slope. For a more insightful understanding of the nature of the correlation using derivatives at $\lambda = 0$, it may then be necessary to proceed to higher orders. This data should be of particular interest for constructing DFAs via interpolation models based on the local AC and is discussed further in Chapter 6.

5.4 The Adiabatic Connection In H_2

The simplest of the small molecules with well–known properties is the Hydrogen molecule – a diatomic with only two electrons and one occupied molecular orbital. However, the dissociating H_2 molecule is a characteristic example of the failure of standard approximations in Kohn–Sham DFT to properly describe *nondynamical correlation*, described briefly in Chapter 2.

5.4.1 The Global Adiabatic Connection Of H_2

There has been much interest in the *global* AC of H_2, i.e. as given by (3.180), because it is thought that changes in the AC with increasing bond length may provide further insight into the nature of nondynamical correlation. H_2 is a particularly interesting case because it is known that the dissociated atoms must each have zero correlation energy, as they only have one electron, so any correlation in the system overall must be the result of nondynamical correlation (with negligible dynamical correlation due to the large separation of the electron densities).

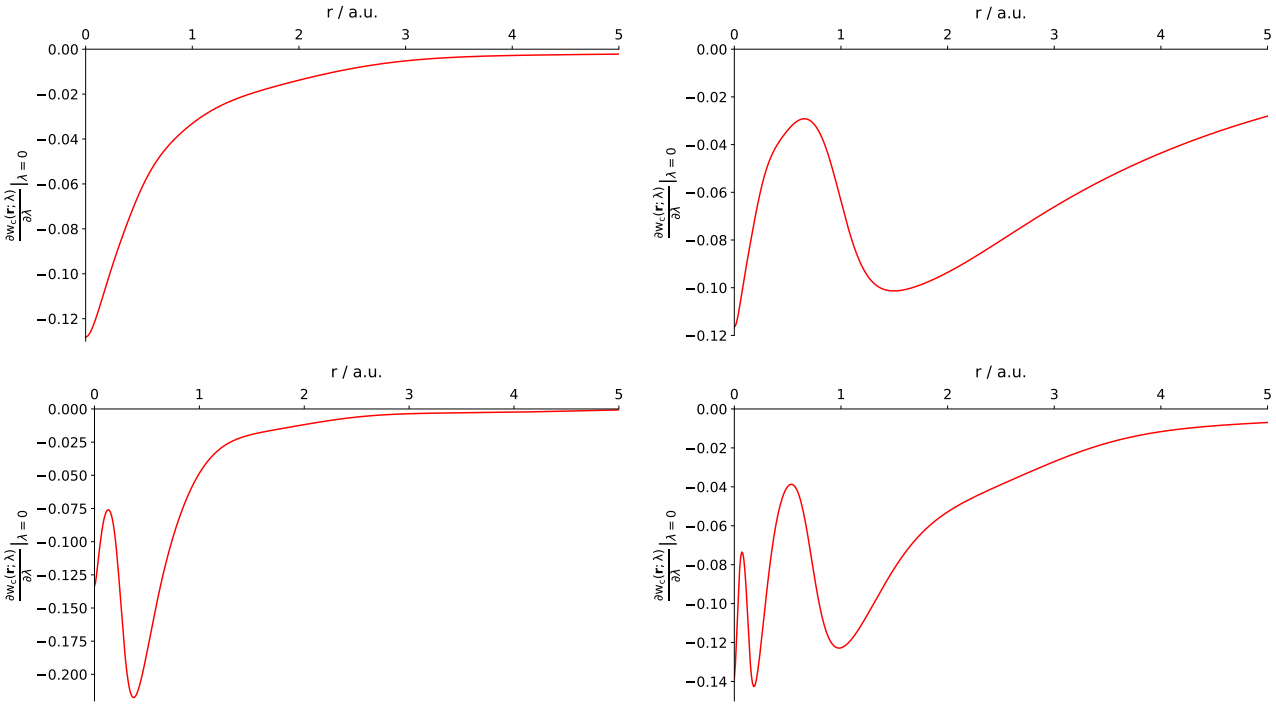

Figure 5.10: The gradient of the correlation energy density with respect to variation in electron-interaction strength, evaluated at the non-interacting limit $\lambda = 0$, for Helium (upper left), Beryllium (upper right), Neon (lower left) and Argon (lower right).

Given that the H_2 molecule close to its equilibrium geometry is well–represented by a single determinant wavefunction, it follows that the overlapping atomic orbital model is a fair one and that there is significant overlap of the electron densities of the two atoms, hence dynamical correlation will be present. It could therefore be said that changes in the AC seen as H_2 transitions from its equilibrium geometry to a dissociated state is an observation of how the behaviour of the AC differs between systems in which the correlation is (almost) entirely dynamical in nature to those in which it is exclusively nondynamical.

In previous studies of the global AC of H_2, notably those by Peach *et. al.*[154] and by Teale *et. al.*[111], it was observed that as the bond length of H_2 was increased from its equilibrium value to a much larger value representing dissociation, the shape of the global AC gradually changed from being approximately linear in λ to having much higher curvature and much greater initial slope, to the extent that for the largest bond–lengths studied, the AC simply appeared as an 'L'–shaped curve, with $\mathcal{W}_{c,\lambda}$ reaching its final value for λ infinitesimally greater than zero. This trend can be seen clearly in Figure 4 of Ref. 111.

5.4.2 Correlation Energy Densities In H_2

In Figures 5.11 & 5.12 below, the correlation energy densities along the principle axis of the H_2 molecule at four geometries are plotted for a range of λ–values between 0 and 1, with the λ–averaged correlation energy density given alongside. The methods used to calculate this data is the same as that described in subsection 5.3.1, however for these calculations the wavefunction method used was that of CCSD[44] (equivalent to FCI for a two–electron system) and the basis set used for expansion of the orbitals and of the potential was the *u*–aug–cc-pCVTZ basis set of Dunning[116].

Many of the observations that may be made of these plots can also be made of the plots pertaining to atoms in Section 5.3. The magnitude of correlation energy densities is always greatest in the locality of the nuclei, and decays with increasing distance from the nuclei. The magnitude of the correlation energy increases with increasing λ but its spatial distribution remains the same, and is reflected in the λ–averaged correlation energy density.

However, new and interesting observations may be made if one compares the equivalent plots for the four different bond–lengths presented here. It appears that, as the bond length is increased, the correlation energy density at the nuclei increases in magnitude and that in the internuclear region rapidly approaches zero, actually reaching zero in the midbond region for the 5.0 a.u. bond length.

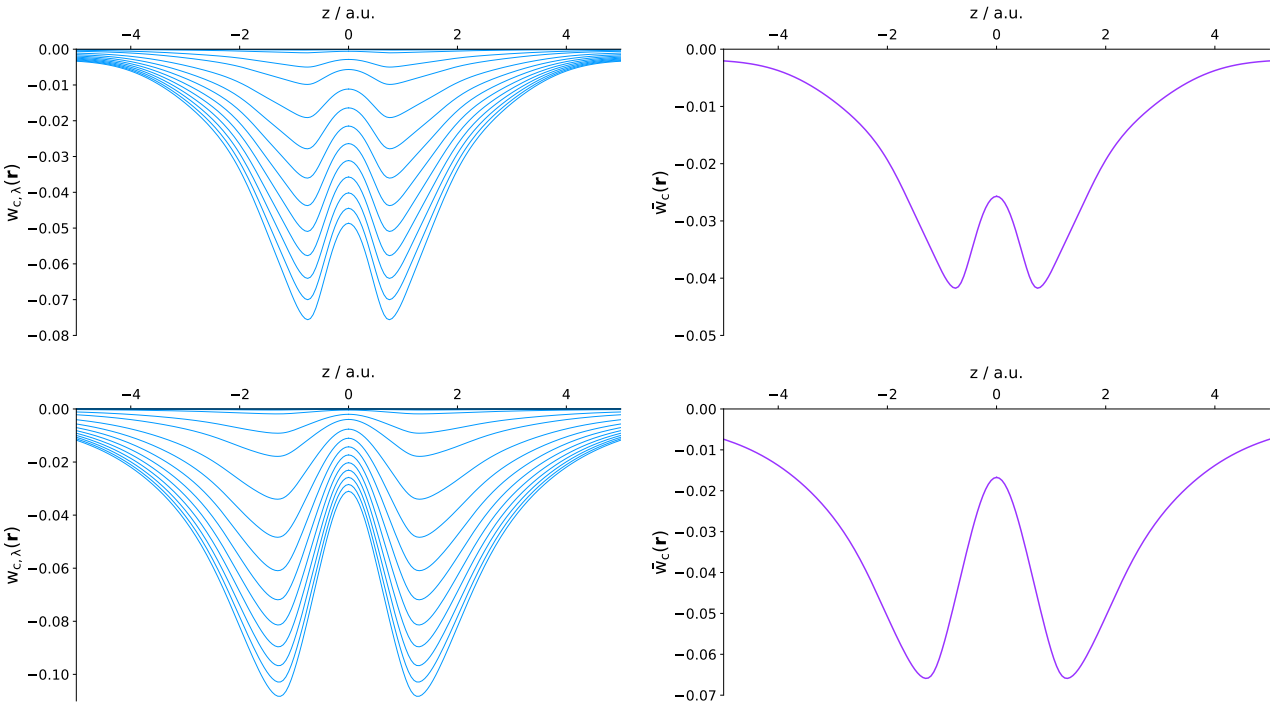

Figure 5.11: The correlation energy density at each value of λ (left) and the λ–averaged correlation energy density (right) along the principle axis of the Hydrogen molecule, with bond length 1.4 a.u. (top) and 2.4 a.u (bottom).

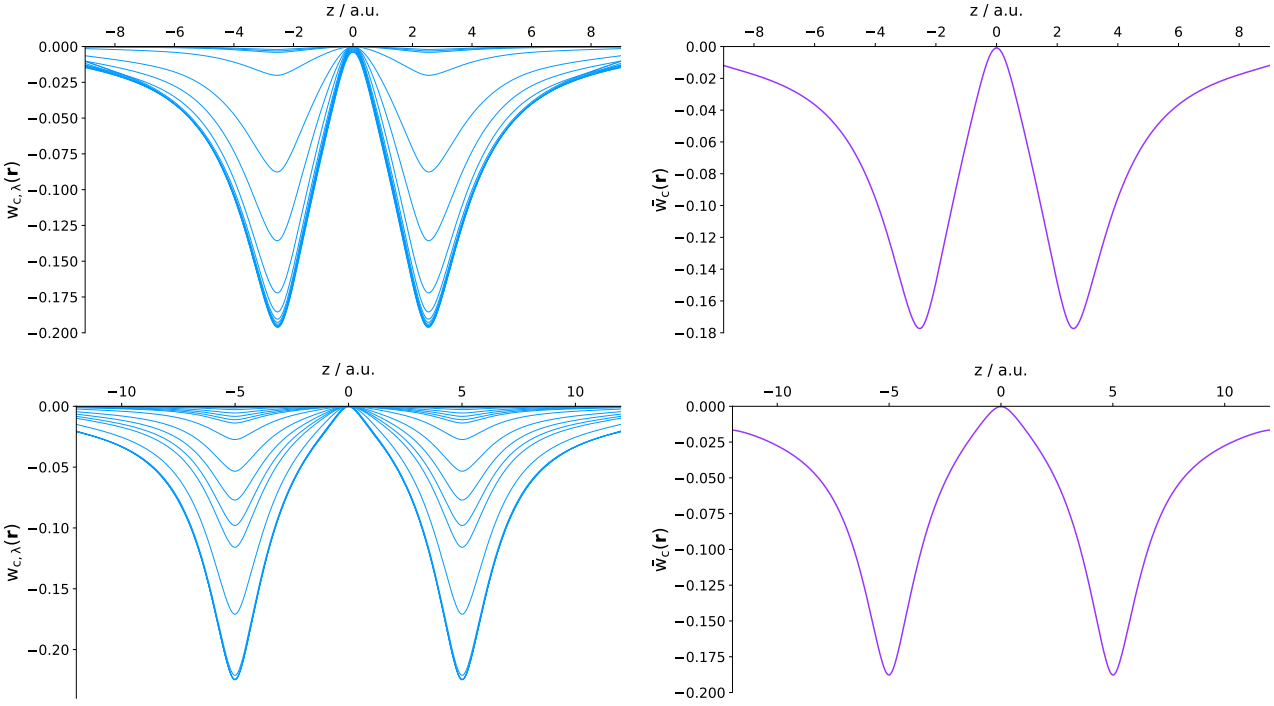

Figure 5.12: The correlation energy density at each value of λ (left) and the λ–averaged correlation energy density (right) along the principle axis of the Hydrogen molecule, with bond length 5.0 a.u. (top) and 10.0 a.u. (bottom).

This may be in part rationalised in terms of charge density; with increasing internuclear distance, there is a rapid decay of charge density in the bonding region hence dynamical correlation effects here (i.e. *purely local* correlation) will become vanishingly small. However, the greatly increasing correlation energy density around each nucleus is much more difficult to explain as a feature of charge density - approaching dissociation, the charge density around each nucleus will become exclusively that of a single electron and there will be no *local* correlation present in the

system.

It follows therefore that the increasing correlation energy density at the nuclei approaching dissociation is by virtue of growing nondynamical correlation effects, that are *non–local* in nature. This type of system is known to exhibit the 'L'–shaped global AC curves referred to above, and indeed it can be seen from the left–hand plots that the correlation energy density converges to the $\lambda = 1$ rapidly and within very few intervals of λ measured for the 5.0 a.u. bond length, whereas for the equilibrium bond length of 1.4 a.u. the magnitude of correlation energy density increases much more smoothly in the range $\lambda = 0 \rightarrow 1$. It appears from the spacing of the $w_{c,\lambda}$ lines that the rate of change is not uniform across all parts of the molecule, and so it is useful to examine the local AC at a sample of points along this plot.

5.4.3 Local Adiabatic Connections In H_2

In Figures 5.13, 5.14, 5.15 & 5.16, the λ–averaged correlation energy density of H_2 in each of the four geometries is reproduced, and alongside each are plots of the local AC at four points as marked on the respective energy density.

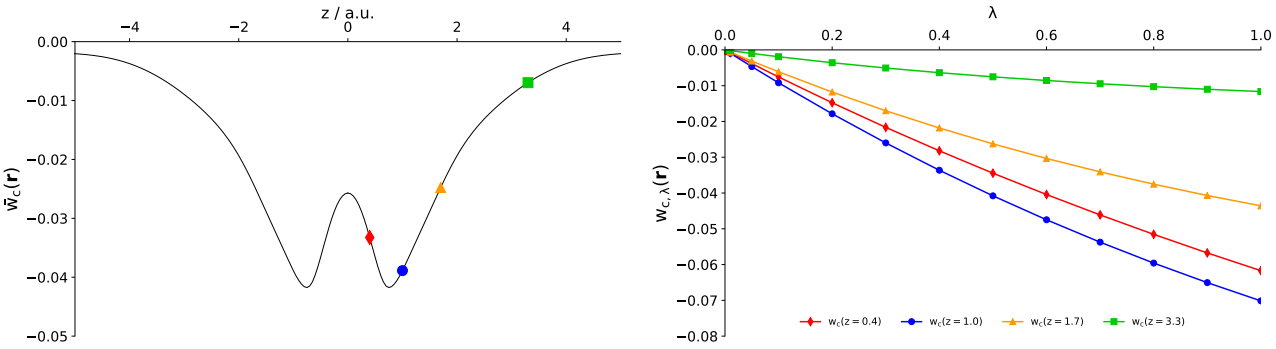

Figure 5.13: The local AC at points along the principle axis of H_2 with bond length 1.4 a.u.; points marked on the left–hand plot correspond to the AC curves on the right–hand plot.

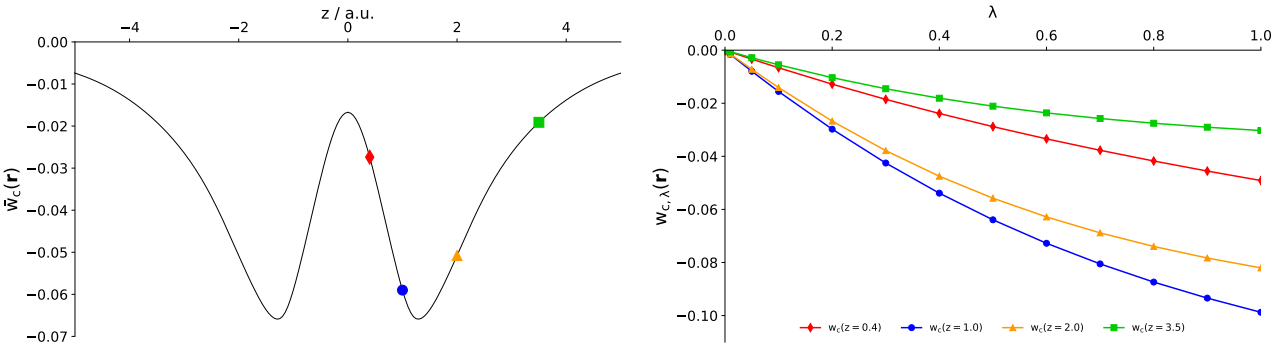

Figure 5.14: The local AC at points along the principle axis of H_2 with bond length 2.4 a.u.; points marked on the left–hand plot correspond to the AC curves on the right–hand plot.

A pattern similar to that shown previously for atoms can be seen in the local AC curves where the H–H bond length is 1.4 a.u in Figure 5.13 ($R_{HH} = 1.4$). The AC is predominantly linear close to the nuclei, as this is a high–density region with mainly dynamical correlation effects, which typically exhibit a linear AC[111]. There is greater curvature in the local AC at a lower density region further away from the H–H bond, and in the very low density regions at the edges of the molecule the local AC is very shallow and roughly linear. However, it also appears that the linearity of the AC remains in the internuclear region, an observation that can be attributed to the relatively high electron density along the bond for $R_{HH} = 1.4$.

In the plots for $R_{HH} = 2.4$ in Figure 5.14, there can be seen several differences in the local ACs compared with those of the equilibrium geometry. The local AC in the internuclear region remains near–linear but is much shallower, reflecting the overall reduction in charge density along the H–H bond. However, the local ACs of the point very close to a nucleus and the point slightly further away from the H–H bond exhibit a much more pronounced curvature than in the equilibrium geometry, appearing to have significant quadratic variation with λ.

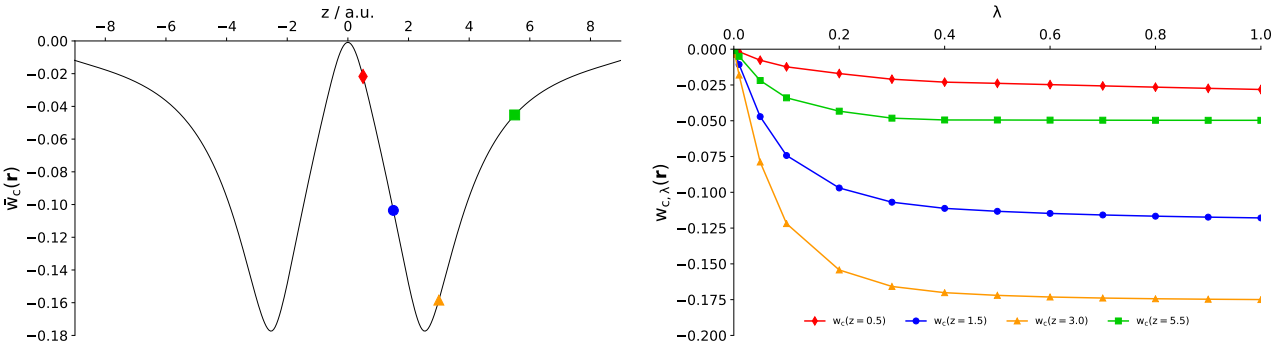

Figure 5.15: The local AC at points along the principle axis of H_2 with bond length 5.0 a.u.; points marked on the left–hand plot correspond to the AC curves on the right–hand plot.

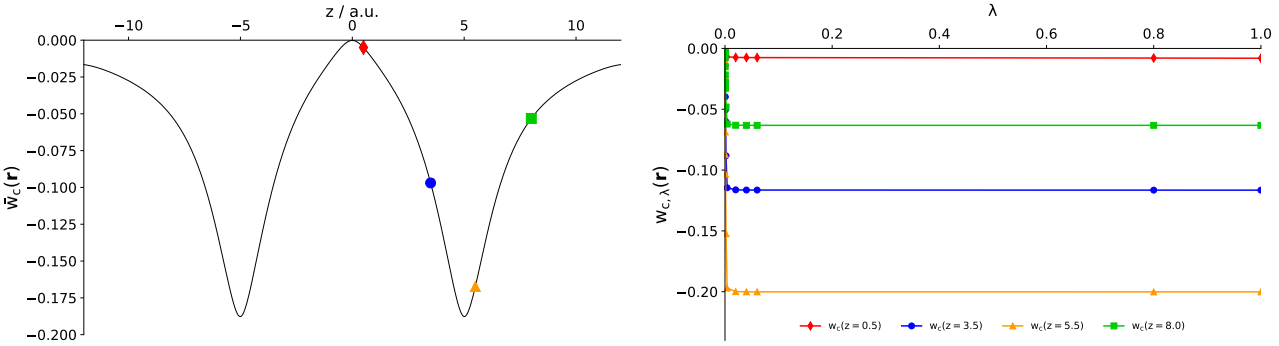

Figure 5.16: The local AC at points along the principle axis of H_2 with bond length 10.0 a.u.; points marked on the left–hand plot correspond to the AC curves on the right–hand plot.

The local AC curves for $R_{HH} = 5.0$ in Figure 5.15 are clearly approaching the form taken by the global AC curves for highly stretched H_2 in Ref. 111. The local AC in the midbond region is relatively shallow and of low curvature, a feature of the very low electron density of the internuclear region at this geometry. However, approaching the nuclei the initial slope of the local AC increases in magnitude, to the extent that it is converging on its $\lambda = 1$ value at $\lambda \leqslant \frac{1}{2}$.

This trend in the limit of dissociation can be observed for $R_{HH} = 10.0$ in Figure 5.16, for which the local AC mirrors the global AC in that it reaches its $\lambda = 1$ value for infinitesimally small, but non–zero, value of λ, producing the characteristic 'L'–shape AC curve. Interestingly, the onset of nondynamical correlation appears to occur throughout the system as the internuclear bond is extended, indicating that the Kohn–Sham wavefunction is a poor model of the interacting wavefunction at all points in space at this geometry.

It is therefore unsurprising that present–day *xc*–functionals completely fail to model this phenomenon, as they model the *local* energy density / potential as a function of *local* quantities only. One approach to the problem of nondynamical correlation comes from considering the case for $\lambda \to \infty$, a model known as the *Strictly Correlated Electron* (SCE) model – described in subsection 6.3.2. The SCE model has recently gained interest, particularly through the work of Seidl *et. al.*[155,156], as it is exactly soluble for a few small model systems. If the AC has the same shape as that of highly–stretched H_2, it follows that the value of the integrand at $\lambda \to \infty$ is equal to that at $\lambda = 1$, and so an accurate density functional may be realised by interpolating the AC between the non–interacting and strictly–interacting systems, see for example Refs. 157,158. Recent work by Vuckovic *et. al.* has shown that the SCE model is particularly successful in modelling the dissociating hydrogen molecule[159].

Some global interpolation models also include the expression for correlation energy derived by Görling and Levy, in which the correlation energy to second–order perturbation theory is equal to half the slope of the global AC at $\lambda = 0$[160]. Interpolation models for the AC have been tested over several years and, although for some applications they greatly outperform standard DFAs, they are fundamentally flawed because they are not size–consistent. These interpolation models are examined in greater detail in Chapter 6; a comprehensive review of the literature on this type of approach can be found in Ref. 145.

The energy densities and local ACs shown Figures 5.13, 5.14, 5.15 & 5.16 suggest that one possible approach to the nondynamical correlation problem could be through local interpolation of the AC, i.e. attempting to reconstruct

the λ–averaged energy density at a given point by interpolating the local AC. This approach is examined further in Chapter 6.

5.5 Energy Densities & The Laplacian Of The Density

Given that the correlation energy densities show a much more complex structure than the corresponding exchange-only energy densities, it is interesting to examine local quantities such as the density and its derivatives in comparison with $\bar{w}_c$ to explore how it may be reconstructed in the form of a simple DFA. In the present work, we consider ρ, $\nabla\rho$, τ and $\nabla^2\rho$ and seek to find correspondences between features in these local quantities and those in $\bar{w}_c(\mathbf{r})$.

We find that, whilst the information contained in the density and its gradient may be sufficient to determine the main features of $\bar{w}_x$ on a local scale, these alone do not provide sufficient information for an accurate reconstruction of the more detailed structure of $\bar{w}_c$. For the kinetic energy density τ, a number of forms may be considered; the most commonly employed forms are those based on the Laplacian of the orbitals and the (everywhere positive) form based on the gradient of the orbitals. Neither of these forms were found to have well–defined features that align well with those in the correlation energy density.

However, comparing $\nabla^2\rho$ with features of $\bar{w}_c(\mathbf{r})$ is much more informative, as the nodes of the Laplacian appear to delimit features of the energy density. The utility of $\nabla^2\rho$ as an indicator for features of the correlation energy density is illustrated in Figure 5.17 for He & Be, and in Figure 5.18 for Ne & Ar. Here, the λ–averaged correlation energy density and $\nabla^2\rho$ are plotted together. The vertical dashed lines indicate the positions of nodes in $\nabla^2\rho$ and can be seen to delimit the core, valence and outer regions of the atom. The correlation energy density clearly undergoes transitions between these regions. This indicates that the seemingly complex structure of the correlation energy density can be understood from a relatively simple quantity depending only on the electronic density, namely $\nabla^2\rho$.

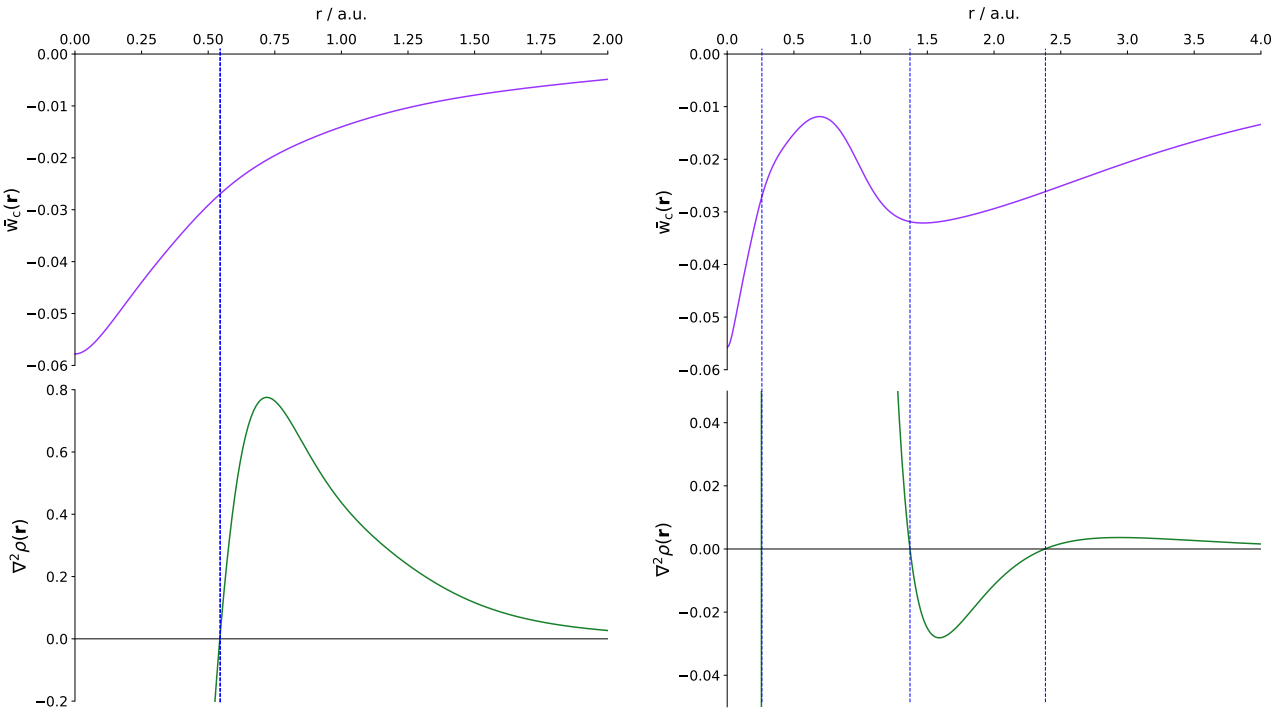

Figure 5.17: For Helium (left) and Beryllium (right), the λ–averaged correlation energy density $\bar{w}_c$ (upper panel) is compared with the Laplacian of the electron density (lower panel). The dashed lines highlight the correspondence with features of $\bar{w}_c$ and nodes of the Laplacian.

This apparent correspondence is more clearly seen in the comparison between energy density and the Laplacian in the diatomic molecules N_2 & HF, where the distribution of energy density along the bond exhibits a more complex structure than that along the radius of an isolated atom.

The correlation energy density for values of λ in the range $0 \leqslant \lambda \leqslant 1$ and the corresponding λ–averaged correlation energy density for the N_2 molecule in its equilibrium geometry is shown in Figure 5.19, calculated in the *u*–aug–cc-pCVDZ basis. Unlike previous examples, for which energy densities were plotted along one dimension, for these two molecules they were plotted on a two–dimensional Cartesian grid. This is because, as can be seen in the figures, the

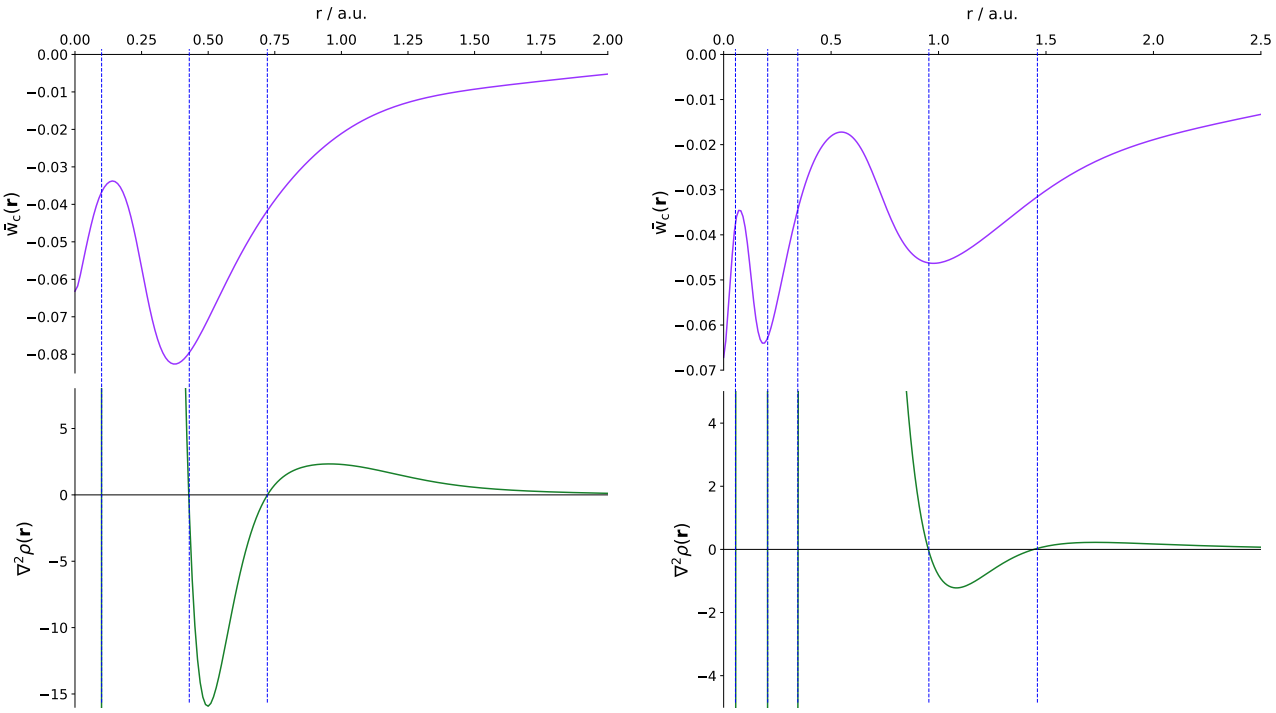

Figure 5.18: For Neon (left) and Argon (right), the λ–averaged correlation energy density $\bar{w}_c$ (upper panel) is compared with the Laplacian of the electron density (lower panel). The dashed lines highlight the correspondence with features of $\bar{w}_c$ and nodes of the Laplacian.

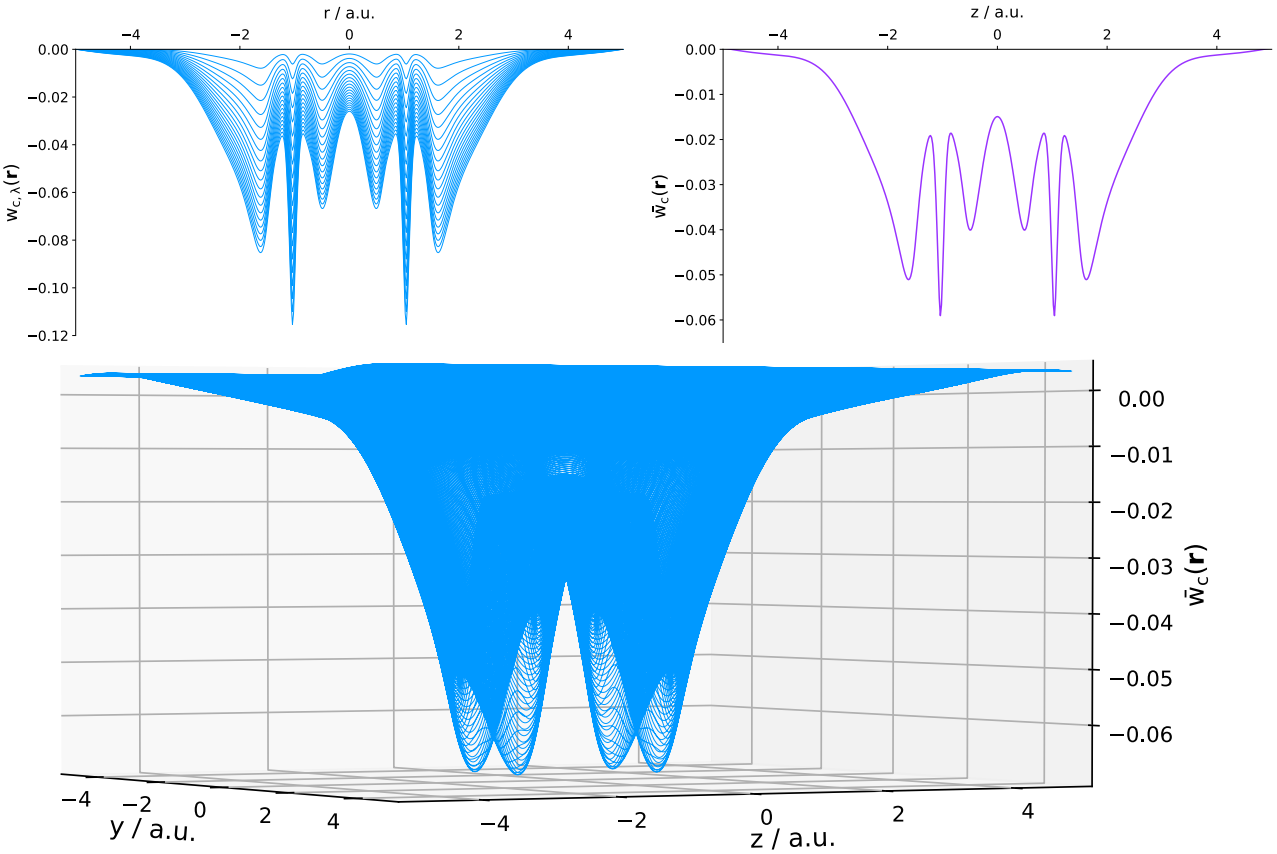

Figure 5.19: The correlation energy density at each value of λ (left) and the λ–averaged correlation energy density (right) along the principle axis of the N_2 molecule, and a 3D plot of the λ–averaged correlation energy below.

energy densities have a larger number of features and molecules have lower symmetry than atoms, so calculating the energy density over a plane allows its distribution to be examined both parallel and perpendicular to the bond.

The plots of $w_{c,\lambda}$ for $0 \leqslant \lambda \leqslant 1$ show relatively evenly spaced lines, suggesting that the local AC over this space will be low–order in λ. The λ–averaged energy density has peaks on either side of the nuclei similar to the inter–shell peaks observed in Figure 5.6 for neon, however their magnitudes are not symmetric around each of the nuclei; they appear to be distorted by the bond such that their magnitudes are lower in the internuclear region. Some sort of distortion to the atomic energy densities would be expected, as the charge density itself is distorted in a molecule compared to that in isolated atoms.

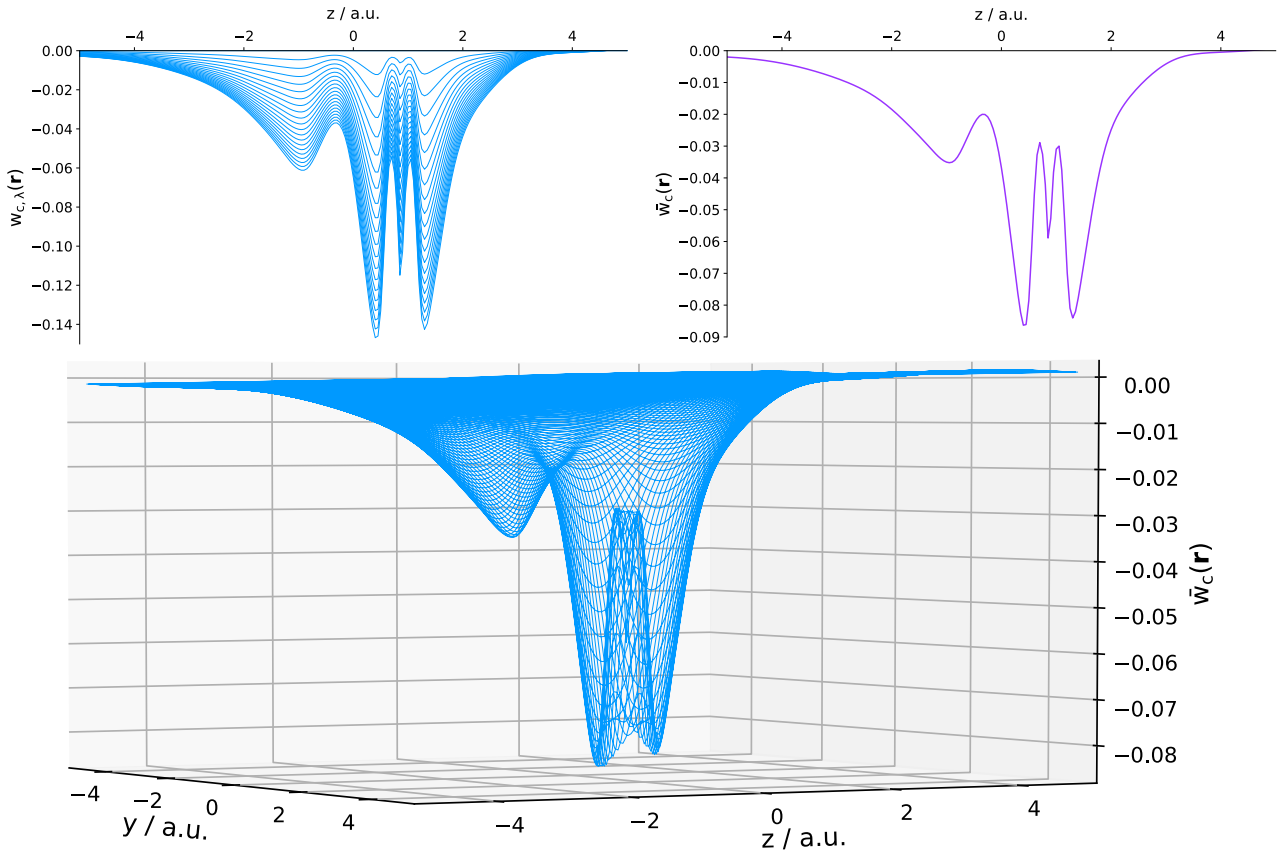

Figure 5.20: The correlation energy density at each value of λ (left) and the λ–averaged correlation energy density (right) along the principle axis of the HF molecule, and a 3D plot of the λ–averaged correlation energy below.

Similar plots for Hydrogen Fluoride (HF) are given in Figure 5.20, which unlike N_2 is heteronuclear so is not symmetric about the midbond. The plots of $w_{c,\lambda}$ for $0 \leqslant \lambda \leqslant 1$ again suggest an AC that is of low–order in λ. The λ–averaged correlation energy density exhibits a pattern that can be rationalised in a similar way to that of N_2, i.e. the energy density around the fluorine nucleus has peaks resembling inter–shell peaks, but distorted. The distortion is much smaller than in N_2, which would be expected hydrogen has a much lower nuclear charge and electronegativity than fluorine.

In Figure 5.21, the correspondence between positions of the nodes of the Laplacian and positions of turning points in the λ–averaged energy density is immediately apparent. In particular, every turning point in the λ–averaged correlation energy density of N_2 is almost exactly accompanied by a node in the Laplacian. For the heteronuclear diatomic, with a non–symmetrical charge density along the bond and a large overall dipole moment, the alignment with nodes of the Laplacian is still observed.

In the construction of most standard GGA functionals dependence on the Laplacian is eliminated by integration by parts, which also changes the gauge of the energy density – making it difficult to compare them with energy densities in the gauge of the exchange–correlation hole. Cancio[161] has suggested that the construction of functionals including the Laplacian should be revisited – noting that the slowly varying limit can still be correctly recovered when the Laplacian is included, but that its inclusion may substantially improve the description of strongly inhomogeneous systems such as atoms, molecules and materials.

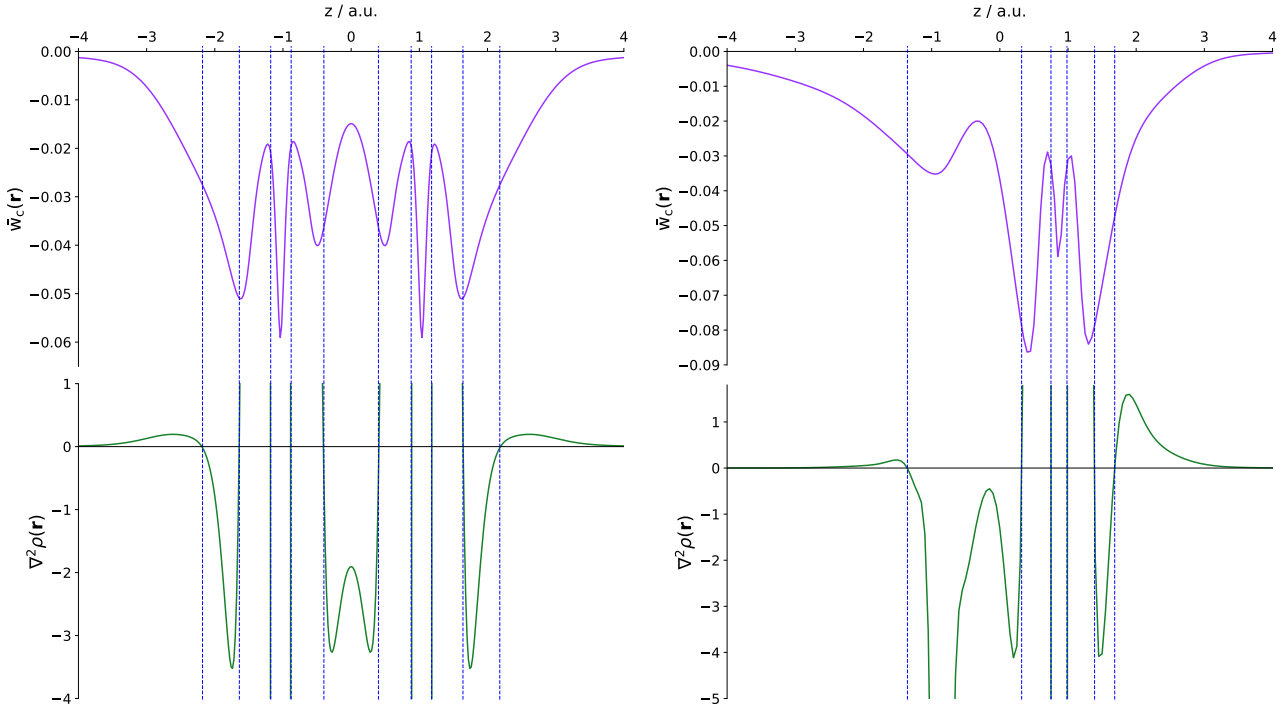

Figure 5.21: Plots showing the λ–averaged correlation energy density of N_2 (left) and HF (right), with the Laplacian of the density of the respective molecule plotted immediately below. The blue dashed lines highlight the parallel between nodes of the Laplacian and features in the λ–averaged correlation energy density.

The results presented here support this idea and suggest that in particular an accurate description of the correlation energy density may be easiest to achieve by introducing a dependence on the Laplacian. The investigation of energy densities for Laplacian dependent correlation models will be the subject of future work.

5.6 Spin–Resolved Energy Densities

In the discussion thus far, only the total energy densities & local ACs have been considered. However, it is common for DFAs of the correlation energy to be constructed as the sum of spin–parallel and spin–antiparallel components, each of which may be approximated by a different expression; the exchange energy by definition only comprises spin–parallel contributions. It follows therefore that resolving the energy densities & local ACs into spin–parallel & spin–antiparallel components may provide additional insights into how the characteristics of these two correlation energy densities compare and thus how their contributions could be accurately treated in DFTs.

The definition for the spin–resolved correlation energy density may be readily obtained by omitting the spin integration in the construction of the two–electron density in (5.3), defining its spin–resolved form as

$$P^{\lambda}_{2,\sigma_1\sigma_2}(\mathbf{r}_1,\mathbf{r}_2) = N(N-1)\int\ldots\int \left|\Psi_\lambda(\tau_1,\tau_2,\ldots,\tau_N)\right|^2 \mathrm{d}\tau_3\ldots\mathrm{d}\tau_N \qquad \tau_i = \{\mathbf{r}_i,\sigma_i\}. \tag{5.44}$$

Substituting this into (5.5) yields the definition of the spin–resolved *xc* energy density at interaction strength λ as

$$w^{\sigma_1\sigma_2}_{\mathrm{xc},\lambda}(\mathbf{r}_1) = \frac{1}{2\rho_{\sigma_1}(\mathbf{r}_1)}\int\frac{P^{\lambda}_{2,\sigma_1\sigma_2}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 - \frac{1}{2}\int\frac{\rho_{\sigma_2}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2. \tag{5.45}$$

Given that the exchange energy density only contributes in the spin–parallel terms, it may be subtracted from the relevant components of (5.45) and the remainder aggregated into spin–parallel and spin–antiparallel correlation energy densities respectively as

$$\begin{aligned} w^{\sigma\sigma}_{\mathrm{c},\lambda}(\mathbf{r}_1) &= \frac{1}{2\rho_\alpha(\mathbf{r}_1)}\int\frac{P^{\lambda}_{2,\alpha\alpha}(\mathbf{r}_1,\mathbf{r}_2)-P^{0}_{2,\alpha\alpha}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 + \frac{1}{2\rho_\beta(\mathbf{r}_1)}\int\frac{P^{\lambda}_{2,\beta\beta}(\mathbf{r}_1,\mathbf{r}_2)-P^{0}_{2,\beta\beta}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 \\ w^{\sigma\sigma'}_{\mathrm{c},\lambda}(\mathbf{r}_1) &= \frac{1}{2\rho_\alpha(\mathbf{r}_1)}\int\frac{P^{\lambda}_{2,\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 + \frac{1}{2\rho_\beta(\mathbf{r}_1)}\int\frac{P^{\lambda}_{2,\beta\alpha}(\mathbf{r}_1,\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 - \frac{1}{2}\int\frac{\rho_\alpha(\mathbf{r}_2)+\rho_\beta(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2. \end{aligned} \tag{5.46}$$

The spin–resolved correlation energy densities have been computed for the Beryllium atom and the N_2 molecule at equilibrium geometry over a range of λ. For both systems, the Lieb optimisation is carried out with the QUEST rapid development electronic structure package[63] using a CCSD(T) reference; the bespoke implementation of coupled cluster analytical gradients implemented in QUEST is used to calculate the spin–resolved & orbital–relaxed one and two–electron densities at the CCSD(T) level, with which the correlation energy densities are computed. The spin–parallel & spin–antiparallel λ–averaged correlation energy densities and local ACs are plotted for Be & N_2 in Figures 5.22 & 5.23 respectively.

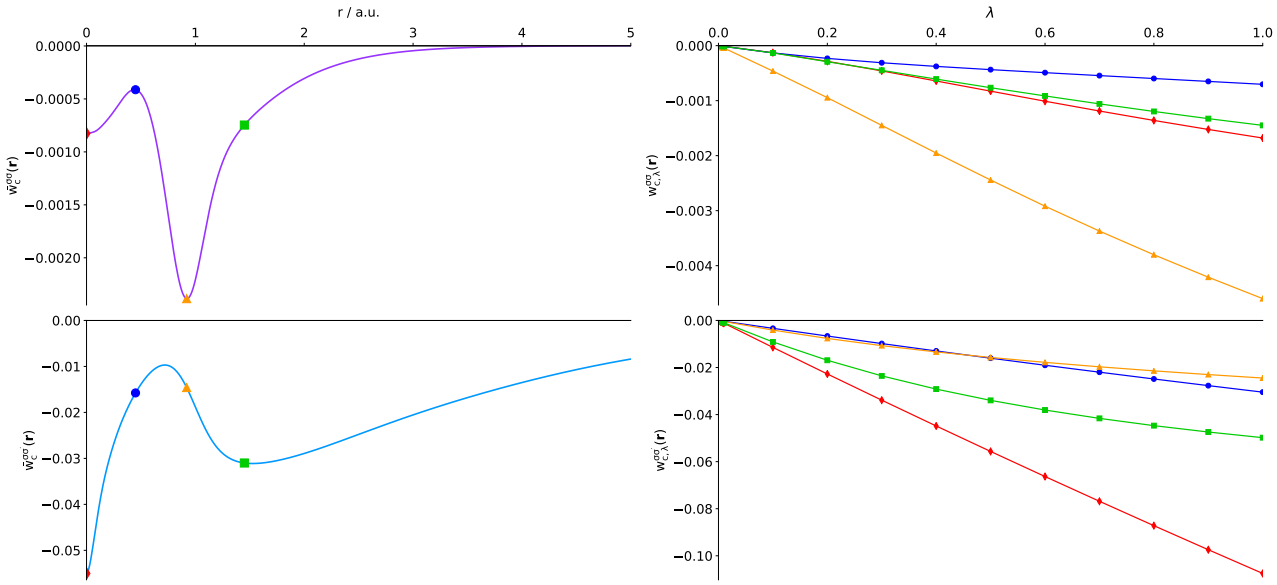

Figure 5.22: The spin–resolved λ–averaged correlation energy density (left) and the local AC at several points in space (right) for the Be atom, computed from a CCSD(T) reference density in the *u*-aug-cc-pCVTZ basis.

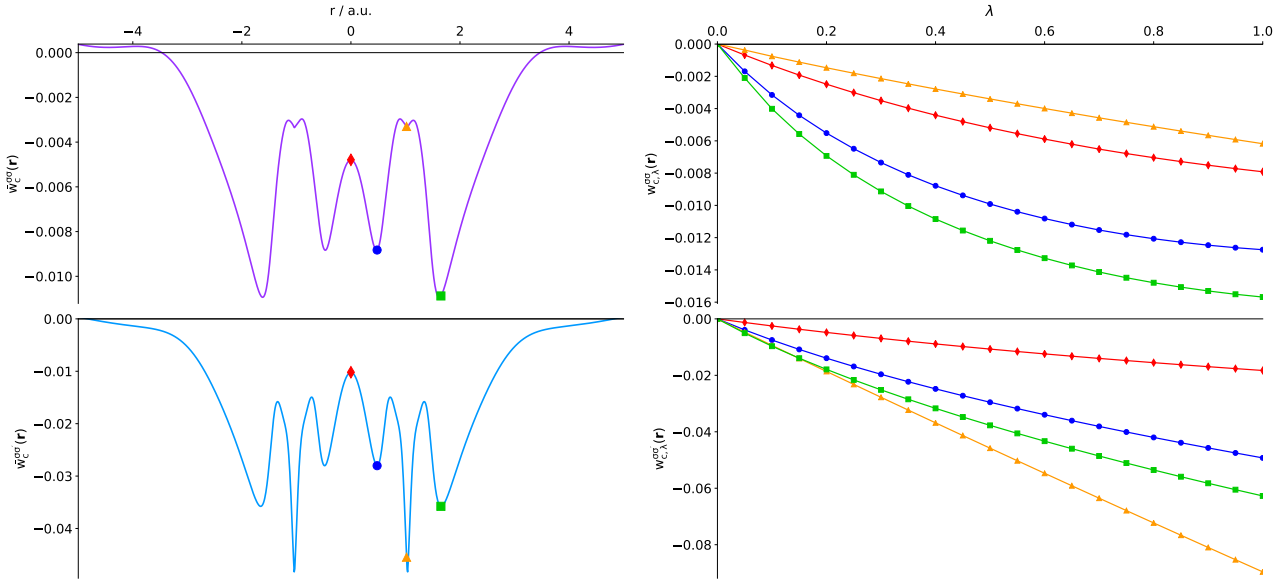

Figure 5.23: The spin–resolved λ–averaged correlation energy density (left) and the local AC at several points in space (right) along the principle axis of the N_2 molecule, computed from a CCSD(T) reference density in the aug-cc-pCVDZ basis.

What is immediately observed in Figures 5.22 & 5.23 is that the spin–parallel λ–averaged correlation energy density is roughly an order of magnitude smaller than the spin–antiparallel equivalent. Furthermore, whilst the spin–antiparallel λ–averaged correlation energy density is of greatest magnitude at the nuclei, the spin–parallel energy density appears to exhibit no significant features at the nuclei and is of greatest magnitude at the inter–shell peaks, delimiting the core, valence and outer regions of the electron density. Whilst a full explanation for these differences could not reasonably be deduced from these early observations, one possible rationalisation of these results may be argued by considering

the spin–parallel term as a higher–order effect of exchange, or a 'correction' to the exchange energy density - which dominates the spin–parallel *xc* energy density $w_{xc,\lambda}^{\sigma\sigma}(\mathbf{r}_1)$ but which is invariant in λ. The spin–antiparallel correlation energy density typically constitutes the primary element of the correlation energy density.

Interestingly, there appear to be differences between the spin–parallel and spin–antiparallel local ACs. Other than differences in magnitude, these may not be discernible in Figure 5.22 but can be seen in Figure 5.23; the spin–parallel local AC tends to exhibit a much more pronounced curvature than its spin–antiparallel counterpart. The contrast in curvature of the local ACs could be seen to tie–in to the above rationalisation of spin–parallel correlation as a higher–order correction to exchange effects. The greater curvature of the spin–parallel local AC suggests that it describes a more nondynamical part of the correlation, for which the exchange–like behaviour was discussed in Section 5.4. Further examination of spin–resolved correlation energy densities in systems known to exhibit significant nondynamical correlation will be the subject of future work.

5.7 Conclusions

We have presented an approach for calculating λ–averaged energy densities in the gauge of the exchange–correlation hole. These quantities may be considered as a target for modelling by DFAs, defined in this gauge. The exchange energy density is relatively smooth with minor features close to the inter-shell structure in atomic densities. Furthermore, we have confirmed that the Becke–Roussel model, which defines an exchange energy density in the gauge of the exchange–correlation hole, delivers a reasonable description of this component of the energy density.

The correlation component of the energy density shows much more complex structure. Interestingly this structure aligns well with the nodes of $\nabla^2\rho$, suggesting that this quantity should be of greater use in the construction of DFAs. This is in agreement with recommendations by Cancio *et al.*[161] in studies of the exchange-only DFAs, based on observations of the full exchange–correlation energy density in Silicon[162]. Further insight from the present work is offered by the ability to resolve the energy density into its exchange and correlation components separately, revealing the more elaborate structure of the latter and its alignment with features of $\nabla^2\rho$, suggesting its possible role as a key variable in the formulation of new DFAs for the more accurate treatment of dynamical correlation. This correspondence is particularly apparent in the comparison between energy densities and the Laplacian in the diatomic molecules N_2 & HF.

A local adiabatic connection at each point in space is defined by the λ–dependence of the energy densities, and the observations for Be and Ne suggest that this local connection captures the nature of the correlation effects. The area enclosed by the local AC gives the value of the energy density at each point in space and, of course, varies substantially across the system.

However, the shape of the local AC still retains extra information about the nature of the correlation; higher order dependence on the interaction strength introduces more rapid curvature in the local AC. Observations for the Be and Ne atoms show that such effects can be localised to different regions of space, with less pronounced curvature in the core regions and more pronounced curvature in the valence regions. This feature is more pronounced in Be than the other atoms, consistent with the near degeneracy effects present in this system. Knowledge of the nature of the local AC should be useful for constructing DFAs via *local* interpolation models, which offer size–consistent models for the exchange–correlation energy in the absence of degeneracy[163,164]. The local ACs constructed for the dissociating H_2 molecule are a striking example of how the curvature of the AC increases as near–degeneracy becomes more pronounced and correlation becomes predominantly nondynamical. The examination of spin–resolved energy densities and local ACs further suggests that distinguishing between spin–parallel and spin–antiparallel correlation may be significant for the accurate treatment of systems with substantial nondynamical correlation.

A particularly interesting prospect arising from the current work is the exploration of this connection between spatial resolution of the correlation energy and its dependence on the interaction strength at each point in space. The results here suggest that the development of a topological-DFT, in which the λ–dependence of the functional is explicitly connected to the topology of the local density, may be fruitful. Such an approach would allow more careful adaptation of the λ–dependence of DFAs to the regions in which they are applied. Furthermore, our results suggest that $\nabla^2\rho$ should be a key variable for identifying appropriate regions.

For general systems with a mixture of dynamical and nondynamical correlation, it may be possible to derive expressions that are explicitly dependent on the Laplacian of the charge density which yield an accurate reconstruction of the correlation energy density per electron, to be spatially integrated with the density using standard DFT numerical quadrature to obtain an accurate estimate of the correlation energy. Equally, consideration of the local adiabatic connection may be necessary if significant nondynamical correlation effects are present, where the inclusion of higher

order terms in λ would be required to accurately predict $\bar{w}_c(\mathbf{r})$. We expect that the methodology outlined here will be an essential tool for these developments.

A number of areas for future work are now being explored. These include: the generalisation of this procedure to deal with generalised range–dependent ACs[101–103], for example with the widely used error–function attenuated interactions[165]; the application of this method to a wide range of molecular systems to provide benchmark data for the development of DFAs; the testing of existing models for exchange–correlation energy densities defined in the gauge of the exchange–correlation hole, and the development of new models in this context; these are explored in Chapter 6.

6 Modelling The Correlation Energy Density

6.1 Introduction

In the previous chapter the λ–averaged correlation energy densities for a number of atoms and small molecules, calculated using the Lieb optimisation, were presented as a novel target to be modelled by new density functional approximations. Examination of local quantities suggested that the Laplacian of the density could be an essential component of functional approximations that accurately reconstruct the local correlation energy density.

Additionally, examination of the form of the local AC in these systems indicates that the extent to which the correlation present may be characterised as dynamical or nondynamical can vary spatially within a system, and that explicit consideration of the curvature of the AC at each point in space may be essential for accurate treatment of strongly correlated systems.

In the present chapter, two different approaches to modelling the correlation energy density are considered. In Section 6.2, an approximation for the correlation energy density obtained by modelling the correlation hole using local quantities is described and its implementation tested for several atoms. In Section 6.3, models for the correlation energy density involving the approximation of the AC at each point in space using interpolation models are examined and tested for small systems.

6.2 A Model Correlation Hole

One difficulty in evaluating functional approximations of the local correlation energy against the accurate λ–averaged correlation energy densities presented in the previous chapter is that very few approximations exist that are defined in the correct gauge – i.e that of the electrostatic potential of the exchange–correlation hole. A large proportion of the correlation functionals that are presently used in DFT calculations are derived from the density gradient expansion, but have been simplified through integration by parts. In this transformation, whilst the global properties of the functional are unchanged, the functional itself is transformed into a different gauge and as such it is not meaningful to draw comparisons between the local values of such approximations and the λ–averaged correlation energy densities of this work.

In this discussion, one of the few correlation functionals known to be in the gauge of the *xc*–hole is examined in comparison to the λ–averaged correlation energy densities to test how accurately these may be modelled by local quantities. The correlation functional to which this discussion pertains is the semi–local first–principles based model of Modrzejewski *et. al.*[166], which was derived by its authors with the intention of creating a functional that would complement the DFT–D3 dispersion corrections of Grimme[167]. This functional of Modrzejewski *et. al.* is a relatively unknown density functional approximation and as yet the literature lacks any particular name with which it is referred to; in the subsequent discussion, the acronym **MCS** will be used to refer to the functional.

6.2.1 The MCS Model

The feature of the MCS functional that makes it an interesting functional to investigate is that its derivation ensures that it is in the gauge of the *xc*–hole. In the simplest sense, the functional is constructed from the analytical form *xc*–hole by subtracting from this the Becke–Roussel (BR) model exchange hole[93] (see subsection 5.3.2 for details of this model), and parametrising the terms that remain with known data for the uniform–electron gas[168].

The starting point for this model was to consider the integrand of the AC $\mathcal{W}_{c,\lambda}$ defined by (3.179) and express this in terms of the spin–resolved correlation hole,

$$\mathcal{W}_{c,\lambda} = \frac{1}{2}\sum_{\sigma\sigma'}\int\int \frac{h_{c,\lambda}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2)\rho_\sigma(\mathbf{r}_1)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \tag{6.1}$$

where the definition of the spin–resolved λ–dependent *xc*–hole and therefore correlation hole is given by

$$\begin{aligned} h_{\mathrm{xc},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2) &= \frac{P_{2\,\sigma\sigma'}^{\lambda}(\mathbf{r}_1,\mathbf{r}_2)}{\rho_\sigma(\mathbf{r}_1)} - \rho_{\sigma'}(\mathbf{r}_2) \\ h_{\mathrm{c},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2) &= h_{\mathrm{xc},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2) - h_{\mathrm{x}}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2). \end{aligned} \tag{6.2}$$

To reduce this to a simpler form, the spherical symmetry of the Coulomb operator exploited such that the correlation hole is replaced by the spherically–averaged correlation hole, with the second electron coordinate replaced by radial distance, s. This transformation allows (6.1) to be re–written as

$$\mathcal{W}_{\mathrm{c},\lambda} = \frac{1}{2}\sum_{\sigma\sigma'}\int\int_0^\infty \frac{h_{\mathrm{c},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,s)\rho_\sigma(\mathbf{r}_1)}{s} 4\pi s^2 \mathrm{d}s\mathrm{d}\mathbf{r}_1 \tag{6.3}$$

with the spherically–averaged correlation hole given by

$$h_{\mathrm{c},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,s) = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} h_{\mathrm{c},\lambda}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_1+s)\sin\theta_s \mathrm{d}\theta_s \mathrm{d}\phi_s. \tag{6.4}$$

From this expression, Modrzejewski *et. al.* then postulated the following model for the correlation hole, comprising of a slightly different form and different dependence on local quantities (to be described by the terms a, b, c and d below) for the spin–parallel and spin–antiparallel correlation holes respectively,

$$\begin{aligned} h_{\mathrm{c},\lambda}^{\alpha\alpha}(\mathbf{r}_1,s) &= s^2(a_{\alpha\alpha} + b_{\alpha\alpha}s + c_{\alpha\alpha}s^2)\exp(-d_{\alpha\alpha}s) \\ h_{\mathrm{c},\lambda}^{\alpha\beta}(\mathbf{r}_1,s) &= (a_{\alpha\beta} + b_{\alpha\beta}s + c_{\alpha\beta}s^2)\exp(-d_{\alpha\beta}s). \end{aligned} \tag{6.5}$$

The short range expansions of the spin–parallel and spin–antiparallel *xc*–holes[169] were already known to be significantly attenuated by cusp conditions[170], suggesting that their short range expansions at a given λ may be reasonably modelled as

$$\begin{aligned} h_{\mathrm{xc},\lambda}^{\alpha\beta}(\mathbf{r},s) &= (B_{\alpha\beta} - \rho_\beta) + \lambda B_{\alpha\beta}s + \ldots \\ h_{\mathrm{xc},\lambda}^{\alpha\alpha}(\mathbf{r},s) &= -\rho_\alpha + \left(B_{\alpha\alpha} - \frac{1}{6}\nabla^2\rho_\alpha\right)s^2 + \frac{\lambda}{2}B_{\alpha\alpha}s^3 + \ldots, \end{aligned} \tag{6.6}$$

which is of the form of the BR model exchange hole,

$$h_{\mathrm{x}}^{\alpha\alpha}(\mathbf{r},s) = -\rho_\alpha - \frac{1}{6}\left[\nabla^2\rho_\alpha - 2\tau_\alpha + \frac{1}{2}\frac{(\nabla\rho_\alpha)^2}{\rho_\alpha}\right]s^2 + \ldots. \tag{6.7}$$

It follows that subtraction of (6.7) from (6.6) leads to a comparable expansion of the correlation hole. Given that two electrons will only experience mutual exchange interactions if they are of parallel spin, the BR model exchange hole is subtracted only from the spin–parallel expansion for the *xc*–hole, with the spin–antiparallel *xc*–hole hence identified as equivalent to the spin–antiparallel correlation hole. The resulting expressions for the spin–antiparallel and spin–parallel model correlation holes, at a given λ, are given respectively as

$$\begin{aligned} h_{\mathrm{c},\lambda}^{\alpha\beta}(\mathbf{r},s) &= (B_{\alpha\beta} - \rho_\beta) + \lambda B_{\alpha\beta}s + \ldots \\ h_{\mathrm{c},\lambda}^{\alpha\alpha}(\mathbf{r},s) &= \left(B_{\alpha\alpha} - \frac{1}{3}D_\alpha\right)s^2 + \frac{\lambda}{2}B_{\alpha\alpha}s^3 + \ldots, \end{aligned} \tag{6.8}$$

where the term D_α is defined as

$$D_\alpha = \tau_\alpha - \frac{|\nabla\rho_\alpha|^2}{4\rho_\alpha} \qquad \tau_\alpha = \sum_i^{N_\alpha}|\nabla\varphi_i^\alpha|^2. \tag{6.9}$$

Modrzejewski *et. al.* then approximated the terms $B_{\alpha\alpha}$ and $B_{\alpha\beta}$ as short–range expansions of the spin–resolved uniform electron gas pair distribution functions, derived by Gori–Giorgi *et. al.*[168]; the numerical expansions may be found in Ref. 166. Expressions for the remaining terms in (6.5) were obtained through treating (6.5) and (6.8) as pairs of simultaneous equations (for spin–parallel and spin–antiparallel models respectively), the introduction of the constraint that each model hole must vanish when integrated over all space and the application of Levy's scaling relations[85] to each of the two model holes respectively. This analysis was found to yield the following identities for

each of the undetermined quantities in (6.5),

$$a_{\alpha\beta} = B_{\alpha\beta} - \rho_\beta \qquad a_{\alpha\alpha} = B_{\alpha\alpha} - \frac{D_\alpha}{3} \tag{6.10}$$

$$b_{\alpha\beta} = \lambda B_{\alpha\beta} + d_{\alpha\beta} a_{\alpha\beta} \qquad b_{\alpha\alpha} = \frac{\lambda}{2} B_{\alpha\alpha} + a_{\alpha\alpha} d_{\alpha\alpha} \tag{6.11}$$

$$c_{\alpha\beta} = -\frac{1}{12}(a_{\alpha\beta} d_{\alpha\beta}^2 + 3 b_{\alpha\beta} d_{\alpha\beta}) \qquad c_{\alpha\alpha} = -\frac{1}{30}(a_{\alpha\alpha} d_{\alpha\alpha}^2 + 5 b_{\alpha\alpha} d_{\alpha\alpha}) \tag{6.12}$$

$$d_{\alpha\beta} = \frac{2.1070}{r_s^{\alpha\beta}} + \frac{\mathcal{G}}{r_s} \frac{\nabla\rho \cdot \nabla\rho}{\rho^{8/3}} \qquad d_{\alpha\alpha} = \frac{2.6422}{r_s^{\alpha\alpha}} + \frac{\mathcal{G}}{r_s} \frac{\nabla\rho \cdot \nabla\rho}{\rho^{8/3}} \tag{6.13}$$

in which the following definitions apply

$$r_s^{\alpha\alpha} = \frac{(3/\pi)^{\frac{1}{3}}}{2\rho_\alpha^{1/3}} \qquad r_s^{\alpha\beta} = \frac{(3/\pi)^{\frac{1}{3}}}{\rho_\alpha^{1/3} + \rho_\beta^{1/3}} \qquad r_s = \left(\frac{3}{4\pi\rho}\right)^{\frac{1}{3}} \tag{6.14}$$

and where $\mathcal{G}$ is an adjustable parameter that can be tuned such that the functional best complements any given treatment of dispersion correction. In their original work, the authors of the MCS functional proposed that a value of $\mathcal{G} = 0.096240$ would be produce the most reliably accurate predictions of molecular interaction energies when used in combination with DFT-D3 dispersion corrections. However, this parameter may be freely varied without compromising the performance of the functional in the slowly–varying density limit; as can be deduced from the model hole (6.5), the result of varying $\mathcal{G}$ is that the rate of decay of the correlation hole with distance changes. The effects of doing so on the properties of the MCS functional are discussed in the following subsection.

6.2.2 Testing The Correlation Hole Model

In order to investigate the model correlation hole derived by Modrzejewski *et. al.*, it must be manipulated into a form that may be evaluated numerically at a given point in space, and numerically integrated over a grid as is standard in DFT calculations to yield the total correlation energy. With this aim, the λ–specific model was transformed through algebraic integration to a λ–averaged equivalent, and the uniform electron gas quantities introduced as parameters in a polynomial expansion of the individual terms at each point in space. The resulting functional, as presented in Ref. 166 for convenient implementation, is given in (6.15) using the parameters in Table 6.1.

$$E_c^{\alpha\beta} = \int_0^1 \mathcal{W}_{c,\lambda}^{\alpha\beta} d\lambda = \int \rho_\alpha \pi \frac{\mathcal{B}_{\alpha\beta} + \mathcal{A}_{\alpha\beta} d_{\alpha\beta}}{d_{\alpha\beta}^3} d\mathbf{r} \tag{6.15a}$$

$$E_c^{\alpha\alpha} = \int_0^1 \mathcal{W}_{c,\lambda}^{\alpha\alpha} d\lambda = \int \rho_\alpha \pi \frac{8\mathcal{B}_{\alpha\alpha} + 4\mathcal{A}_{\alpha\alpha} d_{\alpha\alpha}}{d_{\alpha\alpha}^5} d\mathbf{r} \tag{6.15b}$$

$$\mathcal{A}_{\alpha\beta} = \frac{\rho_\beta}{r_s^{\alpha\beta}} \left[\left(-P_0 + \sum_{k=1}^{4} P_k (r_s^{\alpha\beta})^k \right) \exp\left(-P_5 r_s^{\alpha\beta}\right) + P_0 \right] - \rho_\beta \tag{6.15c}$$

$$\mathcal{B}_{\alpha\beta} = \frac{\rho_\beta}{(r_s^{\alpha\beta})^2} \left[\left(-Q_0 + \sum_{k=1}^{5} Q_k (r_s^{\alpha\beta})^k \right) \exp\left(-Q_6 r_s^{\alpha\beta}\right) + Q_0 \right] + d_{\alpha\beta} \mathcal{A}_{\alpha\beta} \tag{6.15d}$$

$$\mathcal{A}_{\alpha\alpha} = \frac{D_\alpha}{3 r_s^{\alpha\alpha}} \left[\left(-R_0 + \sum_{k=1}^{2} R_k (r_s^{\alpha\alpha})^k \right) \exp\left(-R_3 r_s^{\alpha\alpha}\right) + R_0 \right] - \frac{D_\alpha}{3} \tag{6.15e}$$

$$\mathcal{B}_{\alpha\alpha} = \frac{D_\alpha}{6(r_s^{\alpha\alpha})^2} \left[\left(-S_0 + \sum_{k=1}^{3} S_k (r_s^{\alpha\alpha})^k \right) \exp\left(-S_4 r_s^{\alpha\alpha}\right) + S_0 \right] + d_{\alpha\alpha} \mathcal{A}_{\alpha\alpha} \tag{6.15f}$$

$$E_c = E_c^{\alpha\beta} + E_c^{\beta\alpha} + E_c^{\alpha\alpha} + E_c^{\beta\beta} \tag{6.15g}$$

The MCS functional, in the form given above, was implemented using the XCFun library[171] with which derivatives of the functional can be evaluated easily by use of *automatic differentiation*. This allowed the functional to be tested immediately as part of a fully self–consistent DFT calculation – in this work, such calculations were carried–out with the London quantum chemistry code[172], with the inclusion of full Hartree–Fock exchange. The resulting density, its derivatives, and kinetic energy density were plotted on a Cartesian grid and subsequently read by a small utility that calculated the value of the MCS correlation energy density and correlation energy per volume element, denoted $w_c(\mathbf{r})$ and $e_c(\mathbf{r})$ respectively, over the grid. The values of the MCS correlation energy density plotted in this way were

k	P_k	Q_k	R_k	S_k
0	1·6960000	3·3560000	1·7750000	3·2050000
1	−0·2763000	−2·5250000	0·0121300	−1·7840000
2	−0·0935900	−0·4500000	−0·0047430	0·0036130
3	0·0038370	−0·1060000	0·5566000	−0·0047430
4	−0·0024710	0·0005532		0·5566000
5	0·7524000	−0·0024710		
6		0·7524000		

Table 6.1: Numerical values of the parameters given in (6.15), required for the evaluation of the MCS functional.

numerically integrated with the density over the Cartesian grid to ensure agreement with the global value for the MCS correlation energy computed by LONDON.

Given the correct local evaluation of the MCS correlation energy density, a simple optimisation procedure was then applied to estimate the value of $\mathcal{G}$ that minimised the difference between the numerically–integrated MCS correlation energy with the given density and the reference correlation energy. This new value of $\mathcal{G}$ was then applied in a subsequent self–consistent DFT calculation with the MCS correlation functional to obtain a new density; this optimisation process was repeated until the stationary value of $\mathcal{G}$ that minimised the difference between the MCS correlation energy and reference correlation energy was obtained. Generally, this procedure would be complete in five cycles or less; both the reference the optimised values $\mathcal{G}$ with the corresponding MCS correlation energies for three atomic systems are given in Table 6.2.

Quantity	Helium	Neon	Argon
$\mathcal{G}_{\text{ref}}$	0.096240	0.096240	0.096240
$\mathcal{G}_{\text{opt}}$	0.004630	0.001400	0.000750
$E_{\text{c}}^{\mathcal{G}_{\text{ref}}}$	-0.013871	-0.150275	-0.315758
$E_{\text{c}}^{\mathcal{G}_{\text{opt}}}$	-0.037736	-0.318502	-0.625342
$E_{\text{c}}^{\text{ref}}$	-0.041249	-0.376824	-0.620232

Table 6.2: Correlation energies obtained self–consistently using the MCS functional with the reference value of $\mathcal{G}$ and with the optimised value of $\mathcal{G}$, denoted $\mathcal{G}_{\text{ref}}$ and $\mathcal{G}_{\text{opt}}$ respectively, with the accurate correlation energies from Chapter 5 quoted for comparison.

It can be observed from the present data that the value of $\mathcal{G}$ suggested in Ref. 166 gives poor estimates of the correlation energy in single–point KS DFT calculations. However, decreasing the value of $\mathcal{G}$ (seemingly by a larger amount with increasing atomic number) does appear to yield more accurate estimates of the correlation energy. To some extent, this may be rationalised by referring back to the model itself given in (6.5), from which it may be deduced that a reduced value of $\mathcal{G}$ leads to a reduced spatial attenuation of the correlation hole, amounting to an increase in the total correlation effects that the model attempts to approximate. It follows that lowering $\mathcal{G}$ in these calculations yields a more negative estimate of the correlation energy, improving its agreement with the respective reference correlation energies of these atoms (taken from Chapter 5).

Although this argument to some extent explains the data in Table 6.2, the main feature of the MCS functional that makes it of interest here is that its value can be meaningfully compared locally with the reference correlation energy densities of Chapter 5. The correlation energies per volume element and correlation energy densities calculated for each item in Table 6.2 are plotted radially from the nucleus of the given atom in Figures 6.1, 6.2 & 6.3.

From these three sets of plots, a pattern can be identified between the changes in $w_{\text{c}}(\mathbf{r})$ & $e_{\text{c}}(\mathbf{r})$ and the variation in $\mathcal{G}$ with which a more accurate total correlation energy is obtained. It appears that, with the reference value of $\mathcal{G}$, the similarities between the plots of $w_{\text{c}}(\mathbf{r})$ for the functional and the accurate data are pronounced; the functional is certainly reflecting every feature in the structure of the accurate energy density, however the magnitude of the functional is attenuated more rapidly than the accurate energy density. In the corresponding plots of $e_{\text{c}}(\mathbf{r})$, the values given by the functional are far lower in magnitude than the reference line, although it captures all the features of the reference $e_{\text{c}}(\mathbf{r})$ approximately in proportion, the small errors in $w_{\text{c}}(\mathbf{r})$ (possibly exacerbated by errors in the density arising from the implied inaccuracies in the potential) become much larger when multiplied by the density volume element. It is the integral of $e_{\text{c}}(\mathbf{r})$ that gives the total correlation energy, hence the low values yielded by the reference $\mathcal{G}$.

It follows from this discussion that, for the optimised values of $\mathcal{G}$ with correspondingly more accurate total correlation

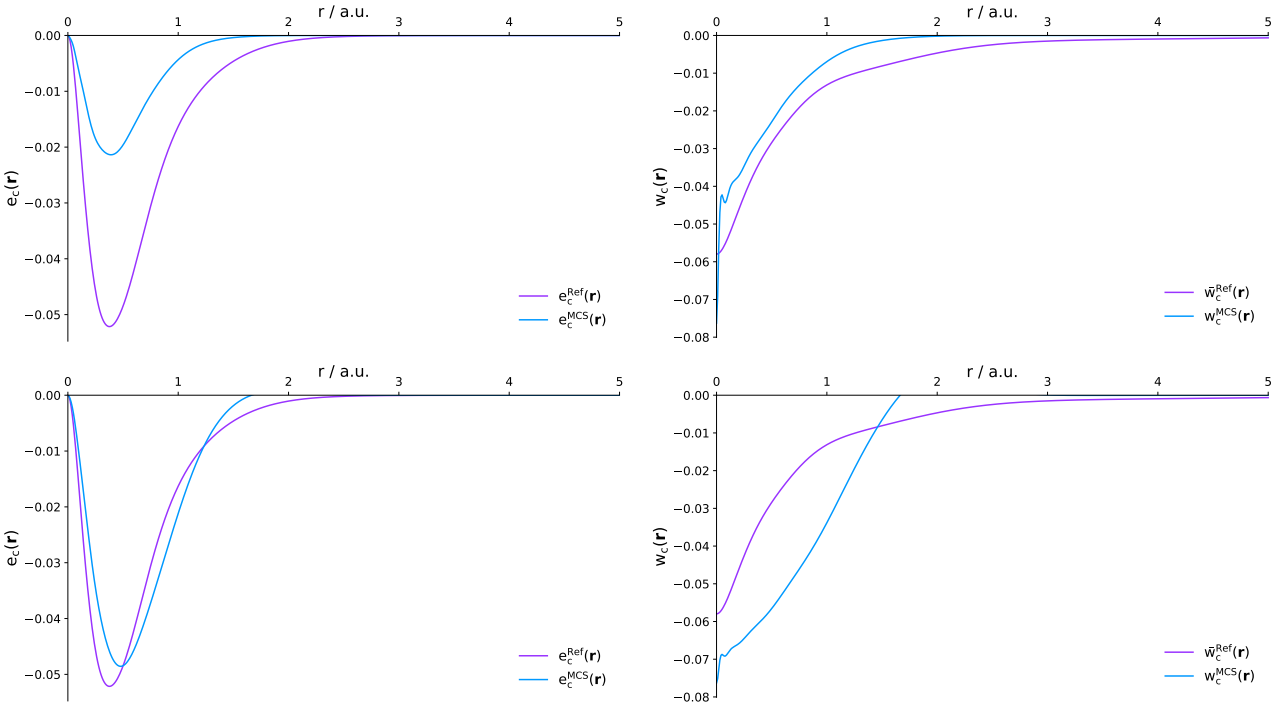

Figure 6.1: The correlation energy per volume element (left) and the correlation energy density (right) in Helium for the MCS model and the reference data from Chapter 5, for the reference value of the variable $\mathcal{G}$ (top) and the optimised value (bottom).

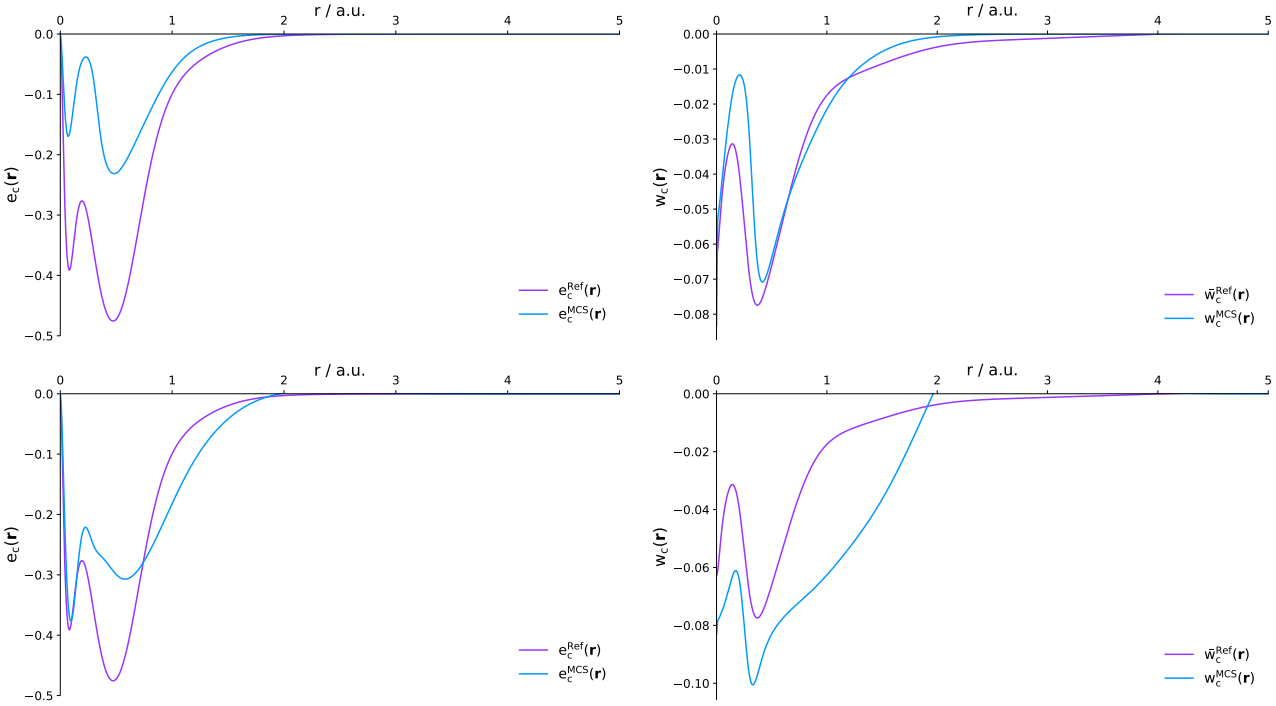

Figure 6.2: The correlation energy per volume element (left) and the correlation energy density (right) in Neon for the MCS model and the reference data from Chapter 5, for the reference value of the variable $\mathcal{G}$ (top) and the optimised value (bottom).

energies, it would be expected that the agreement between the curves of $e_c(\mathbf{r})$ given by the functional and of the accurate data would visibly show greater agreement than those for the reference value of $\mathcal{G}$. Examination of the relevant plots in Figures 6.1, 6.2 & 6.3 confirms this to be the case.

However, the accompanying changes in the character of $w_c(\mathbf{r})$ were slightly less predictable. The decrease in the value

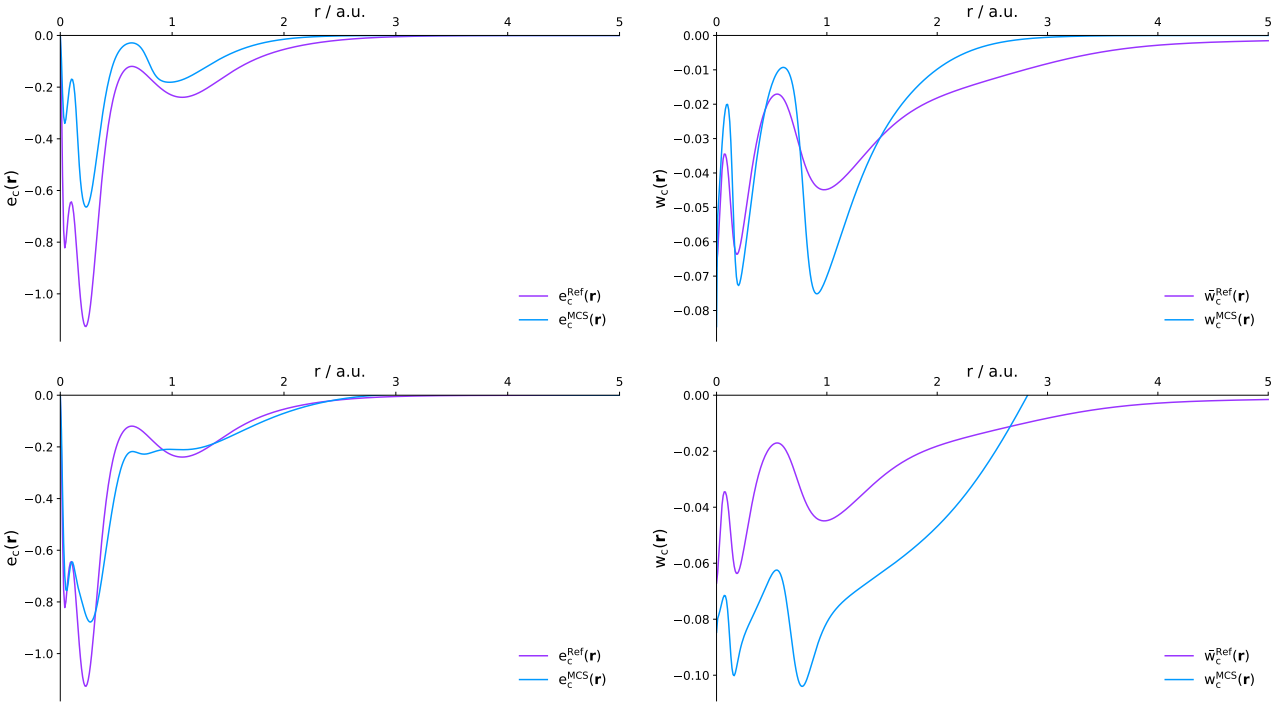

Figure 6.3: The correlation energy per volume element (left) and the correlation energy density (right) in Argon for the MCS model and the reference data from Chapter 5, for the reference value of the variable $\mathcal{G}$ (top) and the optimised value (bottom).

of $\mathcal{G}$ required to improve the correlation energy leads to an increase in the magnitude of $w_c(\mathbf{r})$ in the area relatively close to the nucleus, whilst outside of this region it appears to become positive. This is most likely an artefact of the calculation, arising because the vanishingly small densities in this region reduce the contribution to overall energy from this artefact to a near–negligible level, certainly by comparison to the changes in core region.

Further to this, the $w_c(\mathbf{r})$ yielded seem to become somewhat less resemblant of the accurate picture; the features such as the inter–shell peaks are present, but much less well defined, with the curves appearing more as an 'envelope' that loosely follow the reference energy densities in the core region but with a very large margin of error. This behaviour suggests that the lowering of $\mathcal{G}$ required to improve the total correlation energy has the effect of reducing the attenuation of the correlation hole by such an amount that short range correlation becomes unphysically strong. The restriction of this behaviour to the core region, with artificial positive energy densities instead manifesting at longer range, may be a flaw with the model itself, as the *xc*–hole expansion was heavily truncated such that the cusp condition was satisfied for short–range interaction.

Although the findings discussed here would seem to advocate the use of alternative models to approximate $w_c(\mathbf{r})$, the MCS functional has provided some very encouraging observations that suggest the functional form proposed by Modrzejewski *et. al.* would be a good starting point for further work. Most interestingly the present work shows that it is possible to construct a correlation functional, evaluated from simple local quantities, that can recover the most prominent features of the correlation energy density of an inhomogeneous system such as the inter–shell peaks, despite being parametrised against uniform electron gas data. Further refinement of this model is the subject of ongoing and future study; a new correlation functional derived from MCS could be the first in a class of new DFAs that approximate $w_c(\mathbf{r})$ by explicitly and accurately modelling the correlation hole. In particular, these observations highlight that an improved model for the asymptotic behaviour of the correlation energy density is required.

6.3 Interpolating The Local Adiabatic Connection

6.3.1 Introduction

Since the introduction and widespread application of Kohn–Sham DFT (KS DFT)[65], an increasing number of DFAs for the *xc* energy have been proposed. The simplest DFAs are based on the local density approximation (LDA), as proposed by KS in their 1965 paper[65], in which the *xc* energy is approximated as a functional of the density at a

given point in space. The generalised gradient approximations (GGAs)[91,92,173] go beyond the LDA by modelling the *xc* energy as a functional of the local density and its first derivative. The meta-GGAs[125] are closely related to the GGAs but their functional forms are also dependent on the KS kinetic energy density and/or, less commonly, the Laplacian of the electron density. Further developments led to the introduction of hybrid functionals[95,174], which comprise a mixture of approximate and orbital dependent exchange, and more recently the double–hybrid functionals, that extend the treatment of correlation through inclusion of virtual KS orbitals[175–179]. More recently, local hybrid models such as that proposed in Ref. 180 have been developed, generating interest in this alternative approach with which hybrid functionals can be constructed.

The inclusion of additional dependencies in *xc* functionals has often resulted in significant improvements in their accuracy for general calculations. However, these improvements cannot be described as systematic in the same way that the accuracy of an *ab initio* calculation may be systematically improved by considering a larger number of excited determinants; some DFAs give excellent results for particular systems but perform very poorly otherwise, and *vice versa*. There also remain many problems that none of the currently available DFAs can accurately model. An important example of this, which is pertinent to this work, are *strong correlation* effects, commonly found in systems with near–degenerate orbitals such as the d– and f–block elements, but also in systems where chemical bonds are undergoing formation or fission in the course of a chemical reaction.

In the present work, the problem of constructing DFAs accurate for systems with and without strong correlation is examined using an alternative approach; by modelling the adiabatic connection (AC)[97–100] directly. As described in Section 3.6, the AC provides an exact expressions for the exchange and correlation energies by considering the changes that occur as the strength of electron interaction is smoothly increased from zero to the full physical strength. This formalism has provided the basis for the development of several DFAs[95,181–183], which attempt to interpolate the AC between the non–interacting and physical systems in order to estimate the *xc* energy. An advantage of the AC approach is that it allows the problem of strong correlation to be addressed in a more direct way, by creating interpolation models that are explicitly dependent on the strongly–interacting limit of the AC at which the coupling–constant $\lambda \to \infty$, in addition to its non–interacting limit $\lambda = 0$.

In the literature, DFAs that treat correlation effects via the AC formalism have taken the form of simple models for the AC integrand. These provide approximate descriptions for the AC between the non-interacting KS limit and physical interacting limit. Such models may be parameterised in terms of quantities derived from the KS system and models for points along the AC with higher interaction strength – examples have been constructed using both accurate and approximate data for systems with fractional interaction strength, the physical interaction strength and the strongly–interacting limit. A common observation for all of these approaches is that simple models that are linear in a small set of known parameters are not able to accurately recover all but the simplest dynamical correlation effects. To achieve higher accuracy in the presence of significant strong correlation effects, models that depend in a non–linear manner on the chosen parameters are required. Unfortunately, such models tend to violate size consistency, limiting their possible chemical application. This failing stems from the fact that the parameters used are estimates of global energy contributions.

The present work aims to address this failing by re–formulating AC interpolation models such that a *local* AC is interpolated at each point in space, rather than a global AC interpolated for the entire system; this approach should be more amenable to the construction of models that are size–consistent in the absence of degeneracy[164]. The local AC for several closed–shell atoms has been recently computed[184] to high accuracy between the non–interacting and physical systems using Lieb's convex–conjugate formulation of DFT[67] and the Lieb optimisation method of Refs. 103,111,112,114. These methods and results were discussed in Chapter 5. This approach is used again here to compute the local AC for a number of small systems, from which several parameters are derived that can be used in the construction of local interpolation models. In addition, local information pertaining to the strongly–interacting limit is computed using models for systems at $\lambda = \infty$; together these quantities are used to calculate local analogues of some established global AC interpolation functionals.

In subsection 6.3.2 the *strictly–correlated electron* model is introduced and its potential usefulness in the treatment of strongly interacting systems is discussed. A brief overview of the literature on global AC interpolation models is given in subsection 6.3.3, setting out the rationale behind and the flaws inherent in this approach. The translation of the AC interpolation scheme from a global approach to a local approach is discussed in subsection 6.3.4, in which the rationale for local models is discussed, the local equivalents of some common global AC functionals presented and readily computable approximations for their input parameters examined. The performance of these local AC interpolation models for a range of small systems is presented in subsection 6.3.5, both using accurate input parameters derived from the reference ACs and input parameters approximated with the models described in subsection 6.3.4. In subsection 6.3.6, the AC interpolation approach is used to propose a simple indicator for dynamical and nondynamical correlation and also a lower bound to the exact energy.

6.3.2 The Strictly Correlated Electron Model

The strongly–interacting limit of the AC has recently become the subject of much interest[134,135,156,159,185,186]. The properties of the AC integrand in this limit may reveal highly non–local information relating to correlation effects[155,156,185,187,188] that cannot be derived from local quantities alone. In recent years, study of the strongly–interacting limit in DFT has focused on the *strictly–correlated electron* (SCE) model, in which the electrons are considered to experience an infinite electron–electron interaction strength, $\lambda = \infty$ and which can, in principle, be studied exactly.

Tests on model physical and chemical systems, such as electrons confined in low-dimensional geometries, low-density ultracold dipolar systems, simple stretched bonds and anions, have shown that taking into account this exact behaviour could yield a solution to the strong correlation problem in DFT[159,185,187–191]. However, the exact information encoded in the strongly–interacting limit, described by the SCE functional is not trivial to elucidate; the SCE problem is highly non–local and, although sparse in principle, its non-linearity makes its exact evaluation for general three–dimensional systems computationally complex.

One possible approach to solving the SCE problem relies on the fact that constructing the exact SCE functional for a given density is equivalent to solving an optimal transport problem with a cost function given by the Coulomb interaction[192,193]. This equivalence has led to interest from the applied mathematics community working on optimal transport problems, resulting in the suggestion of several algorithms[190,194,195], together with interesting exact results[196,197].

So far, the SCE problem can be solved exactly for one–dimensional systems[196]. For spherically symmetric systems, a conjectured solution[156] has been proposed and applied[188,191] which is either an exact solution or a highly accurate approximation[197]. By applying methods from optimal transport theory, the SCE problem for the dissociating hydrogen molecule has been solved and both the global[159,190] and local[159] SCE quantities computed.

The SCE functional can be understood as the complement of the KS functional[135,155,156]; the Hohenberg–Kohn universal density functional $F(\rho)$ reduces to the KS functional $T_\mathrm{s}(\rho)$ at the high density, non–interacting limit whilst it reduces to the SCE functional $W_\mathrm{SCE}(\rho)$ in the low density, strongly–interacting limit,

$$F(\rho) \to \begin{cases} T_\mathrm{s}(\rho) = \min\limits_{\Psi_0 \to \rho} \langle \Psi_0 | \hat{T} | \Psi_0 \rangle & \lambda \to 0 \\ W_\mathrm{SCE}(\rho) = \min\limits_{\Psi_\infty \to \rho} \langle \Psi_\infty | \hat{W} | \Psi_\infty \rangle & \lambda \to \infty. \end{cases} \tag{6.16}$$

Given this, it can be argued that the SCE functional is a superior starting point than the KS functional for the description of very strongly correlated systems[159,187,188,190].

In the SCE model, electrons are considered to be infinitely or perfectly correlated at the strongly–interacting limit, thus their positions may be given by an infinite superposition of classical configurations. The fundamental principle is that the positions of all electrons are determined by a collective variable $\mathbf{r}$; a feature that is captured by *co–motion functions* $\mathbf{f}_i(\mathbf{r})$[156]. If a reference electron is at $\mathbf{r}$, then the position of all other electrons in the system will be given by $\mathbf{r}_i = \mathbf{f}_i(\mathbf{r})$[156]. Since the electrons are perfectly correlated, the probability of finding the reference electron at $\mathbf{r}_i$ is equal to the probability of finding the i^th electron at $\mathbf{f}_i(\mathbf{r})$. Therefore, the co-motion functions are defined by the solution to the differential equation[156]

$$\rho(\mathbf{f}_i(\mathbf{r}))\mathrm{d}\mathbf{f}_i(\mathbf{r}) = \rho(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{6.17}$$

A more detailed description of the properties of co–motion functions may be found in Refs. 134,156,188,192.

In terms of co–motion functions, the SCE functional $W_\mathrm{SCE}(\rho)$ may be expressed as[134]

$$W_\mathrm{SCE}(\rho) = \frac{1}{2}\int \rho(\mathbf{r}) \sum_{i=2}^{N} \frac{1}{|\mathbf{r} - \mathbf{f}_i(\mathbf{r})|}\mathrm{d}\mathbf{r}. \tag{6.18}$$

Despite the non–locality of the SCE functional, evident from (6.17), its functional derivative can be readily computed as[185,192]

$$\nabla v_\mathrm{SCE}(\mathbf{r}) = -\sum_{i=2}^{N} \frac{\mathbf{r} - \mathbf{f}_i(\mathbf{r})}{|\mathbf{r} - \mathbf{f}_i(\mathbf{r})|^3}, \tag{6.19}$$

whilst the AC integrand of the SCE functional is simply the *xc* component of the SCE functional, which can be written as

$$\mathcal{W}_\infty(\rho) = W_\mathrm{SCE}(\rho) - J(\rho), \tag{6.20}$$

where $J(\rho)$ is the classical Coulomb energy functional defined in (3.139).

To solve the SCE problem for spherically symmetric systems such as the He and Be isoelectronic series examined in subsection 6.3.4, the approach of Ref. 156 can been used, which is exact if $N = 2$. For atomic densities with $N > 2$, this method represents a close approximation to the true minimum of (6.18), converging on the exact result[197]. For general two and three–dimensional systems, solutions to the SCE problem may be obtained by recasting the problem in the language of mass–transportation theory, in which the SCE functional is computed as the maximum of the Kantorovich dual problem[192]

$$W_{\text{SCE}}(\rho) = \max_{u} \left\{ \int u(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} \;\middle|\; \sum_{i=1}^{N} u(\mathbf{r}) \leqslant \sum_{i=1}^{N-1}\sum_{j>i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \right\} \tag{6.21}$$

where $u(\mathbf{r}) = v_{\text{SCE}}(\mathbf{r}) + c$, $c \in \mathbb{R}$. This optimisation yields both the SCE functional and its functional derivative and may be used to assist in the development of further approximations, a necessary step to widen the potential applications of this approach[185].

6.3.3 DFAs From The Global Adiabatic Connection

The general definition of the global AC integrand at interaction–strength λ was given in Section 3.6 as

$$\mathcal{W}_\lambda(\rho) = \left\langle \Psi_\lambda \middle| \hat{W} \middle| \Psi_\lambda \right\rangle - J(\rho). \tag{6.22}$$

The AC integrand may be characterised by several features that can be exactly defined: the expansion of $\mathcal{W}_\lambda$ in the non–interacting limit is given by[198]

$$\mathcal{W}_\lambda(\rho) = \mathcal{W}_0(\rho) + \mathcal{W}_0'(\rho)\lambda + \ldots \qquad \lambda \to 0, \tag{6.23}$$

whilst its expansion in the strongly–interacting limit can be expressed as[135,155,156]

$$\mathcal{W}_\lambda(\rho) = \mathcal{W}_\infty(\rho) + \mathcal{W}_\infty'(\rho)\lambda^{-1/2} + \mathcal{O}(\lambda^{-n}) \qquad \lambda \to \infty\ ,\ n \geqslant 5/4. \tag{6.24}$$

Here, the value of the integrand and its first derivative in the non–interacting limit, $\mathcal{W}_0(\rho)$ and $\mathcal{W}_0'(\rho)$, are the exchange energy and twice the correlation energy given by second–order *Görling-Levy perturbation theory* (GL2)[160,198] respectively. Görling-Levy perturbation theory allows for the correlation energy in a KS system to be calculated directly from the occupied and virtual KS orbitals as

$$E_{\text{c}}^{\text{GL2}} = \sum_{ia} \frac{\left|\langle \phi_i | \hat{v}_{\text{x}}^{\text{KS}} - \hat{v}_{\text{x}}^{\text{HF}} | \phi_a \rangle\right|^2}{\varepsilon_i - \varepsilon_a} + \frac{1}{4}\sum_{ijab} \frac{|\langle \phi_i\phi_j || \phi_a\phi_b\rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \tag{6.25}$$

where the indices i, j and a, b pertain to occupied and virtual KS spinorbitals respectively, $\hat{v}_{\text{x}}^{\text{KS}}$ is the local KS potential and $\hat{v}_{\text{x}}^{\text{HF}}$ the non–local Hartree–Fock (HF) exchange potential. The second term in (6.25) is analogous to the correlation energy given by second–order Møller–Plesset perturbation theory (MP2)[29], described in Section 2.3, in which ϕ_p and ε_p are canonical HF orbitals and eigenvalues rather than KS ones. The first term accounts for the difference between the KS and HF exchange potentials and has a form similar to a single–excitation term in many–body perturbation theory but which does not feature in the MP2 expression for the correlation energy with canonical HF orbitals due to Brilluoin's theorem.

The analogues of $\mathcal{W}_0(\rho)$ and $\mathcal{W}_0'(\rho)$ at the strongly–interacting limit, $\mathcal{W}_\infty(\rho)$ and $\mathcal{W}_\infty'(\rho)$ respectively, have been studied in Refs. 135,155,156 and the SCE model with which the AC in the strongly–interacting limit may be constructed is discussed in subsection 6.3.2. In addition to these asymptotic limits, the behaviour of $\mathcal{W}_\lambda$ under uniform coordinate scaling is also well–defined[199].

To construct practical DFAs one could consider modelling the AC integrand (6.22) using a function that interpolates between the limits of (6.23) and (6.24). The SCE limit is of particular importance in the present work, however one could also consider models that intercept any other known point on the AC for $\lambda > 0$. Several attempts to develop DFAs based on these ideas have been put forward in the literature, for example in Refs. 112,135,145,155,156,158,181,182,200. Each form makes a choice of a simple model function and a set of parameters on which to base the model. These parameters often include the known exact expressions for the terms in (6.23): $\mathcal{W}_0(\rho) = E_{\text{x}}(\{\phi_i\})$, $\mathcal{W}_0'(\rho) = 2E_{\text{c}}^{\text{GL2}}(\{\phi_p, \varepsilon_p\})$, since these may be computed from the set of Kohn–Sham orbitals ($\{\phi_p\}$) and orbital energies ($\{\varepsilon_p\}$).

The calculation of $\mathcal{W}_0'(\rho)$ requires the GL2 correlation energy[160,198], resulting in a computational cost similar to that of the MP2 model in *ab initio* quantum chemistry[29]. The parameters in the SCE limit entering (6.24) are clearly also of special interest in this context since they can be computed numerically to high accuracy for atomic systems and

molecules with cylindrical symmetry[156,188,191,197]. More frequently, DFAs are derived for points along the AC with $\lambda > 0$, often by employing scaling relations to derive forms from existing DFAs. A similar strategy can also be used to approximate $\mathcal{W}_0'(\rho)$ using a DFA, see for example Ref. 145.

In tandem with choosing a set of exact or approximate values to parameterise a model for the AC, one must also choose an appropriate model function for the AC integrand. A number of these have been suggested and many have been benchmarked in practical applications. One of the simplest and most often used is that of a [1/1] Padé, as suggested by Ernzerhof[181]. A range of forms were suggested by Cohen *et al.* and tested using approximate parameterisations[145], leading to the MCY1 functional in which a [1/1] Padé model is employed. Peach *et al.*[154,201] attempted to disentangle approximations in the choice of parameters from those in the choice of model AC function by utilising near exact KS orbitals and orbital energies derived from full configuration interaction data to calculate $\mathcal{W}_0(\rho)$ and $\mathcal{W}_0'(\rho)$ whilst using the corresponding interacting wavefunctions to evaluate $\mathcal{W}_1(\rho)$ via (6.22). In these studies, it was noted that whilst a particular AC model form may give high accuracy using exact parameters, this may be severely reduced when approximate parameters are employed.

Seidl and co workers[155,200] were the first to make use of the SCE limit in constructing a global AC model, known as the interaction strength interpolation (ISI) functional. The revISI model[135] was later constructed to take account of the $\lambda^{-1/2}$ dependence exhibited by the second term of (6.24), which was not correctly described by the original approach. Teale, Coriani and Helgaker also proposed forms for the AC integrand based on the structure of traditional *ab initio* methodologies[112] and parameterised these forms to intercept values of the AC at any $\lambda > 0$.

The majority of these models for the global AC suffer from the fact they are not size consistent in practice. This deficiency arises from a non–linear dependence on the parameters $\mathcal{W}_0$, $\mathcal{W}_0'$, and a chosen approximation to $\mathcal{W}_{\lambda>0}$. When these global parameters enter in a non–linear fashion (often as ratios), size consistency is generally difficult to achieve. One route forward is to construct local AC models, in which these global parameters are replaced with local values defined at each point in space and which may be more amenable to the construction of models that recover size–consistency (at least in the usual density–functional sense[163]).

6.3.4 Modelling The Local Adiabatic Connection

As was discussed in subsection 5.2.1, the expression for the global AC integrand $\mathcal{W}_\lambda(\rho)$ can be re–written in terms of a local λ–dependent *xc* energy density $w_{\mathrm{xc},\lambda}(\mathbf{r})$ integrated with the density over all space. As subsequently discussed in subsection 5.2.2, the λ–dependent *xc* energy density may be reduced to the form of a local quantity which, when spatially integrated with the density, yields the *xc* energy of the system by taking its integral over $\lambda = 0 \rightarrow 1$ to obtain the coupling–constant averaged (λ–averaged) *xc* energy density,

$$\bar{w}_{\mathrm{xc}}(\mathbf{r}) = \int_0^1 w_{\mathrm{xc},\lambda}(\mathbf{r})\mathrm{d}\lambda. \tag{6.26}$$

This provides a form of the energy density which may be considered as a target to be modelled by *xc* functionals[184]. Given the invariance of the exchange energy density to electron–interaction strength, (6.26) may be trivially resolved into separate exchange and correlation terms as

$$\bar{w}_{\mathrm{c}}(\mathbf{r}) = \bar{w}_{\mathrm{xc}}(\mathbf{r}) - w_{\mathrm{x}}(\mathbf{r}). \tag{6.27}$$

The aim of the local interpolation schemes examined in this work is to approximate $\bar{w}_{\mathrm{xc}}(\mathbf{r})$ and $\bar{w}_{\mathrm{c}}(\mathbf{r})$ through interpolating the local AC. In principle, this approach is analogous to that of the global AC interpolation schemes previously discussed, but rather than depending on quantities pertaining to the global AC, they are instead constructed from their local equivalents. Clearly, a local interpolation will only be physically meaningful if all of the local terms are defined in the same gauge. It is again both convenient and logical to define all local quantities in the gauge of the *xc* hole, in which highly accurate energy densities $w_{\mathrm{xc},\lambda}(\mathbf{r})$ in the range $0 \leqslant \lambda \leqslant 1$ have previously been calculated[184] and additionally can be computed for small systems in the limit $\lambda \rightarrow \infty$[134,159].

Considering the parameters commonly used to interpolate the global AC referred to in the previous section, $\mathcal{W}_0(\rho)$, $\mathcal{W}_0'(\rho)$ and $\mathcal{W}_\infty(\rho)$, local equivalents of each must be defined within the gauge of the *xc* hole to permit the direct translation of global interpolation models into local forms. The local equivalents of each of these quantities and their possible approximations will be examined in the following discussion.

6.3.4.1 The AC In The Non–Interacting Limit

In the global AC, the non–interacting limit is given by the value of the AC integrand at $\lambda = 0$, $\mathcal{W}_0$, which is simply equal to the exchange energy evaluated from the KS orbitals as stated in subsection 6.3.3. The local form of this in the

gauge of the *xc* hole is simply the exchange energy density $w_x(\mathbf{r})$ defined in (5.7). This is equal to half of the non–local Slater potential[86] and is commonly used in local hybrid functionals. Accurate and efficient computational schemes for this quantity have become available in recent years[202,203] and so its use has become more widespread.

6.3.4.2 The Local Slope In The Non–Interacting Limit

The local equivalent of $\mathcal{W}_0'$ is not as simple to define, yet is an essential component of AC interpolation schemes as it provides a measure of the departure from exchange–only behaviour, in other words provides the information from which the correlation energy can be approximated. Whilst in global models it may be calculated by GL2 perturbation theory, there is no analogous expression that yields the local equivalent, which would be defined in the gauge of the *xc* hole as

$$w_0'(\mathbf{r}) = \left.\frac{\partial w_{\mathrm{xc},\lambda}(\mathbf{r})}{\partial\lambda}\right|_{\lambda=0} \equiv \left.\frac{\partial w_{\mathrm{c},\lambda}(\mathbf{r})}{\partial\lambda}\right|_{\lambda=0}, \tag{6.28}$$

and related to the initial slope of the global AC, hence the GL2 correlation energy, by

$$\mathcal{W}_0'(\rho) = 2E_{\mathrm{c}}^{\mathrm{GL2}} = \int w_0'(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r}. \tag{6.29}$$

Numerical Approximation Of The Local Slope

In the absence of an analytical expression for the local initial slope, in this study $w_0'(\mathbf{r})$ is numerically approximated by the method of finite difference, with a series of $w_{\mathrm{xc},\lambda}(\mathbf{r})$ for $\lambda \ll 1$.

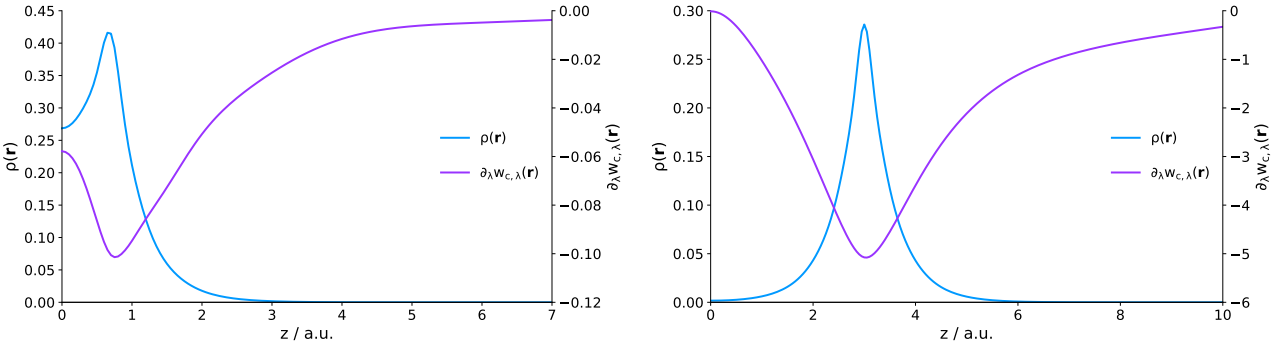

Figure 6.4: Plots comparing the values of $\rho(\mathbf{r})$ and $w_0'(\mathbf{r})$ along the principle axis of the H_2 molecule, measured from the bond midpoint, with bond lengths of 1.4 a.u. (left) and 6.0 a.u. (right).

The approximate local slopes in the H_2 molecule with bond length of 1.4 a.u. and 6.0 a.u. are plotted along the H–H bond in Figure 6.4, along with the densities from which they are calculated, at the FCI level of theory and in the uncontracted aug-cc-pCVTZ basis set[116]. In both cases, the local slope is greatest in magnitude at the nuclei, as has been seen previously in subsection 5.3.4 for atoms[184]. It can further be seen that the magnitude of local slope is significantly larger in the extended H_2 molecule, mirroring observations previously made of the global AC in the dissociating hydrogen molecule[112].

The local slopes in the He and Be isoelectronic series are plotted in Figures 6.5 & 6.6 respectively. It is clear that, with increasing nuclear charge, the charge densities in both series become increasingly contracted. The *x*–axis in both plots has been scaled by nuclear charge, highlighting a contrast in their behaviour with respect to the uniform scaling condition,

$$\lim_{\gamma\to\infty}\frac{E_{\mathrm{xc}}(\rho_\gamma)}{E_{\mathrm{x}}(\rho_\gamma)} = 1, \quad \text{with} \quad \rho_\gamma(\mathbf{r}) = \gamma^3\rho(\gamma\mathbf{r}), \tag{6.30}$$

which is satisfied for non–degenerate KS systems[138]. In Figure 6.5, it can be seen that the slope of the AC becomes less negative with increasing Z, tending to an asymptotic value as $Z \to \infty$, consistent with the scaling relation of (6.30). However, the slope of the AC in the Be isoelectronic series becomes *more* negative with increasing Z, indicating that the scaling relation is not satisfied by this series[110].

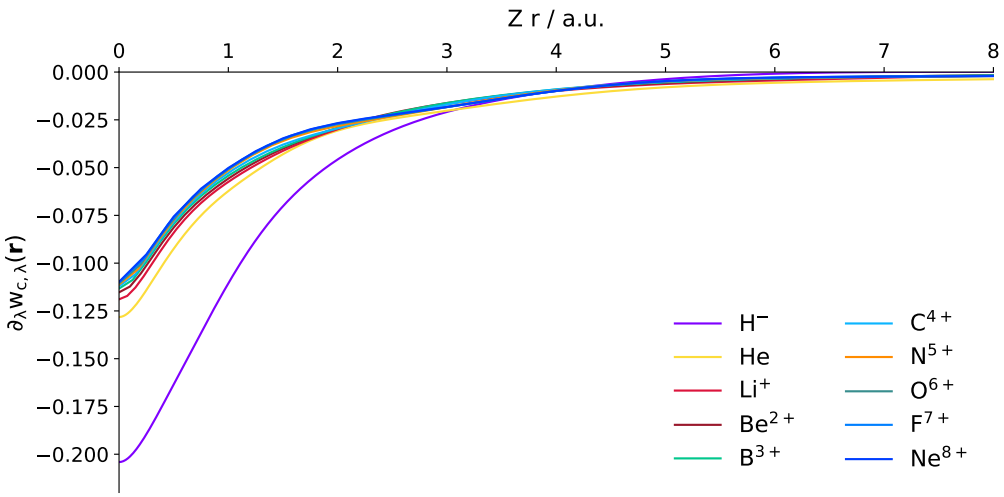

Figure 6.5: Plots of $w_0'(\mathbf{r})$ for the Helium isoelectronic series, with nuclear charges $1 \leqslant Z \leqslant 10$, and with radial distance from the nucleus r / a.u. scaled by nuclear charge.

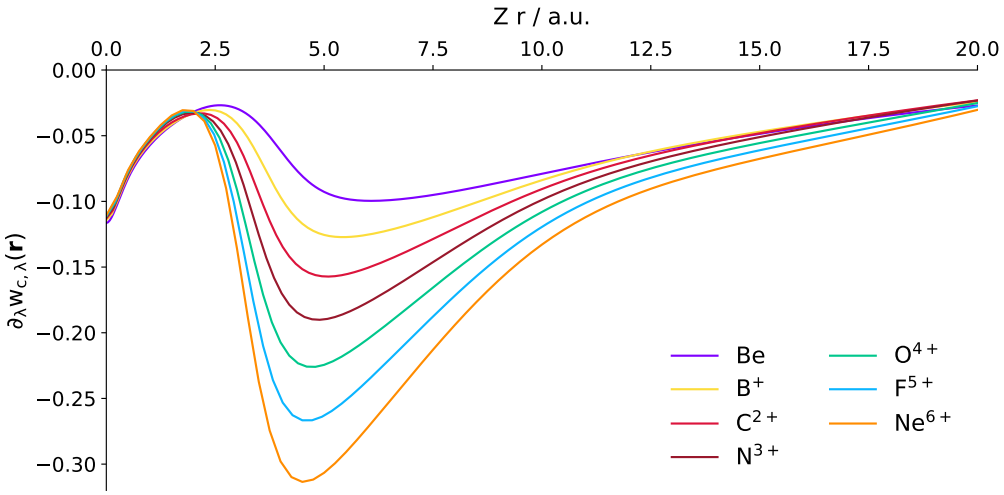

Figure 6.6: Plots of $w_0'(\mathbf{r})$ for the Beryllium isoelectronic series, with nuclear charges $4 \leqslant Z \leqslant 10$, and with radial distance from the nucleus r / a.u. scaled by nuclear charge.

A Functional Approximation For The Local Slope

Whilst it is useful to numerically approximate the local slope for the purposes of evaluating local interpolation schemes, such functionals would only be viable for mainstream use in DFT calculations if they can be described by simple functional forms.

In global models, the initial slope can be calculated directly from the occupied and virtual KS orbitals using GL2 theory as described in subsection 6.3.3. Whilst no local analogue to this readily defined, previous studies of GL2 theory have found that the single–excitation term in (6.25), although not necessarily zero, is nonetheless small in magnitude relative to the MP2–like term[204]. Therefore, it follows that an approximation to the GL2 correlation energy may be obtained by evaluating the MP2 correlation energy[205], $E_c^{\mathrm{GL2}} \approx E_c^{\mathrm{MP2}}$.

Given that an approximation to the global AC slope may be obtained from an MP2–like calculation, it follows that an approximation to the local AC slope may be obtained by deriving a local form of this expression. Whilst MP2 theory treats perturbations of the wavefunction, the analysis may be extended to energy densities in the gauge of the *xc* hole by means of the two–particle density (5.3), as the substitution of (5.4) into (6.28) yields the following,

$$w_0'(\mathbf{r}) = \frac{1}{2\rho(\mathbf{r})} \int \frac{P_2'(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r}', \tag{6.31}$$

where $P_2'(\mathbf{r},\mathbf{r}')$ is the derivative of the two–particle density at $\lambda = 0$,

$$P_2'(\mathbf{r},\mathbf{r}') = \left.\frac{\partial P_2^{\lambda}(\mathbf{r},\mathbf{r}')}{\partial\lambda}\right|_{\lambda=0}. \tag{6.32}$$

Given a non–interacting ground state wavefunction $\Psi^{(0)}$, the perturbed wavefunction Ψ_λ for $|\lambda| \ll |\Psi^{(1)} - \Psi^{(0)}|^2$ can be approximated by the series expansion

$$\Psi_\lambda = \sum_{n=0} \lambda^n \Psi^{(n)}. \tag{6.33}$$

Assuming here that $\Psi^{(0)}$ is non–degenerate and has the form of a Slater determinant, the first–order correction to the wavefunction is given by

$$\Psi^{(1)} = \sum_k \frac{\left\langle \Psi_k^{(0)} | \hat{W} - \hat{V}_{\mathrm{HF}} | \Psi^{(0)} \right\rangle}{\mathcal{E}_0^{(0)} - \mathcal{E}_k^{(0)}} \Psi_k^{(0)}. \tag{6.34}$$

Restricting the space of $\Psi_k^{(0)}$ to only doubly–excited determinants reduces this expression to

$$\Psi^{(1)} = \sum_{ijab} \frac{\left\langle \Psi_{ij}^{ab} | \hat{W} - \hat{V}_{\mathrm{HF}} | \Psi^{(0)} \right\rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \Psi_{ij}^{ab}. \tag{6.35}$$

In MP2 theory, contributions to the correlation energy from singly–excited determinants are necessarily zero due to Brillouin's theorem[29]. However, this is not strictly true in GL2 theory as the singles term in (6.25) makes a small, but non–zero, contribution to the GL2 correlation energy[198]. As such, considering only double–excitations in the model for the local slope can only yield approximations to the local slope; spatial integration of this quantity will not return the exact GL2 correlation energy.

Application of the Slater–Condon rules, described in Section 2.1, to (6.35) allows it to be re–written as

$$\Psi^{(1)} = \frac{1}{4} \sum_{ijab} \frac{\langle \phi_i \phi_j || \phi_a \phi_b \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \Psi_{ij}^{ab}, \tag{6.36}$$

where the coefficient to Ψ_{ij}^{ab} may be identified as an MP2 double–excitation amplitude t_{ij}^{ab},

$$t_{ij}^{ab} = \frac{\langle \phi_i \phi_j || \phi_a \phi_b \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \tag{6.37}$$

To now obtain $P_2'(\mathbf{r},\mathbf{r}')$, it is necessary to take the derivative of the two–particle density corresponding to the perturbed wavefunction, $\Psi_\lambda \approx \Psi^{(0)} + \lambda\Psi^{(1)}$. Substituting this into (5.3) and rearranging the resulting expressions yields the following,

$$\begin{aligned} P_2'(\mathbf{r},\mathbf{r}') &= N(N-1) \int \dots \int \left.\frac{\partial |\Psi_\lambda(\boldsymbol{\tau},\boldsymbol{\tau}',\dots,\boldsymbol{\tau}_N)|^2}{\partial\lambda}\right|_{\lambda=0} \mathrm{d}\sigma\mathrm{d}\sigma'\mathrm{d}\boldsymbol{\tau}_3 \dots \mathrm{d}\boldsymbol{\tau}_N \\ &= N(N-1) \int \dots \int \left| 2\Psi^{(0)}(\boldsymbol{\tau},\boldsymbol{\tau}',\dots,\boldsymbol{\tau}_N)\Psi^{(1)}(\boldsymbol{\tau},\boldsymbol{\tau}',\dots,\boldsymbol{\tau}_N) \right| \mathrm{d}\sigma\mathrm{d}\sigma'\mathrm{d}\boldsymbol{\tau}_3 \dots \mathrm{d}\boldsymbol{\tau}_N \\ &= 2 \left\langle \Psi^{(0)} \left| \hat{P}_2(\mathbf{r},\mathbf{r}') \right| \Psi^{(1)} \right\rangle, \end{aligned} \tag{6.38}$$

in which it is assumed that the wavefunctions are real and where $\hat{P}_2(\mathbf{r},\mathbf{r}')$ denotes the *pair–density operator*

$$\hat{P}_2(\mathbf{r},\mathbf{r}') = N(N-1) \sum_{i\neq j} \delta(\mathbf{r} - \mathbf{r}_i)\delta(\mathbf{r}' - \mathbf{r}_j). \tag{6.39}$$

Substituting (6.36) into (6.38) gives,

$$P_2'(\mathbf{r},\mathbf{r}') = \sum_{ijab} t_{ij}^{ab} \left\langle \Psi^{(0)} \left| \hat{P}_2(\mathbf{r},\mathbf{r}') \right| \Psi_{ij}^{ab} \right\rangle, \tag{6.40}$$

which may then be resolved into the following orbital–explicit expression,

$$P_2'(\mathbf{r},\mathbf{r}') = 2 \sum_{ijab} t_{ij}^{ab} \left\{ \phi_i(\mathbf{r})\phi_j(\mathbf{r}')\phi_a(\mathbf{r})\phi_b(\mathbf{r}')\delta_{\sigma_i\sigma_a}\delta_{\sigma_j\sigma_b} - \phi_i(\mathbf{r})\phi_j(\mathbf{r}')\phi_b(\mathbf{r})\phi_a(\mathbf{r}')\delta_{\sigma_i\sigma_b}\delta_{\sigma_j\sigma_a} \right\}. \tag{6.41}$$

Substituting (6.41) into (6.31) finally results in an expression for the initial slope of the local AC,

$$w_0'(\mathbf{r}) = \frac{1}{2\rho(\mathbf{r})} \sum_{ijab} t_{ij}^{ab} v_{ij}^{ab}(\mathbf{r}), \tag{6.42}$$

where $v_{ij}^{ab}(\mathbf{r})$ is the *antisymmetrised orbital potential*,

$$v_{ij}^{ab}(\mathbf{r}) = \phi_i(\mathbf{r})\phi_a(\mathbf{r}) \int \frac{\phi_j(\mathbf{r}')\phi_b(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \mathrm{d}\mathbf{r}' \delta_{\sigma_i\sigma_a}\delta_{\sigma_j\sigma_b} - \phi_i(\mathbf{r})\phi_b(\mathbf{r}) \int \frac{\phi_j(\mathbf{r}')\phi_a(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \mathrm{d}\mathbf{r}' \delta_{\sigma_i\sigma_b}\delta_{\sigma_j\sigma_a}. \tag{6.43}$$

Multiplying the right–hand side of (6.42) by the density and integrating over all space recovers the MP2–like expression. This is not an exact expression for the local slope, as the single–excitation term of (6.25) is not accounted for. However, the omitted term is generally small relative to the MP2–like term, and vanishes entirely for two–electron systems, hence the expression for the local slope in (6.42) should, in principle, be a fair approximation of the exact local slope.

The approximation in (6.42) will be subject to implementation and testing against the numerical results presented previously in subsequent work. The doubles amplitudes t_{ij}^{ab} are readily obtainable from standard quantum chemical packages and the potential $v_{ij}^{ab}(\mathbf{r})$ can also be readily calculated; however, it would likely be computationally expensive to evaluate on a numerical grid. To reduce this cost a range of techniques, commonly used to accelerate the calculation of integrals in linear–scaling software packages, may be employed [202,206–208].

We note that the behaviour of the local slopes presented in Figures 6.5 & 6.6 may be rationalised in a similar manner to that commonly discussed for global models in terms of (6.25). This is because of the key role of the doubles amplitude t_{ij}^{ab} in (6.42). The doubles amplitude has a dependence on the orbitals and orbital energies that is similar to that of the GL2 energy in (6.25). In Figure 6.5 it is clear that the correlation energy density for the He isoelectronic scales quickly towards an asymptotic constant at Z increases. Furthermore, the local slope decays smoothly with distance from the nucleus, reflecting the behaviour of $v_{ij}^{ab}(\mathbf{r})$.

The behaviour for the Be isoelectronic series in Figure 6.6 is more complex. The HOMO–LUMO gap is known to increase [106] as Z increases from 4 to 10, from which one would expect the correlation energy to become less negative according to the behaviour of t_{ij}^{ab}. In the core region this behaviour is observed, however in the valence region the trend seen is the opposite, with the correlation energy density becoming more negative with increasing Z. This implies that the numerator of t_{ij}^{ab} and the spatial dependence of $v_{ij}^{ab}(\mathbf{r})$ due to the form of the molecular orbitals are dominant in this region.

6.3.4.3 Asymptotic Limits Of The AC

SCE Energy Densities

The SCE theory [135,156,209], introduced in subsection 6.3.2, provides the framework in which exact global and local quantities can be obtained in the $\lambda \to \infty$ limit. From the definition of the global AC integrand in the strongly–interacting limit (6.20), together with that of the SCE functional (6.18), the SCE energy density can be defined as

$$w_\infty(\mathbf{r}) = \frac{1}{2} \sum_{i=2}^{N} \frac{1}{|\mathbf{r} - \mathbf{f}_i(\mathbf{r})|} - \frac{1}{2} v_\mathrm{J}(\mathbf{r}), \tag{6.44}$$

where $v_\mathrm{J}(\mathbf{r})$ is the Coulomb potential. This expression is in the gauge of the electrostatic potential of the *xc* hole (5.2), as shown in Ref. 134. Given that the co–motion functions $\mathbf{f}_i(\mathbf{r})$ parametrise the square of the Ψ_∞ wave function [134,135,156], the SCE energy density can be shown to decay as $w_\infty(\mathbf{r}) \sim -1/2|\mathbf{r}|$, similar to the physical ($\lambda = 1$) and the exchange ($\lambda = 0$) energy densities of (5.5). Its functional derivative (6.19) also exhibits the correct asymptotic decay, $v_\mathrm{SCE}(\mathbf{r}) \sim -1/|\mathbf{r}|$.

Due to the computational difficulties associated with the co–motion functions, SCE energy densities are only relatively easily computed for 1D [189] and spherically symmetric systems [134,156]. To compute the SCE energy density for general 3D systems, the dual Kantorovich SCE formulation [159] may yield a solution, however computational cost still imposes a practical limitation on its use.

In considering approximations to the SCE energy density for use in local interpolation, there are several important properties to examine. The most important of these is that any such approximation must be defined in the gauge of the *xc* hole. Additionally, models that give a reasonably accurate representation of the SCE energy densities, in particular having the correct asymptotic decay, $w_\infty(\mathbf{r}) \to -1/2|\mathbf{r}|$ as $|\mathbf{r}| \to \infty$, are more favourable. It is important to

note however that more accurate $\mathcal{W}_\infty(\rho)$ and $w_\infty(\mathbf{r})$ do not necessarily result in a more accurately approximated $\bar{w}_{\text{xc}}(\mathbf{r})$. This can occur where errors arising from inadequacies of the interpolation model partially cancel errors in the approximate input quantity, leading to a more accurate overall approximation. This will be considered in due course as approximations to the SCE model are examined below.

NLR & PC Energy Densities

Here, the approximations to the SCE functional used in this work, the *nonlocal radius* (NLR)[210] and *point–charge plus continuum* (PC)[200] models, are introduced and their respective properties discussed.

The NLR functional has a non–local structure, inspired by the exact SCE functional[134], with the model *xc* hole[210]

$$h_{\text{xc}}^{\text{NLR}}(\mathbf{r},\mathbf{r}') = -\rho(\mathbf{r}')\,\theta\left(r_{\text{NLR}}(\mathbf{r}) - |\mathbf{r}-\mathbf{r}'|\right), \tag{6.45}$$

where $\theta(x)$ is the step function

$$\theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geqslant 0 \end{cases} \tag{6.46}$$

and $r_{\text{NLR}}(\mathbf{r})$ is the *nonlocal radius*. Wagner and Gori-Giorgi defined $r_{\text{NLR}}(\mathbf{r})$ by generalizing the Wigner–Seitz radius r_{s},

$$r_{\text{s}}(\mathbf{r}) = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{\frac{1}{3}}, \tag{6.47}$$

itself a local function depending only on the density, to nonuniform densities[210] satisfying the relation

$$\int \rho(\mathbf{r}')\,\theta(r_{\text{NLR}}(\mathbf{r}) - |\mathbf{r}-\mathbf{r}'|)\,\mathrm{d}\mathbf{r}' = 1, \tag{6.48}$$

thus defining $r_{\text{NLR}}(\mathbf{r})$ as the radius of a sphere containing one electron. This can equivalently be written as an integral over $\Omega(\mathbf{r})$, the volume of a sphere centred at $\mathbf{r}$ and with radius $r_{\text{NLR}}(\mathbf{r})$,

$$\int_{\Omega(\mathbf{r})} \rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}' = 1. \tag{6.49}$$

For systems with uniform densities, $r_{\text{NLR}}(\mathbf{r})$ simply reduces to the Wigner–Seitz radius. For nonuniform systems, $r_{\text{NLR}}(\mathbf{r})$ contains nonlocal information, as it depends on the density at all the points in the sphere centred at $\mathbf{r}$ and with radius $r_{\text{NLR}}(\mathbf{r})$.

From the model *xc* hole in (6.45), it can be shown[210] that the NLR energy density may be expressed as

$$w_\infty^{\text{NLR}}(\mathbf{r}) = -\frac{1}{2}\int_{\Omega(\mathbf{r})} \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}' \tag{6.50}$$

and thus the global equivalent $\mathcal{W}_\infty^{\text{NLR}}(\rho)$ is given by[210]

$$\mathcal{W}_\infty^{\text{NLR}}(\mathbf{r}) = -\frac{1}{2}\int\int_{\Omega(\mathbf{r})} \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{r}. \tag{6.51}$$

The NLR functional has been implemented in the GAUSSIAN[211] and TURBOMOLE[212] electronic structure packages[180,213]. The NLR model has recently been refined and improved by Bahman, Zhou & Ernzerhof[214] by the addition of a shell of positive charge density, again inspired by the exact[134] SCE *xc* hole and an efficient method of implementing both the NLR and shell models has been derived[214].

The other approximation for the strongly–interacting limit considered here is the PC model of Seidl and coworkers, in which $\lambda \to \infty$ quantities are modelled at a semi–local level[200]. This approximation was developed before empirical results on the exact SCE functional became available[134,156,192] and, in contrast to the NLR approximation, does not model the *xc* hole directly. Detailed analysis does however show that the PC model, at least in the LDA expansion of (6.52), yields energy densities in the gauge of the electrostatic potential of the *xc* hole[134], thus making them suitable for use in local interpolation schemes. Truncating the PC model at the LDA level leads to an energy density expressed as

$$w_\infty^{\text{PC-LDA}}(\mathbf{r}) = -\frac{9}{10}\left(\frac{4\pi}{3}\right)^{\frac{1}{3}}\rho(\mathbf{r})^{\frac{1}{3}}, \tag{6.52}$$

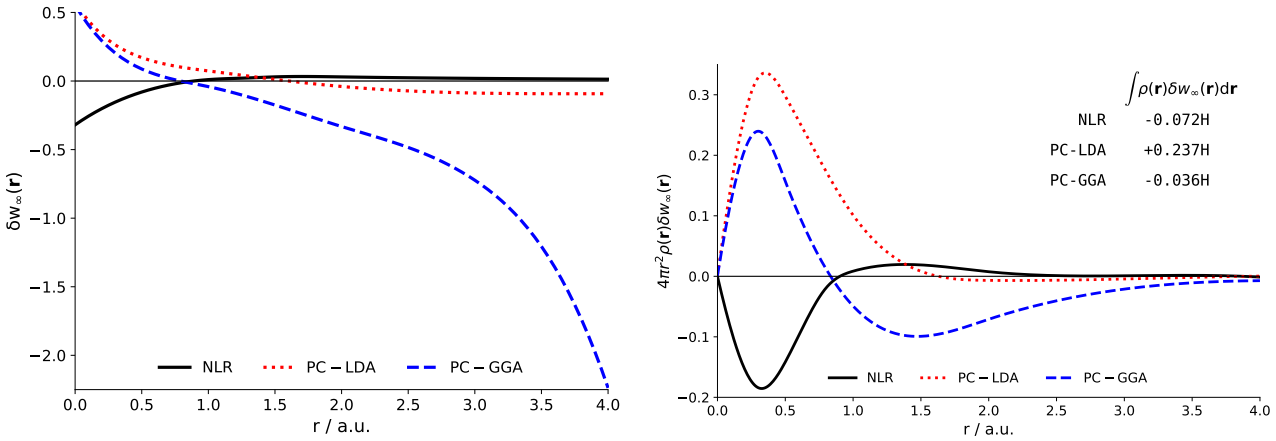

Figure 6.7: Left panel: Plots of the difference between the exact and approximate strongly–interacting limit energy density for the Helium atom, $\delta w_\infty(\mathbf{r}) = w_\infty(\mathbf{r}) - w_\infty^{\text{model}}(\mathbf{r})$, with respect to the distance from the nucleus, r / a.u. Right panel: The quantity $\delta w_\infty(\mathbf{r})$ multiplied by the density and spherical volume element.

whilst truncating at the GGA level yields the energy density,

$$w_\infty^{\text{PC-GGA}}(\mathbf{r}) = w_\infty^{\text{PC-LDA}}(\mathbf{r}) + \frac{3}{350}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}} \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})^{\frac{7}{3}}}. \tag{6.53}$$

In Figure 6.7, the difference between $w_\infty(\mathbf{r})$ computed using SCE theory and its approximate forms are plotted for the Helium atom. It is evident that the NLR energy density is the most faithful to the SCE reference of the three, whilst the two PC forms present a slightly more nuanced picture. At the global level, the PC-GGA approximation appears to have an accuracy superior even to that of the NLR model, with the PC-LDA approximation yielding the greatest error by a considerable margin. However, examination of the errors in the energy densities themselves reveals that the PC-GGA energy density exhibits a highly unphysical asymptotic decay, compared to the broadly reasonable asymptotic decay of the NLR and PC-LDA energy densities, suggesting that its apparent accuracy at the global level may be largely the product of error cancellation[134].

Given the definition of the *xc* hole in (5.2), and by virtue of the *xc* hole sum rule

$$\int h_{\text{xc}}^{\lambda}(\mathbf{r}, \mathbf{r}')\, \mathrm{d}\mathbf{r}' = -1, \tag{6.54}$$

it can be seen that $w_{\text{xc},\lambda}(\mathbf{r}) \to -1/2|\mathbf{r}|$ as $|\mathbf{r}| \to \infty$, for any given λ[134]. In the same way, it can be seen from (6.50) & (6.49) that the NLR energy density will exhibit the correct asymptotic behaviour. However, it is apparent from (6.52) & (6.53) that this will not be the case for the PC-LDA & PC-GGA energy densities respectively.

In Figure 6.8 the SCE, NLR, PC-LDA and PC-GGA energy densities are plotted for the Ne^{6+} ion (an example pertinent to later discussion). It can be seen that, as in the case of the Beryllium atom discussed in Ref. 210, the NLR energy density lies above the SCE energy density in the core region and below in the valence region[210]. Additionally, it can be seen that the PC-LDA and PC-GGA energy densities have the predicted unphysical asymptotic behaviour, with the latter even becoming positive at long range.

This failing can present a challenge to the use of the PC models in local interpolation schemes as it causes both $w_\infty^{\text{PC-LDA}}(\mathbf{r})$ and $w_\infty^{\text{PC-GGA}}(\mathbf{r})$ to intersect $w_{\text{x}}(\mathbf{r})$. In two of the local interpolation schemes used in the present work, the interpolated $w_{\text{xc},\lambda}(\mathbf{r})$ would have a non–zero imaginary component and thus be unphysical in regions where $w_\infty(\mathbf{r}) > w_{\text{x}}(\mathbf{r})$ as these models assume that $w_{\text{xc},\lambda}(\mathbf{r})$ monotonically decreases in λ. In contrast to the global AC, the monotonicity of the local AC (in the gauge of the *xc* hole) has not been formally proven; its assumption in these models is rationalised by the absence of known antithetic examples in Coulombic systems[184,215]. For practical purposes, the interpolation schemes were slightly adapted when using PC model approximations such that $w_{\text{xc},\lambda}(\mathbf{r}) \to w_{\text{x}}(\mathbf{r})$ where $w_\infty(\mathbf{r}) > w_{\text{x}}(\mathbf{r})$.

6.3.4.4 Local Interpolation Models

The local interpolation models examined here are largely simple translations of the well–established global interpolation models into a local form. This was done for the model of Seidl, Perdew & Levy (SPL)[155], the simplified model of

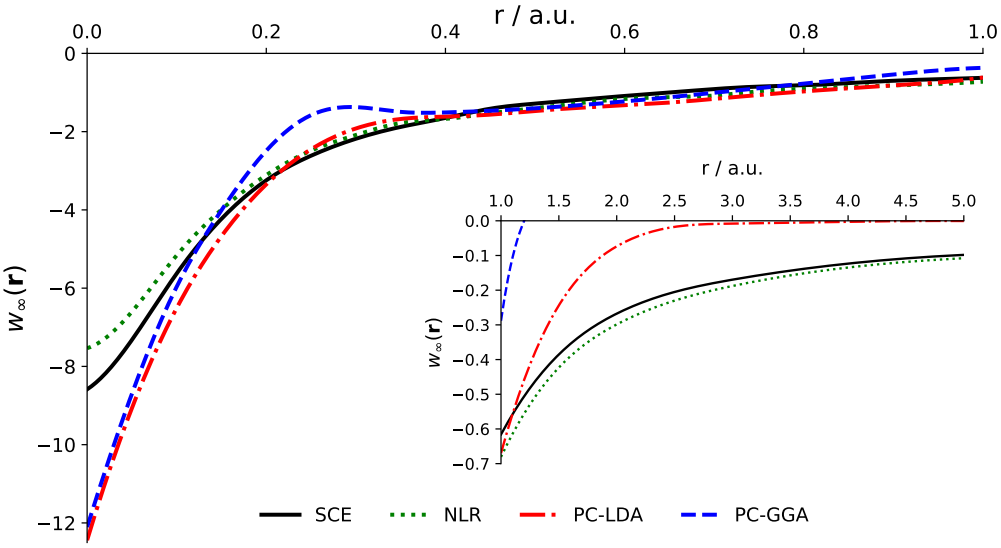

Figure 6.8: The $w_\infty(\mathbf{r})$ energy densities in the Ne^{6+} ion obtained with the following functionals: SCE, NLR, PC-LDA and PC-GGA.

Liu & Burke (LB)[158] and the Padé[1/1][181] model; each of the resulting energy densities are constructed from three local parameters, a, b and c, defined in the gauge of the xc hole. The functional forms of these three models are summarised in Table 6.3.

	$w_{\mathrm{xc},\lambda}(\mathbf{r})$	$a(\mathbf{r})$	$b(\mathbf{r})$	$c(\mathbf{r})$	Refs.
SPL	$a + \frac{b}{\sqrt{1+c\lambda}}$	$w_\infty(\mathbf{r})$	$w_\mathrm{x}(\mathbf{r}) - w_\infty(\mathbf{r})$	$-\frac{2w_0'(\mathbf{r})}{w_\mathrm{x}(\mathbf{r})-w_\infty(\mathbf{r})}$	155,156
LB	$a + b\left(\frac{1}{(1+c\lambda)^2} + \frac{1}{\sqrt{1+c\lambda}}\right)$	$w_\infty(\mathbf{r})$	$\frac{1}{2}\left(w_\mathrm{x}(\mathbf{r}) - w_\infty(\mathbf{r})\right)$	$-\frac{4w_0'(\mathbf{r})}{5(w_\mathrm{x}(\mathbf{r})-w_\infty(\mathbf{r}))}$	158
Padé[1/1]	$a + \frac{b\lambda}{1+c\lambda}$	$w_\mathrm{x}(\mathbf{r})$	$w_0'(\mathbf{r})$	$\frac{w_P(\mathbf{r})-w_\mathrm{x}(\mathbf{r})-w_0'(\mathbf{r})}{w_\mathrm{x}(\mathbf{r})-w_P(\mathbf{r})}$	181,183

Table 6.3: Functional forms of the local AC interpolation models (for the Padé[1/1], $p > 0$, $p \in \mathbb{R}$).

In addition to these, a local form of the two–legged representation[182] was constructed which, given some value of $w_1(\mathbf{r})$, takes the form

$$w_{\mathrm{xc},\lambda}(\mathbf{r}) = \begin{cases} w_\mathrm{x}(\mathbf{r}) + \lambda w_0'(\mathbf{r}) & \lambda \leqslant x_\lambda \\ w_1(\mathbf{r}) & \lambda > x_\lambda \end{cases} \tag{6.55a}$$

$$x_\lambda = \frac{w_1(\mathbf{r}) - w_\mathrm{x}(\mathbf{r})}{w_0'(\mathbf{r})}. \tag{6.55b}$$

In each of these four models, integration of $w_{\mathrm{xc},\lambda}(\mathbf{r})$ with respect to the coupling–constant gives the λ–averaged xc energy density $\bar{w}_\mathrm{xc}(\mathbf{r})$ which, if spatially integrated according to (5.10), yields the xc energy $E_\mathrm{xc}(\rho)$.

An important observation in the translation of global to local models is that, whilst the following global inequalities are always satisfied,

$$\mathcal{W}_0(\rho) \geqslant E_\mathrm{xc}(\rho) \geqslant \mathcal{W}_1(\rho) \geqslant \mathcal{W}_\infty(\rho), \tag{6.56}$$

their local counterparts do not necessarily satisfy these same inequalities. It has previously been observed for the Hooke's atom series that, in the tail regions of the density, $w_\infty(\mathbf{r})$ can be less negative than $w_1(\mathbf{r})$[134]. In this work, the crossing of $w_\infty(\mathbf{r})$ with $\bar{w}_\mathrm{xc}(\mathbf{r})$, $w_1(\mathbf{r})$ and $w_\mathrm{x}(\mathbf{r})$ has only been observed in the tail regions of the density and is thought to be an artefact of the numerical instability that occurs where the density is very small.

6.3.5 Performance Of Local AC Models

Having set–out the local AC interpolation models with exact & approximate forms of parameters required in their construction, the results of testing these models are to be presented here. In the first instance, these models will be

tested for using accurate input parameters to confirm that the translation of AC interpolation schemes from a global to a local form yields workable local models and to test their accuracy. The local interpolation schemes are further tested using various approximate input parameters considered in the previous subsection to examine how this affects the quality of the model, thus indicating their potential viability as new DFAs for use in general calculations.

6.3.5.1 Local Models With Accurate Parameters

Helium Isoelectronic Series

Although the Helium isoelectronic series are a set of only two–electron systems, it is a useful series to consider in evaluating the local interpolation models since most standard DFAs incorrectly characterise the hydride ion (H^-), failing to predict its existence as a bound electronic system[201,216]. Here, local interpolation models are constructed with input quantities in the form of energy densities acquired by Lieb optimisation at the FCI level, as described in subsection 5.2.3, for $0 \leqslant \lambda \leqslant 1$ and at $\lambda = \infty$ by evaluating the SCE functional on the $\lambda = 1$ density – also at the FCI level of theory.

Z	FCI	Local SPL	Global SPL	Local LB	Local Padé[1/1]
1	−0.0409	−0.0367	−0.0368	−0.0398	−0.0401
2	−0.0400	−0.0378	−0.0380	−0.0394	−0.0399
3	−0.0410	−0.0393	−0.0395	−0.0404	−0.0409
4	−0.0416	−0.0402	−0.0404	−0.0411	−0.0415
5	−0.0418	−0.0408	−0.0409	−0.0415	−0.0418
6	−0.0419	−0.0410	−0.0411	−0.0416	−0.0418
7	−0.0414	−0.0407	−0.0408	−0.0412	−0.0414
8	−0.0412	−0.0405	−0.0406	−0.0410	−0.0412
9	−0.0411	−0.0405	−0.0406	−0.0409	−0.0411
10	−0.0411	−0.0405	−0.0407	−0.0408	−0.0411

Table 6.4: Reference and interpolated E_c values, in Hartree, for the He isoelectronic series.

In Table 6.4, the correlation energies given by local forms of the SPL, LB and Padé[1/1] models (parameterised using $w_1(\mathbf{r})$) are given, along with that given by the global SPL model and the FCI correlation energy for comparison. This data shows that the local interpolation correlation energies are in close agreement with the FCI reference values; the mean absolute error (MAE) of the local interpolation models are 1.5 mH, 0.5 mH and 0.1 mH, for the SPL, LB and Padé[1/1] models respectively, with errors generally decreasing with Z.

As would be expected, the local Padé[1/1] is the most accurate of the models, given that it is derived from the full interacting energy density. This data further suggests that the local LB model is marginally superior to the local SPL model. However, comparing the global and local models shows a slightly lower error for the global model; the local SPL model has an MAE of 1.5 mH, compared to 1.3 mH for the global model.

Figure 6.9 compares the FCI $\bar{w}_c(\mathbf{r})$ with that of the local LB and SPL models, for the Helium atom. This reflects the numerical data in Table 6.4, both being very close to the FCI energy density but with slightly lower error in the LB model.

Beryllium Isoelectronic Series

The changes in correlation energy across the Beryllium isoelectronic series are somewhat more complicated than those in the Helium isoelectronic series, with its explanation involving the interplay of several effects. With increasing nuclear charge, the density becomes increasingly contracted, suggesting that the correlation energy should approach the high–density limit for very large Z. However, this is accompanied by a changing HOMO–LUMO gap, here the energy difference between $2s$ and $2p$ orbitals, which increases from $Z = 4 \rightarrow 13$ before decreasing with higher Z values[106].

Table 6.5 shows the reference and interpolated E_c results for the Be isoelectronic series, with Z in the range $4 - 10$. The wave function for λ values between 0 and 1 has been computed at in the same way as for the He isoelectronic series, however at the CCSD level of theory rather than FCI. As for the Helium series, the local Padé[1/1] that uses $w_1(\mathbf{r})$ is the most accurate of the local interpolation models. However, in contrast to the findings for He isoelectronic series, the local interpolation models are much more in–line with the reference data than the global models. For

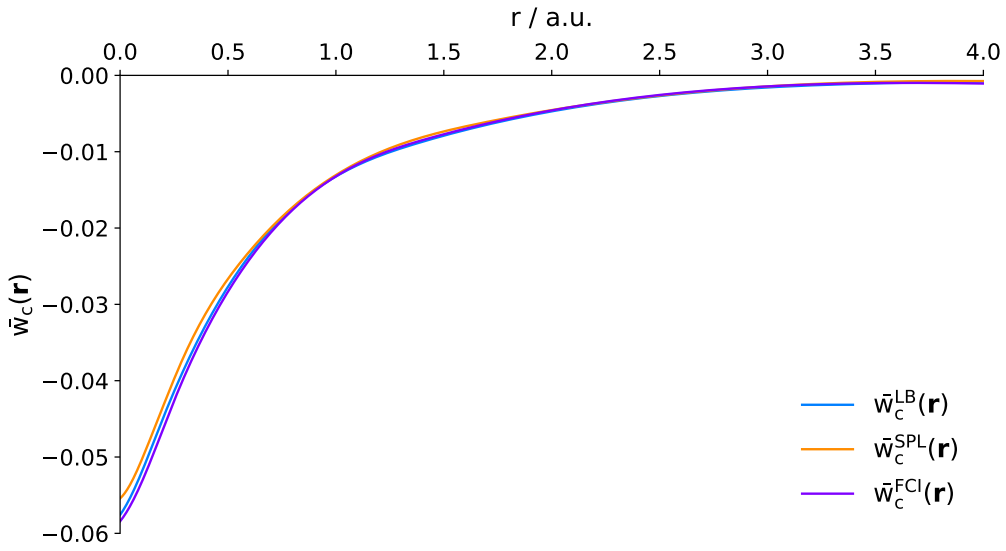

Figure 6.9: Plots of the FCI, local LB and local SPL λ–averaged correlation energy density in the Helium atom.

Z	CCSD	Local SPL	Global SPL	Local LB	Local Padé[1/1]
4	−0.0920	−0.0876	−0.1049	−0.0925	−0.0911
5	−0.1089	−0.1041	−0.1250	−0.1100	−0.1076
6	−0.1244	−0.1202	−0.1455	−0.1271	−0.1229
7	−0.1389	−0.1363	−0.1668	−0.1443	−0.1373
8	−0.1534	−0.1532	−0.1898	−0.1626	−0.1517
9	−0.1683	−0.1717	−0.2157	−0.1826	−0.1666
10	−0.1833	−0.1920	−0.2447	−0.2046	−0.1817

Table 6.5: Reference and interpolated E_c values, in Hartree, for the Be isoelectronic series.

example, in the case of F^{5+} the global SPL model has an MAE of 47.4 mH, whereas the error for the local SPL model is 3.5 mH.

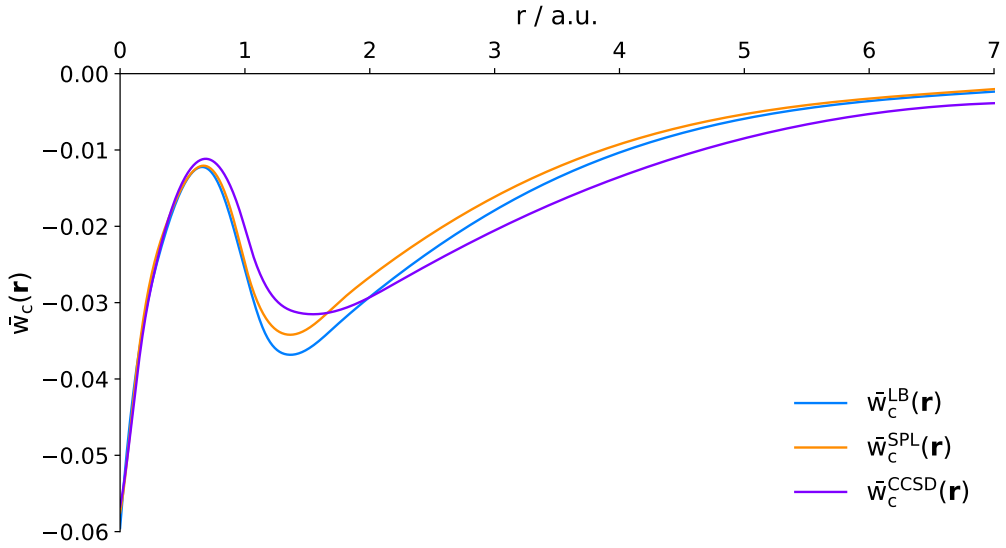

Figure 6.10: Plots of the CCSD, local LB and local SPL λ–averaged correlation energy density in the Beryllium atom.

Figure 6.10 shows the λ–averaged correlation energy densities for the Beryllium atom. The shape of $\bar{w}_c(\mathbf{r})$ reflects the shell structure of the Be atom[110,184]. The local SPL and LB interpolation models appear to qualitatively capture

the shell structure of $\bar{w}_c(\mathbf{r})$, however in some regions they overestimate the reference value whilst in other regions the converse is the case. The error cancellation that results from this is the most likely explanation for the superior accuracy of the local models in comparison to the global models.

Hydrogen Molecule

Despite the development of DFT into the most widely–applied electronic structure method, and the wealth of DFAs for the *xc* energy that have been developed, there are some systems for which no combination of DFAs provide an accurate description. A well–known example of such a system is the dissociating H_2 molecule[217]. Standard DFAs become increasingly inaccurate with greater H–H bond length, reflecting a fundamental flaw of DFAs in their inability to properly treat strong correlation.

The KS SCE approximation, proposed in Ref. 185, uses the SCE functional to approximate the Coulomb and *xc* energy and is equivalent to setting $\mathcal{W}_\lambda(\rho) = \mathcal{W}_\infty(\rho)$ for all λ. It has been seen previously[159,190] that KS SCE correctly predicts the dissociation of H_2 in a spin–restricted formalism, however at equilibrium geometry the energies it predicts are extremely low and the bond lengths predicted are overly short. The overall accuracy of KS-SCE for H_2 dissociation can be substantially improved by the addition of nonlocal corrections[159].

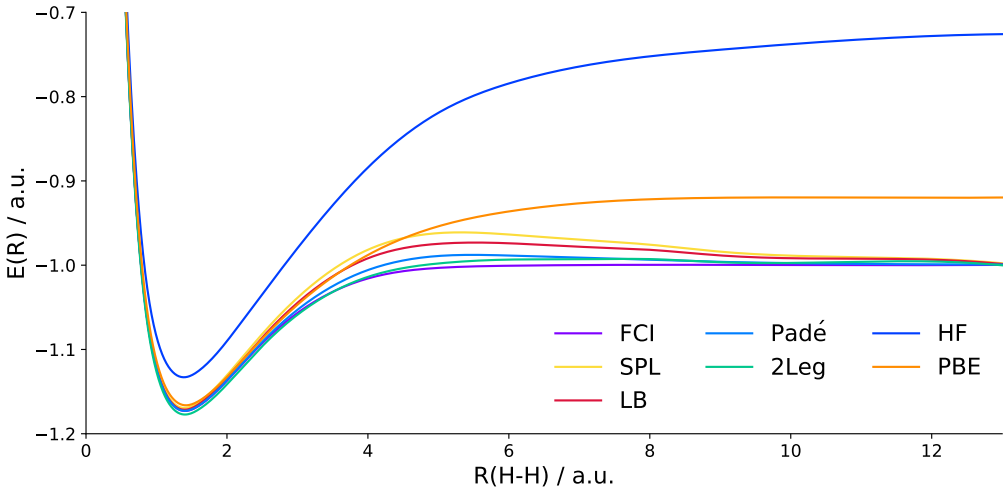

Figure 6.11: Potential energy curves for the H_2 molecule obtained using the local interpolation methods: SPL, Liu-Burke, two-legged representation combined with the Liu-Burke model, Padé[1/1] with $w_1(\mathbf{r})$. Restricted HF, PBE and FCI curves are also shown for comparison.

Figure 6.11 shows the dissociation curves for H_2 given by the local interpolation models, along with those given by HF, FCI and the PBE functional[92] for comparison. The computational details are the same as those of the He isoelectronic series, and the PBE, HF and FCI curves have been obtained from the DALTON quantum chemistry package[150], all within the uncontracted aug-cc-pCVTZ basis set[116]. The SCE energy density has been computed by using the dual Kantorovich method[159], described in subsection 6.3.2.

It can be seen in Figure 6.11 that all of the interpolation models correctly predict the dissociation of H_2, which follows from their inclusion of w_∞. In global AC models, at infinite separation the initial slope diverges as a result of the vanishing HOMO–LUMO gap, and the SPL and LB models reduce to $\mathcal{W}_\infty(\rho)$, yielding the exact energies. However, the dissociation curves produced by the local models approach the FCI curve slowly, resulting in an unphysical 'bump'–like feature. This is a well–known failing of DFT, having been observed with other functionals, such as the random–phase approximation[217] and even the global Padé[1/1] model with $\mathcal{W}_1(\rho)$[154]. It can be seen in Figure 6.11 that this is not remedied by the local interpolation approach, as the curve obtained by the local Padé[1/1] also exhibits this unphysical bump, as does that given by the local SPL model and, to a lesser extent, the local LB model.

Additionally, the two–legged representation of (6.55a) has been used to model the local AC, incorporating the interpolated $w_1(\mathbf{r})$ of the LB model. In regions of the dissociation curve where the unphysical bump is present, this model is more accurate than the local Padé[1/1] with $w_1(\mathbf{r})$. This may be understood by comparing the exact local AC data with the interpolated quantities. The left panel of Figure 6.12 shows the difference $\delta w(\mathbf{r})$ between $\bar{w}_c^{\mathrm{FCI}}(\mathbf{r})$ and that of each of the local interpolation models $\bar{w}_c^{\mathrm{model}}(\mathbf{r})$, along the H–H bond at the 5.0 a.u. geometry, as a function of the distance z from the bond midpoint. This difference is multiplied by the density to represent an energy per volume element. It shows that the local SPL energy density is the one that most overestimates the $\bar{w}_c(\mathbf{r})$. The

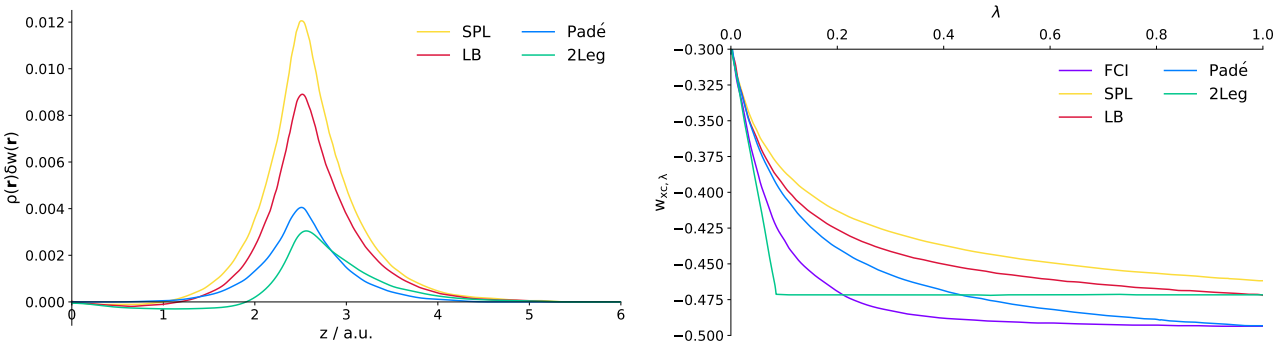

Figure 6.12: Plots of the difference between FCI and interpolated λ–averaged energy densities, $\delta w(\mathbf{r}) = \bar{w}_c^{\mathrm{FCI}}(\mathbf{r}) - \bar{w}_c^{\mathrm{model}}(\mathbf{r})$, with respect to the distance z from the bond midpoint (left), and the local AC curves at one of the nuclei for the FCI and local interpolation models (right), in H_2 with 5.0 a.u. bond length.

error is smaller for the LB model and even more so for the local Padé[1/1] model. The error is smallest with the two–legged model, obtained using the $w_1(\mathbf{r})$ of the local LB approximation. Furthermore, there is error cancellation in the two–legged model, as there are regions where the $\bar{w}_c(\mathbf{r})$ of this model underestimates $\bar{w}_{\mathrm{FCI}}(\mathbf{r})$.

It can also be seen that the curves shown in the left panel of Figure 6.12 have a maximum at the nucleus ($z = 2.5$). Focusing on this region, it appears that the FCI curve meets that of the Padé[1/1] at $\lambda = 1$, and that the two–legged representation curve also meets that of the LB model at $\lambda = 1$. This follows from the construction of the Padé[1/1] and two–legged curves from $w_1^{\mathrm{FCI}}(\mathbf{r})$ and $w_1^{\mathrm{LB}}(\mathbf{r})$ respectively. All curves, except for that of the two–legged model, lie above the FCI curve. In the case of the two–legged interpolation model, the first line segment is below the FCI curve, as a result of (6.55a) and the convexity of the given local AC curve. The second line segment that starts at $x_\lambda \sim 0.1$ is given by $w_1^{\mathrm{LB}}(\mathbf{r})$, and lies above the FCI curve. The resulting error cancellation makes it clear why the two–legged representation appears more accurate than the other models.

6.3.5.2 Local Models With Approximate Parameters

Following the previous discussion of local interpolation models constructed with accurate input parameters, an examination is now made of atomic correlation energies and H_2 dissociation curves computed using local interpolation models constructed from approximate input parameters, comparing the results obtained when using the SCE theory to model the strongly–interacting energy densities to those obtained using its NLR and PC approximations. With this aim, the accurate data for the non–interacting energy density and the local initial slope[215] employed previously are used in the present evaluation of interpolation functionals as this allows the effect of substituting the SCE energy density with an approximation to be isolated & explicitly observed.

In addition, the results of the local interpolation schemes are reported for the LiH dissociation curve, in this case using only the NLR approximation for the energy densities in the strongly–interacting limit, comparing and rationalising the performance of global and local interpolations.

As in the preceding work[215], Lieb optimisation calculations of the type described in subsection 5.2.3 were employed to compute accurate energy densities $w_{\mathrm{xc},\lambda}(\mathbf{r})$ for $\lambda = 0$ and for a range of $\lambda \ll 1$ to numerically approximate the initial slope $w_0'(\mathbf{r})$. This was carried out using the implementation of Refs. 111,112 in a development version of the DALTON quantum chemistry package[150], with the full configuration–interaction (FCI) and coupled–cluster singles and doubles (CCSD)[44] methods being used to compute $\mathcal{E}_\lambda(v)$. All calculations of two–electron systems and those of LiH were undertaken with a FCI wavefunction, whilst the remaining systems were treated at the CCSD level. Additionally, the uncontracted aug-cc-pCVTZ basis set was selected as the orbital and potential basis set for all systems excluding LiH, for which the uncontracted cc-pVDZ basis was used instead[116,218].

To compute SCE quantities for the atomic systems considered here, co–motion functions are obtained by using the conjectured SCE solution for spherically symmetric systems[156], yielding either exact or very accurate energy densities[197,219]. For the H_2 molecule, the SCE energy densities were computed via the dual Kantorovich SCE formulation[192,220], described in Ref. 159. The key quantity necessary in the evaluation of the NLR energy densities is the NLR radius $r_{\mathrm{NLR}}(\mathbf{r})$, which was computed using the method described in Ref. 210. The PBE[92] and FCI dissociation curves for the hydrogen molecule were calculated using the DALTON quantum chemistry package[150].

Atomic Correlation Energies

The correlation energies of several atomic/ionic systems computed with the SPL and LB interpolation schemes and with the SCE, NLR and PC models providing the $\lambda \to \infty$ quantities are presented for global interpolation in Table 6.6 and local interpolation in Table 6.7. Among these systems are two that are typically poorly described by contemporary DFAs; these are the H^- ion, generally not predicted to be bound[216] and the Ne^{6+} ion which belongs to the Beryllium isoelectronic series, a series exhibiting strong near–degeneracy effects with increasing nuclear charge Z[106,215].

It is evident from Tables 6.6 & 6.7 that local interpolation always yields a lower mean absolute error (MAE) than the corresponding global interpolation. Indeed, the least accurate local interpolation (the LB model using $w_\infty^{\mathrm{PC-GGA}}(\mathbf{r})$) has a lower MAE than the most accurate global interpolation (the SPL model using $\mathcal{W}_\infty^{\mathrm{NLR}}(\rho)$).

In Figure 6.13, the global AC (with exchange omitted) obtained by global and local interpolation with the SPL mode, using both the SCE and NLR input quantities, are presented for the Ne^{6+} ion. It can be seen that the curves pertaining to the local interpolation schemes are considerably closer to the reference AC (Lieb/CCSD) than those pertaining to global interpolation. This would indicate that the advantage conferred by a local interpolation approach over a global interpolation approach is significantly greater than any depreciation in accuracy resulting from approximating SCE quantities with those of the NLR model. Furthermore, it can be seen from Table 6.7 that the accuracy of these atomic correlation energies depends more on the interpolation model chosen than the accuracy of the $\lambda \to \infty$ quantities from which they are constructed.

Species	E_c^{ref}	SPL SCE	SPL NLR	SPL PC-GGA	SPL PC-LDA	LB SCE	LB NLR	LB PC-GGA	LB PC-LDA
H^-	-0.0409	-0.0368	-0.0352	-0.0360	-0.0415	-0.0399	-0.0383	-0.0391	-0.0444
He	-0.0400	-0.0381	-0.0371	-0.0376	-0.0401	-0.0396	-0.0388	-0.0392	-0.0413
Be	-0.0920	-0.1049	-0.1025	-0.1042	-0.1095	-0.0925	-0.1071	-0.1085	-0.1128
Ne^{6+}	-0.1833	-0.2447	-0.2399	-0.2435	-0.2527	-0.2526	-0.2487	-0.2517	-0.2590
Ne	-0.3470	-0.3940	-0.3840	-0.3940	-0.4010	-0.4050	-0.3970	-0.4050	-0.4100
Ar	-0.4040	-0.4880	-0.4810	-0.4880	-0.4910	-0.4940	-0.4890	-0.4940	-0.4960
MAE/mH	-	35	32	35	38	37	37	39	43

Table 6.6: The atomic (ionic) correlation energies, in Hartree, obtained by global SPL & LB interpolation with the exact (SCE) and approximate (NLR, PC-LDA & PC-GGA) $\mathcal{W}_\infty$ interpolation quantities.

Species	E_c^{ref}	SPL SCE	SPL NLR	SPL PC-GGA	SPL PC-LDA	LB SCE	LB NLR	LB PC-GGA	LB PC-LDA
H^-	-0.0409	-0.0367	-0.0344	-0.0364	-0.0416	-0.0398	-0.0375	-0.0393	-0.0444
He	-0.0400	-0.0378	-0.0370	-0.0347	-0.0393	-0.0394	-0.0388	-0.0359	-0.0405
Be	-0.0920	-0.0876	-0.0904	-0.0763	-0.0925	-0.1049	-0.0955	-0.0804	-0.0973
Ne^{6+}	-0.1833	-0.1919	-0.1997	-0.1653	-0.2036	-0.2045	-0.2124	-0.1760	-0.2156
Ne	-0.3470	-0.3830	-0.3720	-0.3770	-0.3870	-0.3960	-0.3860	-0.3900	-0.3980
Ar	-0.4040	-0.4450	-0.4360	-0.4350	-0.4510	-0.4590	-0.4510	-0.4940	-0.4640
MAE/mH	-	16	14	17	18	23	21	26	25

Table 6.7: The atomic (ionic) correlation energies, in Hartree, obtained by local SPL & LB interpolation with the exact (SCE) and approximate (NLR, PC-LDA & PC-GGA) w_∞ interpolation quantities.

In addition, it can also be seen that the MAE for the SPL model is smaller than that for the LB model, for both the global and local interpolation schemes. Interestingly, interpolating using the NLR approximation rather than the exact SCE form of $\mathcal{W}_\infty(\rho)$ and $w_\infty(\mathbf{r})$ results in a lower MAE. It has previously been observed that interpolation using SCE quantities often leads to an underestimation of atomic correlation energies[215] and that higher values of $\mathcal{W}_\infty(\rho)$ results in a higher globally interpolated correlation energy. In these systems, $\mathcal{W}_\infty^{\mathrm{NLR}}(\rho) \geqslant \mathcal{W}_\infty(\rho)$ and thus has the effect of partially offsetting the interpolation error. Whilst correlation energies obtained by global interpolation with SCE are a lower bound to those obtained with the NLR model in its place, this is not necessarily the case for the local correlation schemes. As shown in Table 6.7, for Be and Ne^{6+} the correlation energies obtained by interpolation with $w_\infty^{\mathrm{NLR}}(\mathbf{r})$ are lower than those obtained with the exact $w_\infty(\mathbf{r})$. This is reflected in the AC curves for Ne^{6+}, shown in Figure 6.13, in which that obtained by local interpolation with NLR parameters lies below that resulting from local interpolation with SCE parameters, despite the fact that $\mathcal{W}_\infty^{\mathrm{NLR}}(\rho) \sim -11.0\ H$ compared to $\mathcal{W}_\infty(\rho) \sim -11.5\ H$.

To rationalise these observations, it is useful to consider the corresponding energy densities; Figure 6.14 shows the λ–averaged correlation energy density in Ne^{6+}, multiplied by density and spherical volume element, obtained with both SPL and LB interpolation each using $w_\infty(\mathbf{r})$ and $w_\infty^{\mathrm{NLR}}(\mathbf{r})$ as input quantities, with the reference energy density for comparison. It appears from Figure 6.14 that the interpolation accuracy has a greater degree of dependence on interpolation model itself than on the accuracy of the $w_\infty(\mathbf{r})$ input parameter, as the energy densities exhibit a greater difference between SPL & LB than between SCE & NLR. It can also be seen that the two energy density curves obtained by interpolation with NLR quantities appear to be slightly below those obtained with the SCE quantities.

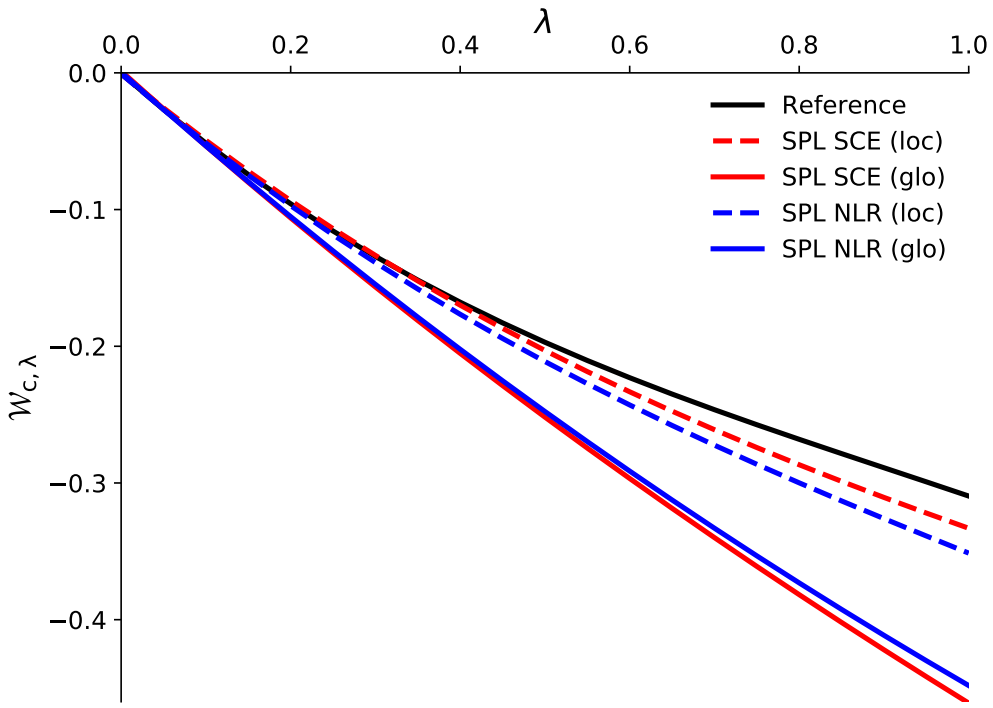

Figure 6.13: The global correlation AC curves for the Ne^{6+} ion obtained from the CCSD reference and from the global and local SPL interpolation model, constructed with interpolation parameters from the SCE model & its NLR approximation.

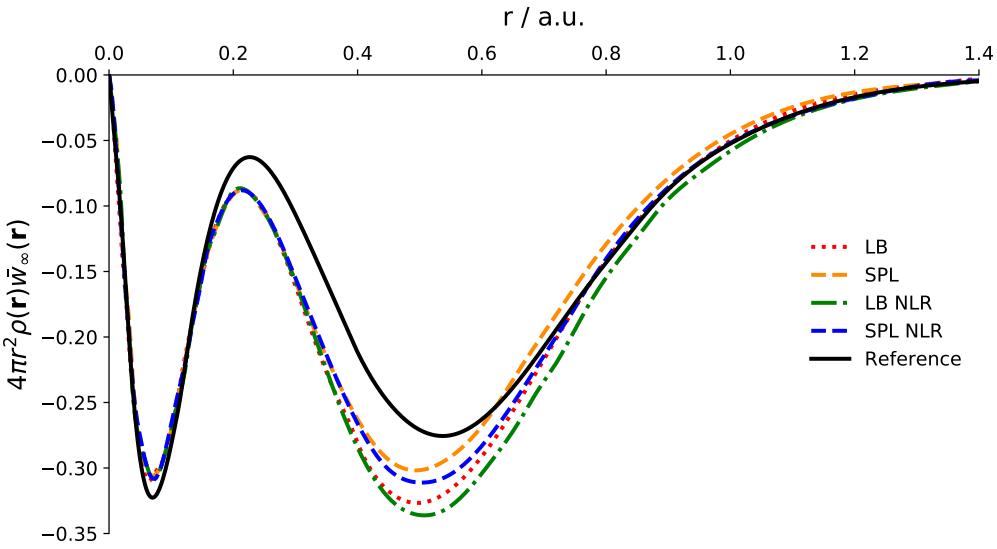

Figure 6.14: The λ–averaged correlation energy densities multiplied by the density and spherical volume element for the Ne^{6+} ion, obtained by both LB and SPL interpolation models and employing both the SCE and NLR $\lambda \to \infty$ energy densities, denoted SPL/LB and SPL NLR/LB NRL respectively. The reference curve has been computed at the CCSD level.

As described in relation to Figure 6.8, in the core region of Ne^{6+} ($r \lesssim 0.32$) $w_\infty^{\mathrm{NLR}}(\mathbf{r}) \geqslant w_\infty(\mathbf{r})$, whereas the opposite is generally true in the valence region. In the SPL and LB models, the sensitivity of the interpolated energy density to $w_\infty(\mathbf{r})$ is dependent on the magnitude of the local slope $w_0'(\mathbf{r})$; in regions where $w_0'(\mathbf{r}) \to 0$, $w_{\mathrm{xc},\lambda}(\mathbf{r})$ simply approaches the exchange energy density as the correlation energy density itself vanishes, hence the accuracy of $w_\infty(\mathbf{r})$ has a minimal effect on the interpolated $w_{\mathrm{xc},\lambda}(\mathbf{r})$. However, the converse is true in regions where $w_0'(\mathbf{r}) \to -\infty$, as the interpolated energy densities for both models would approach $w_\infty(\mathbf{r})$, making them highly sensitive to its accuracy. It can be seen in Figure 6.14 how that reflects on the interpolated energy densities. In the core region, the NLR based interpolated energy densities (labeled SPL NLR and LB NLR) are virtually indistinguishable from those that

are based on the SCE (labeled SPL and LB). Therefore, in the core region the interpolation model conceals the difference between $w_\infty^{\mathrm{NLR}}(\mathbf{r})$ and $w_\infty(\mathbf{r})$. In the valence region, where the local interpolation is much more sensitive to the changes in $w_\infty(\mathbf{r})$, it is apparent that the NLR–based interpolated energy densities are below the SCE–based ones. It is for this reason that, in the case of the Ne^{6+} ion, the NLR–based local interpolation gives a lower correlation energy than the SCE–based local interpolation, exemplifying an interesting difference between the global and local interpolation approaches.

The H_2 Dissociation Curve

Figure 6.15 displays the H_2 dissociation curves obtained by local interpolation with the two–leg and LB models, using both SCE and NLR input parameters, in comparison to those acquired with FCI and the PBE functional.

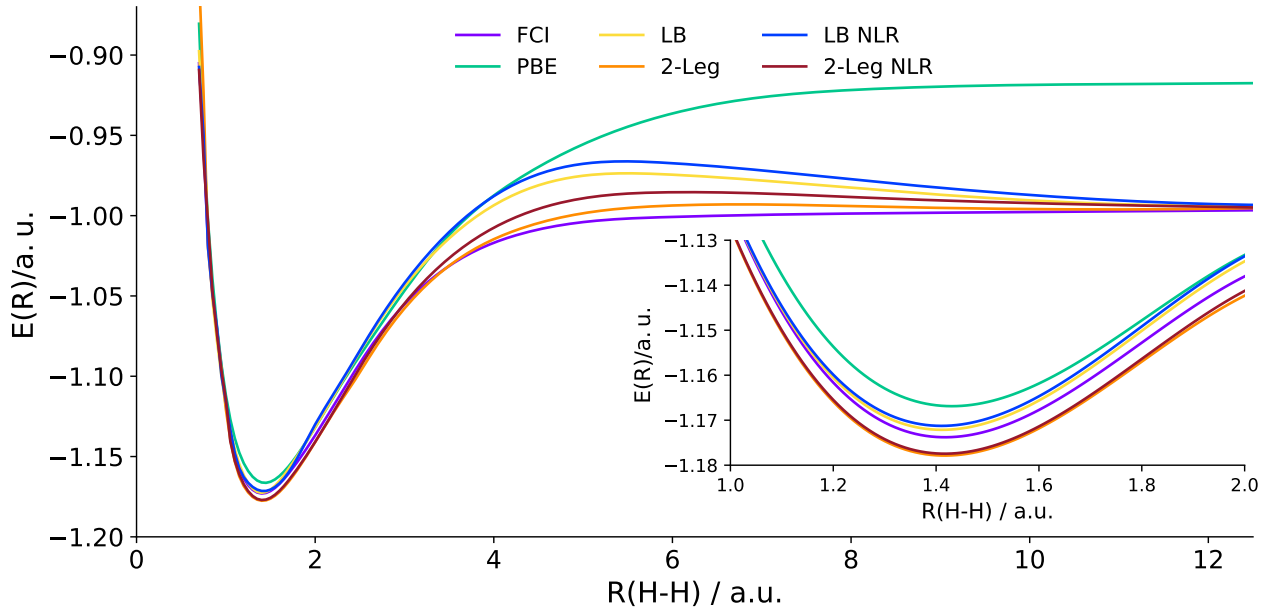

Figure 6.15: The H_2 dissociation curves obtained by the local LB and two-legged model interpolation with the SCE and NLR $w_\infty(\mathbf{r})$ interpolation input parameters, denoted LB/2-leg and LB NLR/2-leg NLR respectively. The PBE and FCI curves are shown for comparison.

It can be seen in Figure 6.15 that the dissociation curves given by the two interpolation models differ only slightly when the NLR model is used in place of the SCE model. Additionally, both models correctly predict the H_2 dissociation limit when using the NLR approximation, reflecting the underlying ability of the NLR model itself to dissociate the H_2 molecule correctly[210,215].

More pertinent however is the region of the dissociation curve at intermediate bond lengths for which DFT, combined with methods constrained to be exact in the dissociation limit, generally exhibits an unphysical 'bump'[215,217]. It was shown previously in subsubsection 6.3.5.1 that the local interpolation approach, even with exact input parameters, does not entirely eliminate this erroneous feature, although it is significantly attenuated with the local two–legged interpolation model[215]. In Figure 6.15, it can be seen that this feature is somewhat worsened when the NLR model is used in place of the SCE model for $w_\infty(\mathbf{r})$ as the input for the local interpolation scheme. In this region, the local initial slope is already large in magnitude (and increasing with bond length) thus there is a strong sensitivity of the interpolation models to the accuracy of $w_\infty(\mathbf{r})$. The approximate nature of the NLR model therefore has its most significant impact in this part of the dissociation curve. In contrast, approximating the SCE energy density with the NLR energy density in the interpolation models has considerably less effect on the energy computed at the equilibrium bond length, reflecting the lower sensitivity to the $\lambda \to \infty$ quantities. It is also noted that all of the local interpolation forms are more accurate than spin–restricted PBE at the equilibrium bond length. The dissociation curves computed using the SPL model are omitted from Figure 6.15 for clarity, however it is noted that their properties are broadly similar to those of the LB curves but with a more pronounced unphysical feature at intermediate bond lengths[215].

Figure 6.16 shows the H_2 dissociation curves obtained by interpolation using the PC model to approximate $w_\infty(\mathbf{r})$. As described in subsection 6.3.4, in the present work the PC model is examined at both the LDA level and at the GGA level, containing gradient–dependent terms. It appears from Figure 6.16 that the most accurate dissociation

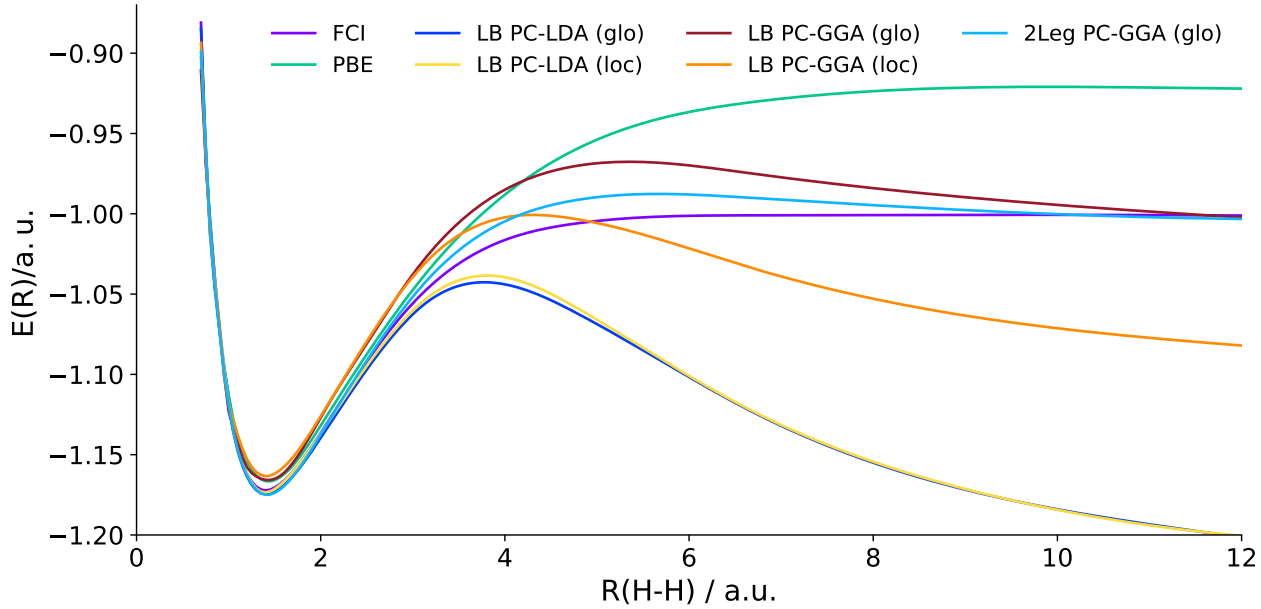

Figure 6.16: H_2 dissociation curves obtained by local (loc) and global (glo) LB and two-legged interpolation models using PC-LDA and PC-GGA input parameters. The PBE and FCI curves are shown for comparison.

curve obtained using PC model energy densities is the global two–legged model using PC-GGA to approximate $w_\infty(\mathbf{r})$. As observed in Figure 6.16, the two–legged model gives a more accurate dissociation curve than the LB interpolation model, however in this case the differences between interpolation models themselves are overshadowed by those arising from the use of different $\lambda \to \infty$ quantities. There is also a marked difference between the dissociation curves obtained by global and local LB interpolation using PC-GGA input quantities; whilst global interpolation yields a qualitatively accurate dissociation curve with only a small underestimation of the dissociation limit, local interpolation employing $w_\infty^{\mathrm{PC-GGA}}(\mathbf{r})$ yields a curve that becomes highly unphysical beyond the equilibrium geometry and results in an energy much lower than that of two hydrogen atoms in the dissociation limit. This is directly attributable to the superior performance of $\mathcal{W}_\infty^{\mathrm{PC-GGA}}(\rho)$ over $w_\infty^{\mathrm{PC-GGA}}(\mathbf{r})$ where, as seen in Figure 6.8, the PC-GGA energy density has erroneous long–range behaviour however global error cancellation results in a factitiously accurate $\mathcal{W}_\infty^{\mathrm{PC-GGA}}(\rho)$. In the local interpolation scheme, $w_{\mathrm{xc},\lambda}(\mathbf{r}) \to w_{\mathrm{x}}(\mathbf{r})$ where $w_\infty^{\mathrm{PC}}(\mathbf{r})$ crosses $w_{\mathrm{x}}(\mathbf{r})$ and as such there is no equivalent error cancellation for local interpolation.

The dissociation curves based on PC-LDA model input parameters are considerably poorer than those yielded by the PC-GGA model, both with a global and local interpolation scheme. For global interpolation, a quantitative comparison of their accuracies at the H_2 dissociation limit can be made by considering that the *xc* energy should cancel the Coulomb energy and that $\mathcal{W}_0'(\rho) \to -\infty$ in this limit, hence $E_{\mathrm{xc}}(\rho) - J(\rho) = \mathcal{W}_\infty(\rho) - J(\rho) = 0$ should be satisfied. For the infinitely stretched H_2, $\mathcal{W}_\infty^{\mathrm{PC}}(\rho)$ will be twice $\mathcal{W}_\infty^{\mathrm{PC}}(\rho)$ evaluated on the density of a hydrogen atom. Whilst this error is relatively small for PC-GGA, $\mathcal{W}_\infty^{\mathrm{PC-GGA}}(\rho) - J(\rho) = -0.3\ mH$, it is much larger for PC-LDA, $\mathcal{W}_\infty^{\mathrm{PC-LDA}}(\rho) - J(\rho) = 312\ mH$.

Whilst local interpolation using $w_\infty^{\mathrm{NLR}}(\mathbf{r})$ in place of the exact SCE quantities give reasonably accurate dissociation curves for H_2, those obtained with the computationally cheaper $w_\infty^{\mathrm{PC}}(\mathbf{r})$ appear volatile and unphysical and as such presently seem an inappropriate choice to substitute SCE energy densities in local interpolation models. The use of the PC model in global interpolation schemes appears to show more promise, yielding qualitatively accurate H_2 dissociation curves when using $\mathcal{W}_\infty^{\mathrm{PC-GGA}}(\rho)$[134]. The accuracy of global interpolations that include the PC model and the issues coming from the lack of size consistency have been recently investigated in Ref. 221.

The Lithium Hydride Dissociation Curve

Size consistency within the global interpolation models that used in this work is still preserved for systems that dissociate into equal fragments, assuming that the interpolation input quantities are size consistent themselves – which is not necessarily the case for the exchange energy and GL2 correlation energy in a spin-restricted framework, as discussed in Ref. 221. For this reason, it would be interesting to compare the performance of the local interpolation models against the global ones in the case of heterolytic dissociation. Figure 6.17 shows the dissociation curves

obtained by the local and global SPL interpolation and the reference (FCI) curve for comparison. For both the global and local interpolation schemes, the NLR approximation is used to provide the input quantities: $\mathcal{W}_\infty^{\text{NLR}}(\rho)$ and $w_\infty^{\text{NLR}}(\mathbf{r})$, respectively.

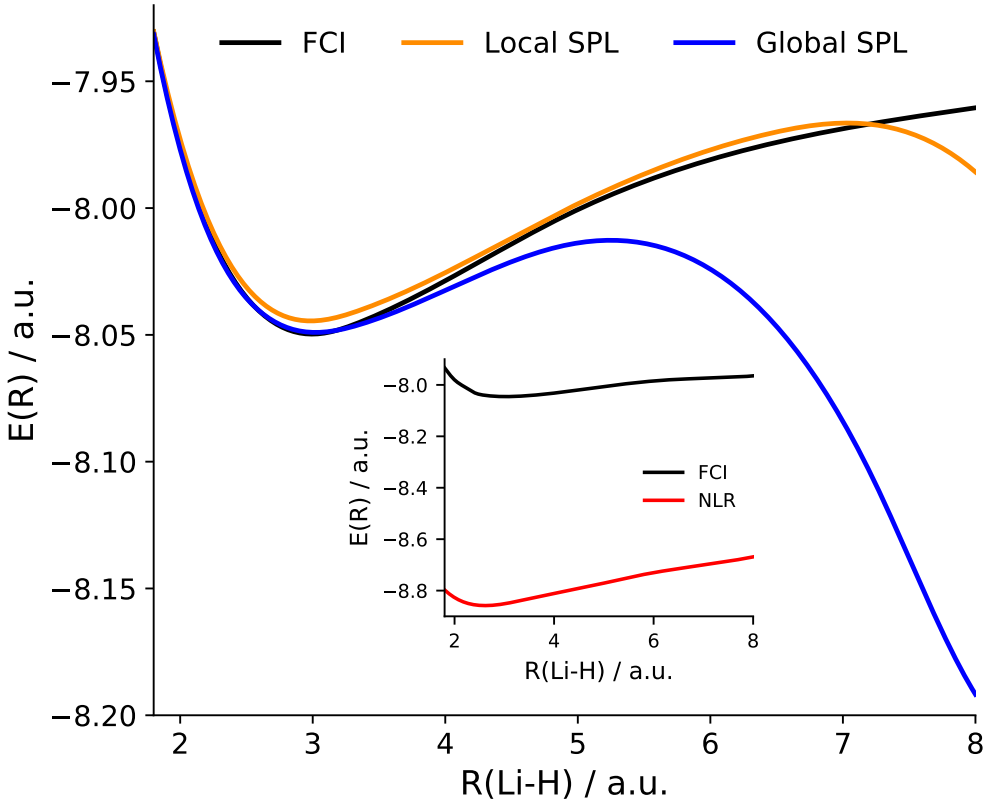

Figure 6.17: The LiH dissociation curves obtained by FCI, NLR and the global and local SPL interpolation with the NLR quantities.

It can be seen that the energies of the stretched LiH obtained by the global SPL interpolation are erroneously low. In the dissociation limit of LiH there is a step in the KS potential closer to the hydrogen, the more electronegative of the two atoms[222–224]. This step ensures that in the dissociation limit the atomic HOMO orbital energies are re–aligned and the molecule correctly dissociates into neutral atoms[222–224]. This is also why the KS HOMO-LUMO gap vanishes in the dissociation limit of LiH, as it happens in the Hydrogen molecule. As the gap closes and $\mathcal{W}_0'(\rho)$ diverges, the global SPL interpolation of $\mathcal{W}_\lambda(\rho)$ reduces to $\mathcal{W}_\infty^{\text{NLR}}(\rho)$. The energy of the latter are extremely low as can be seen from the inner panel of Figure 6.17, which shows the NLR dissociation curve corresponding to the approximation: $\mathcal{W}_\lambda(\rho) \approx \mathcal{W}_\infty^{\text{NLR}}(\rho)$.

In contrast to the global interpolation model, it can be observed that the locally interpolated energies are fairly concordant with the reference data up to a bond length of $R \sim 7$ a.u. However, as the HOMO-LUMO gap vanishes, the local slope will eventually become divergent everywhere[215], with the result that the locally–interpolated energies tend towards the NLR energies of the two fragments. The NLR energy is correct for the H atom (as NLR is exact for any one-electron density), but it over-correlates the Li atom. It is thus seen that the local slope is yet not sufficient to indicate the level of static correlation at a given point in space as it depends on the HOMO-LUMO gap, which is a global quantity. Other means of indicating the presence of strong correlation such as those recently proposed in Refs. 180,225,226, or the introduction of a local form of the HOMO-LUMO gap as described in Ref. 227 may provide a way of ameliorating the problem with the local slope; this may be considered in future work.

6.3.6 Correlation Indicator & Lower Bound To The Energy

In this subsection, the two–legged interpolation model between zero correlation and strong correlation is used to define and compute what may be considered a correlation type indicator and additionally to propose a lower bound to the exact energy tighter than those previously established[185].

6.3.6.1 Simple Correlation Type Indicator

In quantum chemistry, electron correlation is usually classified into the intuitive concepts of dynamical and nondynamical correlation, as introduced in Chapter 2 and further discussed in Chapter 5. Dynamical correlation is considered to be a short–range effect that is well treated by perturbation theories such as Møller–Plesset perturbation theory[29], which uses Hartree-Fock as a reference, or Görling-Levy perturbation theory[198], which uses the KS system as a reference. Nondynamical correlation however is associated with near–degeneracy effects, where a wavefunction comprised of multiple Slater determinants is necessary to properly describe the system. These are cases in which a single determinant reference is a poor starting point for perturbation theory. In terms of the AC, a system dominated by dynamical correlation has a λ dependence of $\mathcal{W}_\lambda(\rho)$ that is predominantly linear for λ between 0 and 1, while a system with substantial nondynamical correlation has a $\mathcal{W}_\lambda(\rho)$ with a much greater degree of curvature.

Burke, Ernzerhof and Perdew[182] previously observed that the point of intersection between the two line segments used in their two-legged interpolation model has either a minimal value, representing only nondynamical correlation, or a maximal value, representing only dynamical correlation. It is shown here that the point of intersection of the two line segments in a different but comparable two-legged interpolation model[215] is a simple parameter that indicates the correlation type; $x_{\text{corr}}(\mathbf{r})$ of (6.55b). The global equivalent to this local quantity, X_{corr} can be defined as

$$X_{\text{corr}} = \frac{\mathcal{W}_1 - \mathcal{W}_0}{\mathcal{W}_0'}. \tag{6.57}$$

Assuming the convexity of $\mathcal{W}_\lambda$ and $w_{\text{xc},\lambda}(\mathbf{r})$, X_{corr} and $x_{\text{corr}}(\mathbf{r})$ will have values between 0 and 1. If $X_{\text{corr}} = 1$, then the shape of the AC integrand will be linear for λ values between 0 and 1 and so the GL2 correlation energy captures all the correlation in the system. It can be said therefore that if $X_{\text{corr}} = 1$, then the correlation in the system is purely dynamical. However if $X_{\text{corr}} = 0$, then the AC curve is L-shaped and all the correlation present in the system is nondynamical.

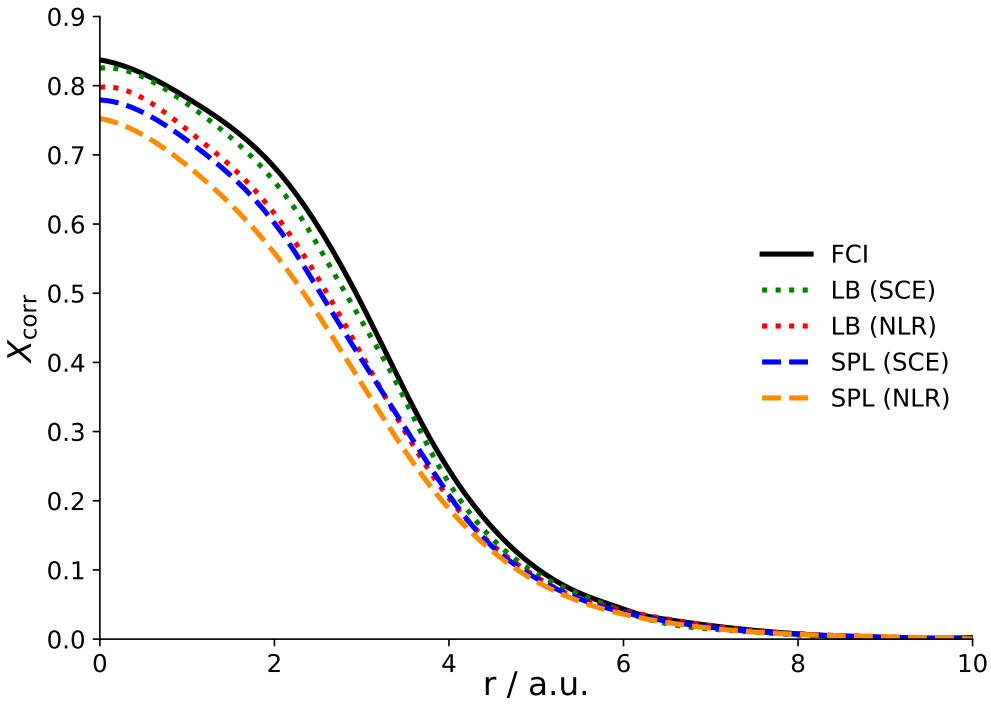

Figure 6.18: Plots of the X_{corr} correlation–type indicator of (6.57), as a function of the H_2 bond length.

In Figure 6.18, the reference X_{corr} for the hydrogen molecule as a function of the bond length is shown. At short bond lengths, the value of X_{corr} is relatively close to 1, reflecting the dominance of dynamical correlation in this region. Moreover, at short bond lengths the value of X_{corr} is close to that for the He atom, $X_{\text{corr}} = 0.84$. As the H_2 bond is extended, the value of X_{corr} decreases and finally drops to vanishes in the dissociation limit of H_2, at which point only nondynamical correlation is present.

By virtue of (6.57), the three input quantities required to compute X_{corr} are $\mathcal{W}_0(\rho)$, $\mathcal{W}_0'(\rho)$ and $\mathcal{W}_1(\rho)$. If $\mathcal{W}_1(\rho)$ is unavailable, it can instead be approximated by interpolation with the SPL or LB model using $\mathcal{W}_0(\rho)$, $\mathcal{W}_0'(\rho)$ and $\mathcal{W}_\infty(\rho)$ as inputs. In Figure 6.18, the reference X_{corr} and those constructed using approximations to $\mathcal{W}_1(\rho)$ are shown

for H_2, where the SPL and LB interpolation methods are employed using both the SCE and NLR–derived $\mathcal{W}_\infty(\rho)$ quantities. It can be seen in Figure 6.18 that all the approximate curves follow the trend of the reference X_{corr} curve. In this case, the LB interpolation model is more accurate than the SPL interpolation model. What can also be deduced is that the approximation of $\mathcal{W}_1(\rho)$ is more sensitive to the interpolation form than the accuracy of the $\mathcal{W}_\infty$ quantity; the LB X_{corr} curve based on $\mathcal{W}_\infty^{\text{NLR}}(\rho)$ is more accurate than the SPL X_{corr} curve constructed with the SCE–derived $\mathcal{W}_\infty(\rho)$.

Aside from indicating a type of correlation, the X_{corr} parameter may be useful to evaluate the accuracy of the GL2 approximation for a given system. The closer the X_{corr} value is to 0, the poorer the description of correlation in the system by this perturbation theory. A better starting point for the correlation description in this case would be the KS SCE theory, which gives unphysically low energies for the systems with an X_{corr} value close to 1.

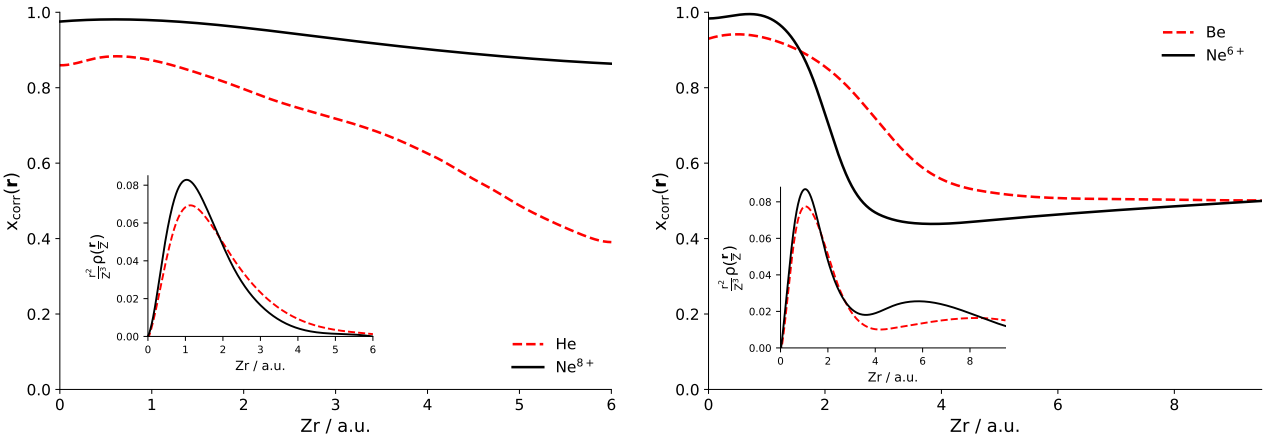

Figure 6.19: The $x_{\text{corr}}(r)$ local correlation–type indicator of (6.55b) shown for the He atom and Ne^{8+} ion (left) and for the Be atom and Ne^{6+} ion (right), where r is the radial distance from the nucleus.

Grimme and Hansen[228] have recently introduced a position–dependent indicator based on fractional orbital occupation, aiming to detect so–called molecular *hot regions* that have a high nondynamical correlation contribution. In the present work, $x_{\text{corr}}(\mathbf{r})$ may be interpreted as a local form of the correlation–type indicator X_{corr}. It may be easier to visualise $x_{\text{corr}}(\mathbf{r})$ than $w_{\text{xc},\lambda}(\mathbf{r})$ since the latter depends on both λ and $\mathbf{r}$.

In the left panel of Figure 6.19, $x_{\text{corr}}(r)$ is plotted for the He atom and the Ne^{8+} ion, both from the Helium isoelectronic series. It may be seen for the He atom, with global $X_{\text{corr}} = 0.84$, that $x_{\text{corr}}(r)$ decreases moving away from the nucleus, but in the energetically most important regions ($r \lesssim 2.0$ *a.u.*), $x_{\text{corr}}(r)$ is relatively close to 1. The $x_{\text{corr}}(r)$ curve for the Ne^{8+} ion is higher and flatter than that of He and the corresponding global $X_{\text{corr}} = 0.97$ – much closer to 1 than that for He. This indicates that dynamic correlation is more dominant in the Ne^{8+} ion than in the He atom; this behaviour is as expected for the Helium isoelectronic series with increasing nuclear charge[215]. It is important to note that in the asymptotic low-density regions, the local correlation–type indicator becomes highly sensitive to numerical errors since both numerator and denominator or (6.55b) become very small and should be treated with caution in these energetically unimportant regions.

In the right panel of Figure 6.19, $x_{\text{corr}}(r)$ is shown for the Be atom and the Ne^{6+} ion, both members of the Beryllium isoelectronic series. It is observed that these two $x_{\text{corr}}(r)$ curves exhibit an almost step–like structure, clearly distinguishing between the core region ($Z\ r \lesssim 4.0$ *a.u.* for Be & $Z\ r \lesssim 3.2$ *a.u.* for Ne^{6+}) and the valence region. In the core region, $x_{\text{corr}}(r)$ is close to 1.0 for Ne^{6+} ion and is around 0.9 for Be. This indicates the presence of almost purely dynamical correlation. The trend noted for the Helium isoelectronic series above, that increasing the value of Z value results in higher values of $x_{\text{corr}}(r)$, remains true for this series in the core region. However, in the valence region the value of $x_{\text{corr}}(r)$ for both systems is significantly lower, with values in the region of 0.5, indicating the presence of substantial nondynamical correlation – reflected by the pronounced curvature of the corresponding local AC curves[184].

Interestingly, the relative positions of the two $x_{\text{corr}}(r)$ curves in the core region are the opposite of those in the valence regions, where the curve of $x_{\text{corr}}(r)$ for Ne^{6+} crosses then lies below that of Be as distance from the nucleus increases. This observation suggests that, in the valence region of Ne^{6+}, there is a greater contribution of nondynamical correlation. This may be understood by consideration of the trend in the KS HOMO-LUMO gap for the Beryllium isoelectronic series with increasing nuclear charge[106,215]. Whilst the KS HOMO-LUMO gap (2*s*-2*p* gap) in Be is smaller (0.133 *H*) than that of Ne^{6+} (0.481 *H*), the absolute values of the Ne^{6+} orbital energies are much larger. Therefore, a more meaningful comparison may be made by defining a relative KS orbital energy gap, $\frac{\epsilon_{2s}-\epsilon_{2p}}{\epsilon_{2s}}$. The size

of this relative gap for Ne^{6+} is much smaller than that of Be, having values of 0.070 and 0.364 respectively. These KS energy gaps have been calculated with the Lieb optimisation at $\lambda = 0$, using the CCSD/aug-cc-pCVTZ level of theory to obtain $\mathcal{E}_0(v)$.

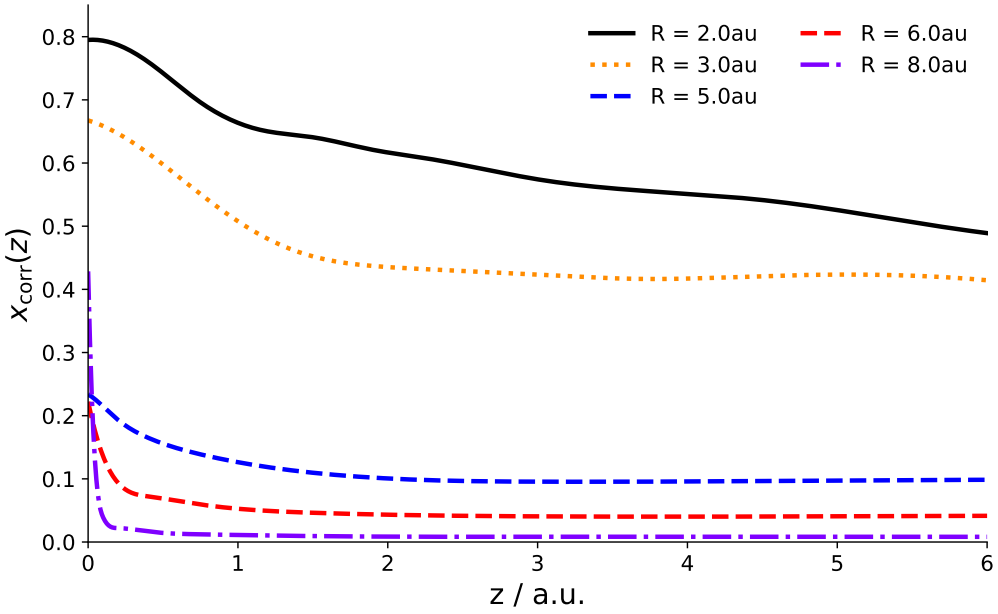

Figure 6.20: The $x_{corr}(z)$ local correlation-type indicator of (6.55b) shown for the H_2 molecule along the internuclear axis at different bond lengths, where z is the distance from the bond midpoint.

In Figure 6.20, the $x_{corr}(z)$ curves for the H_2 molecule along the internuclear axis for several bond lengths are plotted as a function of the distance from the bond midpoint, z. It is observed that, at the smaller bond length ($R = 2.0$) the structure of the $x_{corr}(z)$ curve is similar to that for the Helium atom. Furthermore, as the bond is extended, the structure of the curves can be seen to become increasingly linear. For the most stretched H_2, this indicates that *hot nondynamical correlation regions* are present at almost all points in space. Very small $x_{corr}(z)$ values indicate that, at all these points in space, the local AC has the characteristic L–shape[111,215]. The exception is the energetically unimportant bond midpoint of stretched H_2, at which the local AC becomes $w_{xc,0\leqslant\lambda\leqslant1}(z) = w_x(z)$ in the limit of dissociation, with $x_{corr}(z) = 1$ due to the antisymmetry of the correlation hole at the bond midpoint of H_2 in the limit $R \to \infty$[229].

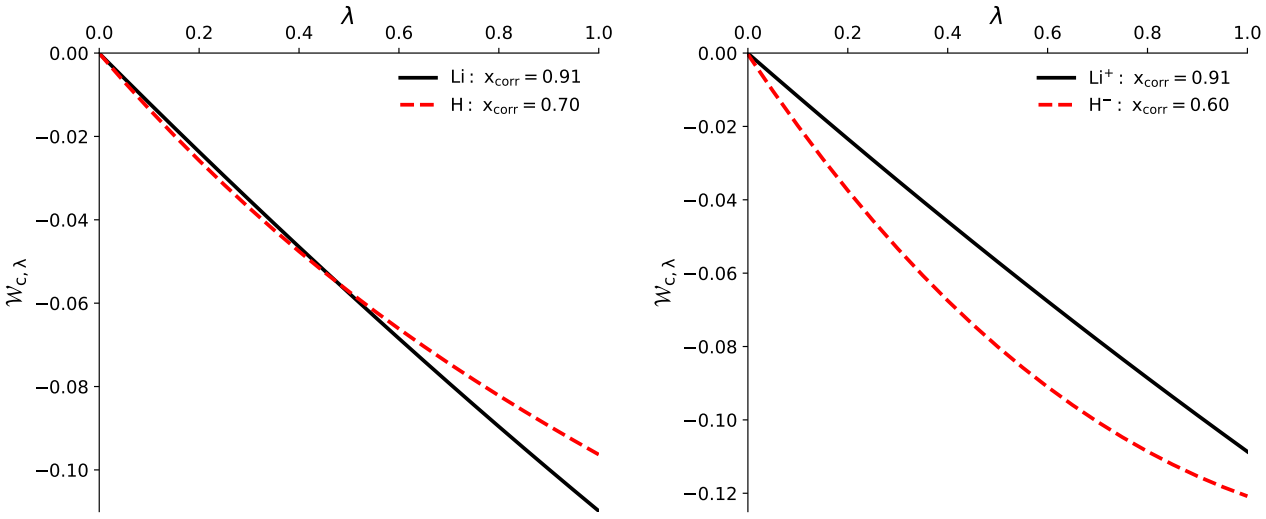

Figure 6.21: The local correlation AC curves at the two nuclei in LiH with bond length $R = 3.0$ a.u (left) and at the nuclei of H^- and Li^+ ions (right).

In the case of LiH, an interesting difference between shapes of the local AC curves can be observed at the Hydrogen and Lithium nuclei. The left panel of Figure 6.21 shows the correlation part of the local AC curves at both the

Hydrogen and Lithium nuclei in the LiH molecule at near-equilibrium geometry, $R = 3.0$ *a.u.* It can be seen that the curvature of the local AC at the Hydrogen nucleus is much more pronounced than that at the Lithium nucleus. As a result of the electronegativity difference between Lithium and Hydrogen, it would be expected that LiH at equilibrium would have significant ionic character; the Hydrogen atom would be expected to have slightly anionic character and the Lithium atom would be expected to have slightly cationic character. This is mirrored by the observation of the corresponding local AC integrands and associated $x_{\text{corr}}(r)$ values. From Figure 6.21 it can be seen that the local AC at the Li nucleus of LiH with bond length $R = 3.0$ *a.u.* is almost linear, as is the case for the local AC at the nucleus of Li^+ ion, shown in the right panel in Figure 6.21. The two corresponding $x_{\text{corr}}(r)$ are the same and both are close to 1. On the other hand, the pronounced curvature of the local AC at the H nucleus of LiH is similar to that in the H^- ion. The $x_{\text{corr}}(r)$ value at the H nucleus in LiH is 0.7, indicating the presence of nondynamical correlation in this region, but still somewhat less than that present at the nucleus of the H^- ion, where $x_{\text{corr}}(r) = 0.6$.

Both X_{corr} and $x_{\text{corr}}(r)$ quantities can be used in the context of interpolation along the AC. It may be the case that particular interpolation forms are better suited for a particular correlation regime than for the others. For instance, the SPL model appears to be superior for atoms with X_{corr} generally close to 1, while the LB and two-legged interpolation performed better than the SPL for the intermediate correlation regimes such as H_2 with a bond length of $R = 5.0$ *a.u.* It was shown in the present work that the quality of interpolation approximations can be more sensitive to the form of the interpolation model than to the exactness of the $\lambda \to \infty$ interpolation parameters. To tune the interpolation accuracy, a new *xc* functional may be constructed in which the correlation energies obtained by different interpolation schemes (or the correlation energy densities in the local interpolation variant) are mixed linearly. The linear mixing parameters could depend on X_{corr} and thus be system–dependent; e.g. for large X_{corr} the total correlation would have a higher portion of the SPL interpolated correlation, whilst for systems with a lower X_{corr} more of the correlation would be obtained by the LB or the two-legged interpolation models instead. These ideas may be pursued in future work.

6.3.6.2 Lower Bound To The Exact Energy

In wavefunction theory, the energies obtained by variational methods are always upper bounds to the exact ground state energy. The larger the space for the trial wavefunction, the closer the approximate energy is to the exact energy. This useful property is lost in KS DFT using DFAs, as the approximate energies resulting from this approach can be both higher and lower than the exact energies. It was shown in Ref. 185 that approximating the AC integrand with the single line segment $\mathcal{W}_\lambda(\rho) = \mathcal{W}_\infty(\rho)$ always yields a lower bound to the exact energy. In this approximation, referred to as *KS SCE*, the *xc* functional is simply given by $E_{\text{xc}}(\rho) = \mathcal{W}_\infty(\rho)$. The fact that the global AC curve is monotonically decreasing ensures that the KS SCE energies (obtained both self–consistently and by evaluation with the exact densities) are always lower than the exact energies. For systems in which nondynamical correlation is dominant, the KS SCE method yields reasonable energies[159,188,189], tending towards the exact energy in the low–density limit. In contrast, for systems with relatively little nondynamical correlation, the KS SCE energies are much lower than the exact energies[159,189,190]. In these scenarios, the KS SCE model still provides a lower bound to the exact energies – albeit a very loose bound. The KS SCE energies can be improved by adding corrections[159] or by using them as input parameters in an interpolation scheme[215] such as those discussed previously. In this case, the energies obtained may be substantially improved, however as for other DFAs, they may be higher or lower than the respective physical energies and can no longer be regarded as a lower bound.

In the present work, a method to tighten the lower bound given by the KS SCE energy for a given density is proposed by redefining the two-legged interpolation model, using $\mathcal{W}_0(\rho)$, $\mathcal{W}_0'(\rho)$ and $\mathcal{W}_\infty(\rho)$ as inputs,

$$\mathcal{W}_\lambda = \begin{cases} \mathcal{W}_0 + \lambda \mathcal{W}_0' & \lambda \leqslant X_{\text{corr}}^{\text{SCE}} \\ \mathcal{W}_\infty & \lambda > X_{\text{corr}}^{\text{SCE}} \end{cases} \tag{6.58}$$

where,

$$X_{\text{corr}}^{\text{SCE}} = \frac{\mathcal{W}_\infty - \mathcal{W}_0}{\mathcal{W}_0'}. \tag{6.59}$$

This model is referred to here as the 2-leg SCE interpolation model. For $\mathcal{W}_\lambda$ of (6.58) to be a rigorous lower bound to the exact $\mathcal{W}_\lambda$, two conditions have to be satisfied. The first is the monotonically decreasing nature of $\mathcal{W}_\lambda$[112,230], which ensures that the second segment of (6.58) lies below the curve of the exact integrand, $\mathcal{W}_\infty \leqslant \mathcal{W}_\lambda$. The second condition is the convexity of the $\mathcal{W}_\lambda$ integrand. If $\mathcal{W}_\lambda$ as function of λ is convex, then $\mathcal{W}_0 + \lambda \mathcal{W}_0' \leqslant \mathcal{W}_\lambda$.

The convexity of $\mathcal{W}_\lambda$ is often assumed to be true, but it can be violated in the case of phase transitions along the AC path. In these cases, the AC curve could have discontinuities and would then be only piecewise convex. For example if the physical $\lambda = 1$ system is the uniform electron gas at a density lower than that at which the ferromagnetic

transition occurs, then the AC could contain a discontinuity. Phase transitions that occur when the external potential is changed smoothly, described in Refs. 231–235, are usually accompanied by a change in the density. However, as set–out in Section 3.6, this work assumes the density-fixed adiabatic connection, for which the density constraint is expected to reduce to occurrence of phase transition discontinuities. It is further noted if $\mathcal{W}_\lambda$ is only piecewise convex, then the X_{corr} indicator defined in the previous section may be greater than 1. In most cases, the density constraint on a chemical system should prevent this from occurring, although exceptions may still occur. Besides being monotonically decreasing, it is also known that $\mathcal{W}_\lambda$ is bounded from below by the *Lieb–Oxford inequality*[84]

$$\mathcal{W}_\lambda(\rho) \geqslant -C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r}, \tag{6.60}$$

where C_{LO} is a constant, rigorously shown to be between 1.4119[236] and 1.6358[237]. In addition to the 2-leg SCE lower bound, another lower bound can be constructed by replacing $\mathcal{W}_\infty(\rho)$ appearing in (6.58) with $-C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r}$,

$$\mathcal{W}_\lambda = \begin{cases} \mathcal{W}_0 + \lambda \mathcal{W}_0' & \lambda \leqslant X_{\text{corr}}^{\text{LO}} \\ -C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r} & \lambda > X_{\text{corr}}^{\text{LO}} \end{cases} \tag{6.61}$$

where,

$$X_{\text{corr}}^{\text{LO}} = \frac{-C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r} - \mathcal{W}_0}{\mathcal{W}_0'}. \tag{6.62}$$

This construct is referred to as the 2-leg LO interpolation model. Since $\mathcal{W}_\infty(\rho) \leqslant -C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r}$, correlation energies from a 2-leg LO interpolation would also be lower bounds to the exact correlation energies, assuming convexity of the global AC integrand. In this work the LO constant takes the value $C_{\text{LO}} = 1.4174$, which was obtained by explicitly computing the indirect energies of uniform electron spheres using the SCE methodology and extrapolating the value of C_{LO} in the $N \to \infty$ limit[236]. The resulting value is close to the lowest rigorously known value $C_{\text{LO}} = 1.4119$[236] and is also lower than that previously believed to its lower bound $C_{\text{LO}} = 1.444$ – obtained from the total energy of a base centred cubic crystal of the uniform electron gas. Lewin and Lieb have recently shown that this value does not correspond to an indirect energy[238].

Species	$E_{\text{c}}^{\text{ref}}$	KS SCE	LO	2-leg SCE	2-leg LO
H^-	-0.0409	-0.1860	-0.2662	-0.0571	-0.0571
He	-0.0400	-0.4733	-0.6689	-0.0450	-0.0450
Be	-0.0920	-1.3464	-1.7775	-0.1234	-0.1234
$Ne^{6}+$	-0.1833	-3.9666	-5.1464	-0.2779	-0.2779
Ne	-0.3470	-7.9160	-9.0470	-0.4370	-0.4370
Ar	-0.4040	-20.944	-23.272	-0.5110	-0.5110
H_2 $R = 1.4$	-0.0400	-0.3020	-0.4310	-0.0490	-0.0490
H_2 $R = 2.8$	-0.0680	-0.2250	-0.3580	-0.1020	-0.1020
H_2 $R = 5.0$	-0.1840	-0.2340	-0.4010	-0.2210	-0.3610
H_2 $R = 7.0$	-0.2340	-0.2500	-0.4320	-0.2480	-0.4250
H_2 $R = 9.0$	-0.2560	-0.2620	-0.4480	-0.2620	-0.4460

Table 6.8: The correlation energy obtained from the KS SCE model, LO bound, 2-leg SCE model and 2-leg LO model compared with the reference value.

Table 6.8 shows the CCSD reference correlation energies, the KS SCE correlation energies (i.e, $\mathcal{W}_\infty(\rho) - \mathcal{W}_0(\rho)$), the correlation energies obtained from the LO bound (i.e., $-C_{\text{LO}} \int \rho^{4/3}(\mathbf{r})\,\mathrm{d}\mathbf{r} - \mathcal{W}_0(\rho)$), together with the correlation energies obtained by the 2-leg SCE and 2-leg LO interpolation schemes. It can be seen that already the KS SCE correlation energies are significantly above those computed using the LO bound in its place. Furthermore, it is apparent that the 2-leg SCE substantially tightens the lower bound of KS SCE, by up to an order of magnitude in some cases. Whilst the accuracy of the 2-leg SCE is not as high as that of other interpolation models presented in this work, it does have the advantage that it recovers the KS SCE property of yielding correlation energies that are a lower bound to the physical counterparts.

When $X_{\text{corr}}^{\text{SCE}} \geqslant 1$, the correlation energy obtained from the interpolated $\mathcal{W}_\lambda(\rho)$ of (6.58) becomes $\mathcal{W}_0'(\rho)/2 = E_{\text{c}}^{\text{GL2}}(\rho)$. This is true for the atoms given in Table 6.8 and is why both 2-leg SCE and 2-leg LO yield the same correlation energies for the given atoms. However, the role of $\mathcal{W}_\infty(\rho)$ in the 2-leg SCE interpolation is to correct $E_{\text{c}}^{\text{GL2}}(\rho)$ energy if it becomes too low, as happens for example in the stretched H_2 molecule with $R = 10.0$ *a.u.* (where $X_{\text{corr}}^{\text{SCE}} \ll 1$), for which the GL2 correlation energy is too low $E_{\text{c}}^{\text{GL2}}(\rho) \sim -80$ *H* but the correlation energy of the 2-leg SCE interpolation model is much more accurate.

It is apparent that the 2-leg SCE and the other interpolation schemes employed here benefit from the complementary information provided by $\mathcal{W}_\infty(\rho)$ and $\mathcal{W}'_0(\rho)$. For the stretched H_2, the 2-leg SCE correlation energies are much closer to the physical ones than those of the 2-leg LO interpolation. The latter could be tightened if instead of using $C = 1.41$ the value $C = 1.21$ was used; this should be the optimal constant for $N = 2^{236}$, but would lead to an overall loss of generality of the model with respect to the number of electrons.

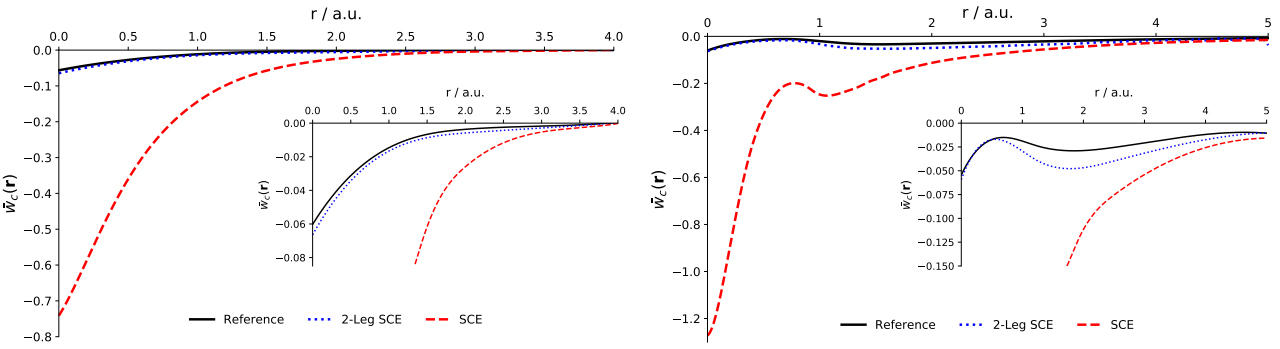

Figure 6.22: The λ–averaged correlation energy density $\bar{w}_c(\mathbf{r})$ obtained by the 2-leg SCE interpolation (the local variant of (6.58)) shown together with the reference $\bar{w}_c(\mathbf{r})$ and the SCE correlation energy density $\bar{w}_\infty(\mathbf{r})$ for the He atom (left, with FCI reference) and Be atom (right, with CCSD reference).

It is instructive to examine whether the energy densities obtained by the local variant of (6.58) yield a lower bound to $w_{xc,\lambda}(\mathbf{r})$ and thus also a lower bound to $\bar{w}_{xc}(\mathbf{r})$. For this to be the case, two conditions must be satisfied; these are that $w_{xc,\lambda}(\mathbf{r})$ is both monotonically decreasing and is convex with respect to λ. Numerical results for Coulomb systems suggest that both conditions are satisfied [184,215]. The known counterexample to the monotonicity of $w_{xc,\lambda}(\mathbf{r})$ is a system in which the external potential is not Coulombic; it occurs in the Hooke's atom series, where in the energetically unimportant region, $w_\infty(\mathbf{r})$ can be above $w_1(\mathbf{r})$ [134].

In Figure 6.22, the λ–averaged correlation energy densities $\bar{w}_c(\mathbf{r})$ obtained by the local variant of the 2-leg SCE interpolation, together with the reference $\bar{w}_c(\mathbf{r})$ and that corresponding to the $\lambda \to \infty$ limit $w_{c,\infty}(\mathbf{r}) = w_\infty(\mathbf{r}) - w_x(\mathbf{r})$ for the He atom (left panel) and the Be atom (right panel). In both cases, the SCE correlation energy densities are a lower bound to the reference $\bar{w}_c(\mathbf{r})$, but clearly this bound is loose, particularly in the region near the nuclei. With the 2-leg SCE local interpolation method, the local bound appears substantially tightened. No attempt is made here to construct a local variant of the 2-leg LO bound, as the rigorous local variant of the LO inequality that binds the energy density in the gauge of the *xc* hole is not known [134].

6.4 Conclusion

This Chapter reports on different approaches that may be taken to model the λ–averaged energy densities presented in Chapter 5. The first approach, discussed in Section 6.2, was that of modelling the correlation hole, then reconstructing an an approximation to the λ–averaged correlation energy density from this model. The approximated energy densities appeared to recover the main features of the accurate reference energy densities, such as the inter–shell peaks, however the total integrated correlation energy was poor. Tuning the model to obtain the most accurate correlation energy possible resulted in an energy density much less resemblant of the reference energy density, suggesting that the improvement in the correlation energy was at the expense of much more error cancellation in the integral of the energy density.

The second approach considered in this work, presented in Section 6.3, was that of interpolating the local AC using local quantities defined in the gauge of the electrostatic potential of the *xc* hole. In order to obtain approximations to the local AC over the physical regime ($0 \leqslant \lambda \leqslant 1$), interpolation models were constructed between the non–interacting and strongly–interacting limits of DFT. The non–interacting energy densities were obtained using the Lieb optimisation with respect to an *ab initio* reference, whilst the energy densities in the strongly–interacting limit were obtained using the strictly–correlated electron (SCE) approach. The inclusion of the SCE information in DFAs helps ensure their ability to capture strong correlation effects.

Unlike previous attempts in this direction that used global input quantities to interpolate the AC, the local approach is more amenable to the construction of approximations that do not violate size consistency, at least in the usual DFT sense [163]. This approach shows promise for the accurate treatment of strong correlation effects whilst potentially maintaining size consistency. In this work, highly accurate solutions to the SCE problem for atomic and diatomic

systems were employed, allowing for assessment of the local interpolations without spurious (or fortuitous) error cancellations. In addition to this, approximations to the computationally expensive SCE energy densities were discussed and their effects on the quality the local interpolation schemes considered. In this work, several different models for the SCE energy densities in the gauge of the *xc* hole were employed; these were the nonlocal radius (NLR) functional and the point-charge plus continuum (PC) functional. In addition to approximations for the strongly–interacting limit of the local AC, an approximate expression for the local slope of the AC in terms of occupied and unoccupied KS orbitals was derived; results obtained using this approximation will be the subject of future work.

The results of interpolating the local AC for the He and Be isoelectronic series and the Hydrogen molecule, using exact energy densities in the gauge of the electrostatic potential of the *xc* hole, are presented in subsection 6.3.5. Local interpolation models using an approximation for the strongly–interacting limit were then tested for atomic systems, together with the Hydrogen molecule and the Lithium hydride dissociation curves. These tests showed that the NLR energy densities are an excellent alternative to the SCE energy densities for use in local interpolation schemes. The energy densities obtained with the PC model are computationally simple to construct, but the overall performance of the local interpolation based on the PC model was not satisfactory; they are adequate for atoms but they introduce a large error when tested on molecular dissociation curves. The global interpolation based on the PC-GGA model seems to be more promising despite the size consistency issues with global AC interpolation functionals. Extensive testing of the PC model in the context of global interpolation has been recently carried on in Ref. 221, where some of the limitations of global interpolation models have been carefully studied.

These interpolation models have also been used to propose a correlation–type indicator and also a lower bound to the exact *xc* energy. These will be considered further in future work to see if they can be used to improve the accuracy of the local interpolation functionals. In addition, the recently proposed model of Bahmann, Zhou, and Ernzerhof[214] may be examined as it should in principle provide results superior to the NLR approximation.

One further concept arising from this Chapter which may be the subject of future study is the possibility of some form of synthesis between an MCS–type functional and a local interpolation–type functional. It would appear thus far that the two different approaches to reconstructing the correlation energy density somewhat complement each other; the MCS functional shows promise in higher density regions closer to the nucleus but are poor in the asymptotic region, whilst local interpolation models appear to perform relatively well in the limit of low density but can be significantly erroneous in higher density regions. Since all models here are defined in the gauge of the electrostatic potential of the *xc* hole, a synthesis of the two approaches should be straightforwardly possible.

7 Quantum Chemistry In A Magnetic Field

In the material studied thus far, approaches to understanding & modelling the solutions to the non–relativistic electronic Schrödinger Equation have been examined; the only interaction considered other than electron–electron interaction is that between the electrons and the external potential pertaining to the nuclear charges in a molecular system. Whilst this is appropriate for problems in which only the ground–state energy is sought, it provides no means by which the *response* of the molecule to changes in its environment can be accounted for. In the present Chapter, the treatment of external magnetic fields is considered by the inclusion of field–dependent terms in the electronic Hamiltonian – thus allowing the behaviour of a molecule in an external magnetic field to be computed.

For the present discussion, the external magnetic field $\mathbf{B}$ will be given in atomic units unless specified otherwise; in terms of the SI unit for magnetic field strength – the Tesla (T), $1\,\text{a.u.} = 235'000\,\text{T}$. The atomic unit of magnetic field strength is therefore extremely large in magnitude when, by comparison, some of the strongest magnetic fields created in laboratories are in the order of 100 T. Magnetic fields with strengths approaching 1 a.u. are only known to exist on celestial bodies such as white–dwarf stars, whilst magnetic fields with strengths several orders of magnitude higher are located on neutron stars & pulsars.

The presence of an external magnetic field can affect the chemistry of molecular systems, however the way in which this occurs is not necessarily the same for all magnetic field strengths. The strength of an external magnetic field and its effects on the chemistry of molecular systems can be characterised in general terms by the three regimes,

(i) *Coulomb regime*, $|\mathbf{B}| \ll 1\,\text{a.u.}$: this regime describes the standard conditions for systems present on Earth, in which the effects of an external magnetic field are small relative to Coulomb interactions so can be treated perturbatively and which results in a familiar chemistry being exhibited,

(ii) *Intermediate regime*, $|\mathbf{B}| \approx 1\,\text{a.u.}$: this regime describes conditions on celestial bodies such as white dwarf stars, where magnetic field strengths are much greater in magnitude than those present on Earth to the extent that their interaction with electrons is comparable in magnitude to Coulomb interactions, resulting in an exotic chemistry,

(iii) *Landau regime*, $|\mathbf{B}| \gg 1\,\text{a.u.}$: this regime describes conditions on celestial bodies such as neutron stars and pulsars, where magnetic field strengths can reach hundreds or even thousands of atomic units and as a result are the dominant form of interaction effecting electrons, permitting Coulomb interaction to be treated perturbatively and leading to an alien chemistry being exhibited.

Methods for the treatment of an external magnetic field in the Coulomb regime are well–established [239–241] and are readily employed to compute molecular properties from their response to a magnetic perturbation. However, this is not the case when considering an external magnetic field in the intermediate regime, in which the magnetic field cannot be treated perturbatively; finite–field approximations may be used for this purpose however the convergence of the energy in basis set size is extremely slow, precluding calculations on general systems [242–245].

In this work the alternative approach of Ref. 172 is considered, in which self–consistent field calculations are conducted using atomic orbitals that respond to an external magnetic field in a manner that is correct to first order in the field strength [246–249]. This non–perturbative approach permits the examination of general systems in the presence of strong magnetic fields since the convergence of the energy with basis set size is much more rapid than that for finite–field approximations. Additionally, this presents an opportunity to study systems for which the response to the external magnetic field is highly nonlinear and cannot be approximated by perturbative methods. A brief overview of the theory underlying this approach is given in the present chapter, whilst the main technical challenge in its practical implementation is addressed in Chapter 8.

7.1 The Molecular Schrödinger Equation In An Electromagnetic Field

The interaction between electrons and an electromagnetic field, whilst being pertinent to the quantum mechanical description of the molecular system, is typically rationalised using the arguments of classical mechanics. The Hamiltonian formulation of Newton's equations of motion allows the transposition of the classical argument to quantum mechanics to be most readily seen and will be employed in subsection 7.1.1 below. It is important to note that the spin of an electron, described in subsection 1.2.1, substantially affects the interaction between an electron and an electromagnetic field. The form of this interaction was shown by Dirac as part of the relativistic treatment of

electrons[250]. Relativistic effects are not considered in the present work and as such the interaction between electron spin and field resulting from the non–relativistic reduction of the Dirac equation will be assumed.

7.1.1 Interaction Of A Charged Particle & Electromagnetic Field

In Newtonian mechanics, the motion of a particle with mass m can be described by the *equation of motion*

$$\mathbf{F}(\mathbf{r}, \mathbf{v}, t) = m\mathbf{a}, \tag{7.1}$$

in which $\mathbf{F}$ is the force acting on the particle, which has position $\mathbf{r}$ and moves with velocity $\mathbf{v}$, whilst $\mathbf{a}$ is the acceleration of the particle and t is the time. In a conservative system, the force is independent of both the particle velocity and time whilst in a nonconservative system the force on the particle is a function of its velocity and potentially is also time dependent. An electromagnetic field is nonconservative, where a particle of charge Z is subject to the *Lorentz force*

$$\mathbf{F}(\mathbf{r}, \mathbf{v}, t) = Z[\mathbf{E}(\mathbf{r}, t) + \mathbf{v} \times \mathbf{B}(\mathbf{r}, t)], \tag{7.2}$$

where $\mathbf{E}$ is the *electric field strength* and $\mathbf{B}$ is the *magnetic induction*, which together satisfy *Maxwell's equations*,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{7.3}$$

$$\nabla \times \mathbf{B} - \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \tag{7.4}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{7.5}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}, \tag{7.6}$$

where $\rho(\mathbf{r}, t)$ & $\mathbf{J}(\mathbf{r}, t)$ are the charge & current densities respectively whilst the constants ϵ_0 & μ_0 are the *permittivity of free space* and the *permeability of free space* respectively. Maxwell's equations complement Newton's equation of motion since they provide the means by which $\mathbf{E}$ & $\mathbf{B}$ can be obtained for a known ρ & $\mathbf{J}$, whilst Newton's equation with the Lorentz force allows the determination of ρ & $\mathbf{J}$ if $\mathbf{E}$ & $\mathbf{B}$ are known. This implies that Maxwell's equations for an electromagnetic field can only be solved in concert with Newton's equations of motion for the particles in the given field. For these purposes, an assumption is made that the electromagnetic field is fixed and unaffected by the particles, allowing the problem to be simplified to just considering the equation of motion of particles subject to the Lorentz force.

Nonetheless, Maxwell's equations may be used to derive a more convenient set of identities for $\mathbf{E}$ & $\mathbf{B}$; it can be shown that (7.5) is satisfied if $\mathbf{B}$ can be written in terms of some *vector potential* $\mathbf{A}$ as

$$\nabla \cdot \mathbf{B} = 0 \qquad \Longrightarrow \qquad \mathbf{B} = \nabla \times \mathbf{A}, \tag{7.7}$$

which may be substituted into (7.6), the result of which may be written in terms of some *scalar potential* ϕ as

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = \mathbf{0} \qquad \Longrightarrow \qquad \mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{7.8}$$

The *homogeneous* pair of Maxwell's equations (7.5) & (7.6) are therefore necessarily satisfied when the electric field and the magnetic induction can be written in terms of the scalar & vector potentials $\phi(\mathbf{r}, t)$ & $\mathbf{A}(\mathbf{r}, t)$ respectively. In addition, these potentials contain only four components (ϕ, A_x, A_y, A_z) rather than the six $(E_x, E_y, E_z, B_x, B_y, B_z)$ and as such constitute a more efficient representation of the electromagnetic field.

An inherent property of the potentials $\phi(\mathbf{r}, t)$ & $\mathbf{A}(\mathbf{r}, t)$ is that they are not unique; they may be transformed by an arbitrary gauge function $f(\mathbf{r}, t)$ whilst still yielding the same observable electric and magnetic fields $\mathbf{E}$ & $\mathbf{B}$ respectively, which are uniquely determined by the physical system. Considering the transformed potentials

$$\phi' = \phi - \frac{\partial f}{\partial t} \qquad \& \qquad \mathbf{A}' = \mathbf{A} + \nabla f, \tag{7.9}$$

substituting these into (7.8) & (7.7) yields

$$\mathbf{E}' = -\nabla\phi' - \frac{\partial \mathbf{A}'}{\partial t} = -\phi + \frac{\partial f}{\partial t} - \frac{\partial \mathbf{A}}{\partial t} - \frac{\partial \nabla f}{\partial t} = \mathbf{E}, \tag{7.10}$$

$$\mathbf{B}' = \nabla \times \mathbf{A}' = \nabla \times (\mathbf{A} + \nabla f) = \mathbf{B} + \nabla \times \nabla f = \mathbf{B}, \tag{7.11}$$

thus showing that the gauge transformations do not affect the physical fields. However, the gauge function $f(\mathbf{r}, t)$ may be utilised to apply additional constraints ϕ & $\mathbf{A}$, a pertinent example of which is the constraint that the vector potential have a divergence of zero which defines the *Coulomb gauge*

$$\nabla \cdot \mathbf{A} = 0. \tag{7.12}$$

Returning to the equation of motion (7.1), in *Hamiltonian mechanics* the motion of a particle with position $\mathbf{r}$ and with *conjugate momentum* $\mathbf{p}$ can be described by a scalar Hamiltonian function $\hat{\mathcal{H}}(\mathbf{r}, \mathbf{p})$ for which

$$\frac{\partial \mathbf{r}}{\partial t} = \frac{\partial \hat{\mathcal{H}}}{\partial \mathbf{p}} \qquad \frac{\partial \mathbf{p}}{\partial t} = -\frac{\partial \hat{\mathcal{H}}}{\partial \mathbf{r}}, \tag{7.13}$$

thus defining Hamilton's equations of motion. In a conservative field, the force acting on a particle may be defined in terms of a scalar potential V as

$$F(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}, \tag{7.14}$$

with which it can be seen that the Hamiltonian

$$\hat{\mathcal{H}}(\mathbf{r}, \mathbf{p}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \qquad \Longrightarrow \qquad \frac{\partial \mathbf{r}}{\partial t} = \frac{\partial \hat{\mathcal{H}}}{\partial \mathbf{p}} = \frac{\mathbf{p}}{m} \quad \& \quad \frac{\partial \mathbf{p}}{\partial t} = -\frac{\partial \hat{\mathcal{H}}}{\partial \mathbf{r}} = F(\mathbf{r}) \tag{7.15}$$

yields an equation of motion equivalent to (7.1) since the acceleration is the derivative of velocity with respect to time.

Applying this to the equation of motion of a charged particle in an electromagnetic field, it is necessary to construct a Hamiltonian such that Hamilton's equations of motion are equal to (7.1) where $\mathbf{F}$ is the Lorentz force (7.2). Substituting $\mathbf{E}$ & $\mathbf{B}$ in this expression with their definitions in terms of the potentials in (7.8) & (7.7) respectively, it can be shown that the classical Hamiltonian function becomes

$$\hat{\mathcal{H}}(\mathbf{r}, \mathbf{p}) = \frac{\boldsymbol{\pi}^2}{2m} + Z\phi(\mathbf{r}), \tag{7.16}$$

where the term $\boldsymbol{\pi}$ is the *kinetic momentum*,

$$\boldsymbol{\pi} = \mathbf{p} - Z\mathbf{A}. \tag{7.17}$$

Transposition of this classical equation into a quantum mechanical one is effected by *quantisation*, or the substitution of the momentum term in (7.16) with a quantum mechanical operator

$$\mathbf{p} \to -\imath\hbar\nabla \qquad \Longrightarrow \qquad \boldsymbol{\pi} \to -\imath\hbar\nabla - Z\mathbf{A}. \tag{7.18}$$

For an electron the charge Z is given by $-e$, where e is the proton charge. The resulting form of the Hamiltonian (7.16) for an electron however remains incomplete without any treatment of electron spin; the form that electron spin enters the Hamiltonian was proposed by Pauli as[251]

$$\hat{\mathcal{H}} = \frac{\boldsymbol{\pi}^2}{2m} + \frac{e\hbar}{2m}\mathbf{B} \cdot \boldsymbol{\sigma} - e\phi, \tag{7.19}$$

where $\boldsymbol{\sigma}$ is the set of *Pauli spin matrices*

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -\imath \\ \imath & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.20}$$

The Pauli spin term is often re–written to include the quantisation pre–factor as $\mathbf{s} = \hbar\boldsymbol{\sigma}/2$ to simplify the expression for the Hamiltonian. The non–relativistic Hamiltonian for an electron in an electromagnetic field is therefore written as

$$\hat{\mathcal{H}} = \frac{\boldsymbol{\pi}^2}{2m} + \frac{e}{m}\mathbf{B} \cdot \mathbf{s} - e\phi \qquad \boldsymbol{\pi} = -\imath\hbar\nabla + e\mathbf{A}. \tag{7.21}$$

7.1.2 Molecular Hamiltonian In The Coulomb Gauge

Given the form of the non–relativistic molecular Hamiltonian for an electron in the presence of an electromagnetic field (7.21), a simplification may be made by assuming the Coulomb gauge (7.12) as

$$\begin{aligned} \boldsymbol{\pi}^2 &= (\mathbf{p} + e\mathbf{A}) \cdot (\mathbf{p} + e\mathbf{A}) \\ &= \mathbf{p}^2 + e\mathbf{p} \cdot \mathbf{A} + e\mathbf{A} \cdot \mathbf{p} + e^2\mathbf{A}^2 \\ &= \mathbf{p}^2 + 2e\mathbf{A} \cdot \mathbf{p} + e^2\mathbf{A}^2. \end{aligned} \tag{7.22}$$

Furthermore, in a molecular system it can be assumed that the scalar potential ϕ is dominated by Coulomb potential pertaining to the atomic nuclear charges,

$$\phi(\mathbf{r}) = -\frac{e}{4\pi\epsilon_0}\sum_k \frac{Z_k}{|\mathbf{r}-\mathbf{R}_k|} + \phi_{\text{ext}}(\mathbf{r}), \tag{7.23}$$

where ϕ_{ext} represents any other terms present in the scalar potential. For a system of many electrons, the electronic Hamiltonian can therefore be written in a similar style to (1.3) but with additional terms to describe the interaction with an electromagnetic field, grouped by order in $\mathbf{B}$, as

$$\begin{aligned}\hat{\mathcal{H}} = &\underbrace{\frac{1}{2m}\sum_i \mathbf{p}_i^2 - \frac{e^2}{4\pi\epsilon_0}\sum_{k,i}\frac{Z_k}{|\mathbf{r}_i-\mathbf{R}_k|} + \frac{e^2}{4\pi\epsilon_0}\sum_{i>j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}}_{\text{zeroth-order}} \\ &+ \underbrace{\frac{e}{m}\sum_i \mathbf{A}_i\cdot\mathbf{p}_i + \frac{e}{m}\sum_i \mathbf{B}_i\cdot\mathbf{s}_i - e\sum_i \phi_{\text{ext},i}}_{\text{first-order}} + \underbrace{\frac{e^2}{2m}\sum_i \mathbf{A}_i^2}_{\text{second-order}}.\end{aligned} \tag{7.24}$$

Switching to atomic units, (7.24) can be broken–down as the sum of a zeroth–order Hamiltonian $\hat{\mathcal{H}}_0$ plus a first–order & second–order perturbative Hamiltonian $\hat{\mathcal{H}}^{(1)}$ & $\hat{\mathcal{H}}^{(2)}$ respectively,

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}^{(1)} + \hat{\mathcal{H}}^{(2)} = \hat{\mathcal{H}}_0 + \underbrace{\sum_i \mathbf{A}(\mathbf{r}_i)\cdot\mathbf{p}_i}_{\text{orbital paramagnetic}} + \underbrace{\sum_i \mathbf{B}(\mathbf{r}_i)\cdot\mathbf{s}_i}_{\text{spin paramagnetic}} + \underbrace{\frac{1}{2}\sum_i \mathbf{A}^2(\mathbf{r}_i)}_{\text{diamagnetic}} - \sum_i \phi_{\text{ext}}(\mathbf{r}_i). \tag{7.25}$$

The magnetic perturbations appearing in the Hamiltonian can be characterised as indicated in (7.25), where

(i) *paramagnetic terms* are those that are linear in the magnetic field $\mathbf{B}$ thus may result in either a raising or lowering of the energy of the system relative to zero–field; the orbital paramagnetic term and spin paramagnetic term arise from the interaction between the angular momentum of the electronic orbital and the spin of the electron respectively,

(ii) *diamagnetic terms* are those that are quadratic in the magnetic field $\mathbf{B}$, hence always result in a raised energy relative to the system at zero–field.

The non–relativistic electronic Hamiltonian (7.25) may be written in a more simple form by making implicit the summation over electrons as

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \mathbf{A}(\mathbf{r})\cdot\mathbf{p} + \mathbf{B}(\mathbf{r})\cdot\mathbf{s} + \frac{1}{2}\mathbf{A}^2(\mathbf{r}). \tag{7.26}$$

7.1.3 Gauge Origin Invariance

In the presence of a static & uniform external magnetic field, hence for which $\mathbf{B} = \nabla\times\mathbf{A}$ is constant, the vector potential can be expressed as

$$\mathbf{A}(\mathbf{r};\mathbf{O}) = \frac{1}{2}\mathbf{B}\times(\mathbf{r}-\mathbf{O}), \tag{7.27}$$

where $\mathbf{O}$ is the *gauge origin*, the location of which is arbitrary. Substituting this definition of the vector potential into (7.26) and re–arranging to give

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \frac{1}{2}\mathbf{B}\cdot(\mathbf{r}-\mathbf{O})\times\mathbf{p} + \mathbf{B}\cdot\mathbf{s} + \frac{1}{8}[\mathbf{B}^2(\mathbf{r}-\mathbf{O})^2 - [\mathbf{B}\cdot(\mathbf{r}-\mathbf{O})]^2], \tag{7.28}$$

it can be seen that a change in the position of the gauge origin $\mathbf{O}$ would constitute a gauge transformation of the Hamiltonian. As described previously, a gauge transformation may be parametrised by a gauge function $f(\mathbf{r},t)$, with the resulting transformed potentials defined in (7.9). The transformed Hamiltonian $\hat{\mathcal{H}}'$ can be shown to be related to the original $\hat{\mathcal{H}}$ by the unitary transformation of the related term $\hat{\mathcal{H}} - \imath\partial/\partial t$ as

$$\left(\hat{\mathcal{H}}' - \imath\frac{\partial}{\partial t}\right) = \exp(-\imath f)\left(\hat{\mathcal{H}} - \imath\frac{\partial}{\partial t}\right)\exp(\imath f). \tag{7.29}$$

In order that the Schrödinger Equation still be satisfied, the eigenfunctions of the transformed Hamiltonian Ψ' must be related to the eigenfunctions of the original Hamiltonian Ψ by a compensating unitary transformation,

$$\Psi' = \exp(-\imath f)\Psi. \tag{7.30}$$

As a result, all observables of the system are unaffected by gauge transformations of the Hamiltonian & wavefuncton; this can be demonstrated simply for the density as

$$\rho'(\mathbf{r}) = [\Psi'(\mathbf{r})]^*[\Psi'(\mathbf{r})] = [\exp(-\imath f)\Psi(\mathbf{r})]^*[\exp(-\imath f)\Psi(\mathbf{r})] = \Psi^*(\mathbf{r})\Psi(\mathbf{r}) = \rho(\mathbf{r}). \tag{7.31}$$

For the static magnetic field described by a vector potential of the form (7.27), the gauge function characterising a shift in gauge origin is defined by

$$\mathbf{A}(\mathbf{r};\mathbf{G}) = \mathbf{A}(\mathbf{r};\mathbf{O}) - \mathbf{A}(\mathbf{G};\mathbf{O}) = \mathbf{A}(\mathbf{r};\mathbf{O}) + \nabla f \qquad \Longrightarrow \qquad f(\mathbf{r}) = -\mathbf{A}(\mathbf{G};\mathbf{O}) \cdot \mathbf{r}. \tag{7.32}$$

Thus a shift in the gauge origin from $\mathbf{O}$ to $\mathbf{G}$ effects a gauge transformation of the wavefunction given by

$$\Psi' = \exp(\imath\mathbf{A}(\mathbf{G};\mathbf{O}) \cdot \mathbf{r})\Psi. \tag{7.33}$$

Such behaviour, required to make observables gauge origin independent, cannot be readily modelled by standard atomic orbitals in a practical calculation. The means by which this is addressed here is by the inclusion of the phase–factor $\exp(\imath\mathbf{A}(\mathbf{G};\mathbf{O}) \cdot \mathbf{r})$ in the atomic orbitals themselves as originally proposed by London in Ref. 246. These modified functions, here referred to as *London atomic orbitals* (LAOs), respond to an applied magnetic field in a manner that is correct to first order in $\mathbf{B}$ and their use permits calculations of a system to be rigorously gauge origin independent.

As described in the introduction, the use of LAOs is not a new innovation and there is extensive literature of their deployment in gauge origin independent calculations of magnetic response properties[239–241,252–254]. The limitations of this perturbative approach preclude the study of systems in the presence of strong magnetic fields beyond the Coulomb regime and additionally make difficult the calculation of highly nonlinear magnetic response properties.

The approach of Ref. 172 seeks to address these limitations by employing LAOs for the full self–consistent field procedure, providing a non–perturbative treatment of the effects of arbitrarily strong magnetic fields on electronic systems entirely independent of the gauge origin. Such an approach can largely be applied within the general self–consistent field framework set–out in Chapter 1, since the only material difference in the wavefunction that must be taken account of in the self–consistent field procedure is that it is generally complex and cannot be assumed to have a vanishing imaginary component.

A significant technical challenge however remains; the methods employed for the computation of molecular integrals over the basis functions in standard calculations (see subsection 1.2.3) are only appropriate for real–valued basis functions, not those such as LAOs which contain a complex phase factor. In Ref. 172, a generalisation of standard integral algorithms to complex–valued basis functions was presented that addressed this problem to an extent. However, the approach taken in Ref. 172 provides a suboptimal method for integrating LAOs thus significantly limiting the range of systems to which the non–perturbative treatment of strong magnetic fields can be applied. Chapter 8 presents an efficient and straightforward alternative formulation of the calculation of molecular integrals using LAOs and its implementation in the quantum chemical development code QUEST[63].

8 Integrals Over London Atomic Orbitals

8.1 Introduction

In recent years, there has been a great deal of interest in the non–perturbative calculation of molecular energies and properties in the presence of arbitrarily strong magnetic fields[172,255–265]. Investigations have included the implementation of electronic structure methods at the Hartree-Fock[172,261], configuration–interaction[257], coupled-cluster[262], coupled-cluster equation of motion[265] and current density-functional[259,263,264] levels of theory. Underpinning the implementation of these methods has been the use of Gaussian–type *London atomic orbitals*[239,246] (LAOs). LAOs consist of standard Gaussian basis functions multiplied by a field-dependent plane-wave phase factor, allowing results independent of gauge-origin to be obtained using finite basis sets.

One of the principle challenges in these calculations is the the evaluation of molecular integrals over LAOs. A number of algorithms have been put forward for the evaluation of electron repulsion integrals (ERIs) over LAOs including the generalised Obara–Saika[266–268], accompanying coordinate expansion[269,270], Rys quadrature[261,271,272], and McMurchie–Davidson[172,273–275] schemes. In particular, the McMurchie–Davidson scheme has been employed in the LONDON quantum chemistry program[276] and a version of the Rys quadrature scheme in the BAGEL program[277].

A number of advantageous features for integral schemes to perform non-perturbative calculations with arbitrary strength magnetic fields can be identified. The first of these is that the algorithmic complexity should not be significantly increased over the corresponding zero-field scheme. The second is that they should be applicable to arbitrarily high angular momenta functions, in order to allow sufficient flexibility for the description of anisotropic changes in the electronic structure upon application of a magnetic field. The third feature is that they should be efficient for application to both contracted and primitive functions, since the latter may provide additional basis set flexibility in high field applications.

In common with integral schemes at zero field, no single algorithm can satisfy all of these requirements simultaneously. Indeed, each algorithm inherently has its own strengths and weaknesses; often the most effective approach is to use several algorithms and select the most efficient for each integral class extemporaneously. In this work we explore generalised algorithms based on the McMurchie–Davidson (MD), Head-Gordon Pople (HGP) and Rys quadrature methods.

It has been shown that the generalised MD algorithm is significantly more complex than the standard equivalent, requiring many more intermediates, thus does not exhibit the same efficiency or scaling. In contrast, the HGP method can be generalised without increasing the underlying complexity of the algorithm; here we present an implementation of this approach for the first time. A generalised Rys quadrature has been explored previously, however its use was limited to low field strengths owing to the fact that the required quadrature roots and weights were being approximated by a 2D interpolation scheme. Whilst adequate for this purpose, the application of this approach to arbitrary field strengths is problematic as it would necessitate the storage of very large interpolation grids. In addition, it was noted in Ref. 261 that high-precision arithmetic was necessary to determine the required parameters. In this work, we generalise the Rys quadrature approach to a much wider range of field strengths using the approach put forward by Flocke[278] to compute the roots and weights when required.

We commence in Section 8.2 by discussing some preliminaries, including the use of shell-pair quantities and the transformation between Cartesian and spherical harmonic Gaussians. In Section 8.3, we discuss the calculation of the one-electron integrals required for practical calculations, including the generalisation of the kinetic energy integrals associated with the introduction of a vector potential describing the magnetic field. In Section 8.4 we describe three approaches to the calculation LAO-ERIs, comparing their relative algorithmic complexity. In Section 8.5, relative timings for each approach are discussed, illustrating the relative advantages and disadvantages of each algorithm. A mixed scheme, which selects the most efficient algorithm for each class of integral extemporaneously is presented. This mixed approach provides an an effective approach to minimising the computational cost of the LAO-ERI evaluation.

8.2 Preliminaries & Shell–Pair Data

In this work we are concerned with the evaluation of molecular integrals over Gaussian–type LAOs. A standard unnormalised Cartesian Gaussian–type orbital (GTO) has the general form

$$\phi_a(\mathbf{r}) = (x - \mathrm{A}_x)^{a_x} (y - \mathrm{A}_y)^{a_y} (z - \mathrm{A}_z)^{a_z} \sum_{k=1}^{K_a} d_k \exp\Big(-\alpha_k |\mathbf{r} - \mathbf{A}|^2\Big) \tag{8.1}$$

where the function is centred at $\mathbf{A} = (\mathrm{A}_x, \mathrm{A}_y, \mathrm{A}_z)$, has angular momentum $\mathbf{a} = (a_x, a_y, a_z)$ and has exponents $\{\alpha_k\}$ with respective contraction coefficient $\{d_k\}$. Where the contraction length K_a is equal to 1, the GTO only has one exponent with a corresponding contraction coefficient of 1.0 and is a *primitive* GTO, whereas if $K_a > 1$ the GTO is *contracted*. Contracted basis sets such as those proposed by Pople[279,280] & Dunning[116] are often used in quantum chemical computations as they allow an increased number of Gaussian exponents to be included in the basis set without increasing the overall number of basis functions, thus providing the opportunity to increase accuracy without significant increases in computational cost. Gaussian–type LAOs are similar to standard GTOs, differing by a phase factor

$$\omega_a(\mathbf{r}) = \phi_a(\mathbf{r}) \exp(-\imath \mathbf{k}_a \cdot \mathbf{r}) \tag{8.2}$$

where $\mathbf{k}_a$ is the wave vector of the London plane wave, $\mathbf{k}_a = \frac{1}{2}\mathcal{B} \times (\mathbf{A} - \mathbf{O})$, depending on the *external magnetic field* $\mathcal{B}$ and relative to the *gauge–origin* $\mathbf{O}$.

In contemporary molecular integral codes, integrals are almost always computed in shell–batches rather than individually[281]; a shell comprises all basis functions with common centre, exponent and total angular momentum $L_a = a_x + a_y + a_z$ but with different distributions between the three Cartesian components. It is highly advantageous to compute integrals in shell–batches as integrals between individual GTOs within the shells have many common intermediates which may be re–used for all integrals in the batch.

Prior to the calculation of any molecular integrals, the important task of computing shell–pair quantities is carried–out. This serves two main purposes: firstly to streamline subsequent integral evaluation by having key quantities precomputed, reducing the cost of their evaluation from $\mathcal{O}(n^4)$ in the two–electron integrals to $\mathcal{O}(n^2)$ and secondly, to reduce the number of integrals to be calculated by discarding negligible shell–pairs. As each shell–pair describes one charge distribution (i.e. it is simply the product of two shells), if the overlap between the two shells is negligible, it may be discarded. This is particularly important for larger systems, as the number of significant shell–pairs scales only *linearly* with molecule size[281].

In the evaluation of integrals over LAOs, the computation of shell–pair data can be of further advantage, as the field–related terms can be sequestered within the standard pair quantities, allowing all subsequent integrals (with the exception of kinetic–energy integrals) to be evaluated *without* explicit reference to the complex phase factor.

8.2.1 Shell–Pairs

In order to discuss the relevance of shell–pairs, it is first necessary to define the charge distributions comprised of products of LAOs,

$$\omega_a^*(\mathbf{r})\omega_b(\mathbf{r}) = \sum_{\mu=1}^{K_a} \sum_{\nu=1}^{K_b} [\mathbf{a}_\mu \mathbf{b}_\nu| = (\mathbf{ab}| \tag{8.3}$$

where the notation $[\mathbf{a}_\mu \mathbf{b}_\nu|$ represents the product of the μ^{th} and ν^{th} individual contractions of ω_a and ω_b respectively, whilst $(\mathbf{ab}|$ is the overall inner product of the two LAOs; if both are primitive, the two definitions are equivalent. As will be discussed in the following sections, algorithms for the computation of molecular integrals contain steps that may only be applied to primitive functions, hence shell–pair quantities are calculated for each $[\mathbf{a}_\mu \mathbf{b}_\nu|$ in the charge distribution.

Given an individual pair of primitive functions, centred on $\mathbf{A}$ and $\mathbf{B}$, with exponents α and β and contraction coefficients d_a and d_b respectively, the *Gaussian product theorem* allows their product to be re–written as a single Gaussian function multiplied by a pre–exponential factor,

$$d_a \exp\big(-\alpha|\mathbf{r} - \mathbf{A}|^2\big) \cdot d_b \exp\big(-\beta|\mathbf{r} - \mathbf{B}|^2\big) = d_a d_b \underbrace{\exp\left(-\frac{\alpha\beta}{\alpha+\beta}|\mathbf{A} - \mathbf{B}|^2\right)}_{\text{pre-exponential factor}} \underbrace{\exp\big(-\zeta|\mathbf{r} - \mathbf{P}|^2\big)}_{\text{product Gaussian}}, \tag{8.4}$$

where $\mathbf{P}$ is the Gaussian product centre. For the product of two primitive GTOs without phase factor, the following pairwise quantities are computed

$$\zeta = \alpha + \beta \qquad \mathbf{P} = \frac{1}{\zeta}\left(\alpha\mathbf{A} + \beta\mathbf{B}\right) \qquad U_{\mathrm{P}} = d_a d_b \left(\frac{\pi}{\zeta}\right)^{\frac{3}{2}} \exp\left(-\frac{\alpha\beta}{\zeta}\left|\mathbf{AB}\right|^2\right), \tag{8.5}$$

where $\mathbf{AB} = \mathbf{A} - \mathbf{B}$. With the presence of the complex phase factor, additional terms appear in (8.4); generalising (8.4) to simple Gaussian functions each with a complex phase factor as defined in (8.2) yields

$$\begin{aligned} d_a \exp\left(-\alpha|\mathbf{r}-\mathbf{A}|^2\right) \exp(-\imath\mathbf{k}_a\cdot\mathbf{r})^* d_b \exp\left(-\beta|\mathbf{r}-\mathbf{B}|^2\right) \exp(-\imath\mathbf{k}_b\cdot\mathbf{r}) &= d_a d_b \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right) \exp(-\imath(\mathbf{k}_b-\mathbf{k}_a)\cdot\mathbf{r}) \exp\left(-\zeta|\mathbf{r}-\mathbf{P}|^2\right) \\ &= d_a d_b \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right) \exp\left(-\frac{|\mathbf{k}_b-\mathbf{k}_a|^2}{4\zeta} - \imath\mathbf{P}\cdot(\mathbf{k}_b-\mathbf{k}_a)\right) \exp\left(-\zeta\left|\mathbf{r}-\mathbf{P}-\imath\frac{\mathbf{k}_b-\mathbf{k}_a}{2\zeta}\right|^2\right), \end{aligned} \tag{8.6}$$

which has the form of a Gaussian function with its origin in the complex plane. Defining the following additional pairwise quantities

$$\boldsymbol{\chi}_{\mathrm{P}} = \frac{1}{2}\mathcal{B}\times(\mathbf{B}-\mathbf{A}) \qquad \mathcal{K}_{\mathrm{P}} = \exp\left(-\frac{1}{4\zeta}\boldsymbol{\chi}_{\mathrm{P}}\cdot\boldsymbol{\chi}_{\mathrm{P}} - \imath\mathbf{P}\cdot\boldsymbol{\chi}_{\mathrm{P}}\right), \tag{8.7}$$

the LAO shell–pair for each pair of primitive functions requires only computation and storage of the following

$$2\alpha \qquad 2\beta \qquad \frac{1}{2\zeta} \qquad \tilde{\mathbf{P}} = \mathbf{P} - \frac{\imath}{2\zeta}\boldsymbol{\chi}_{\mathrm{P}} \qquad \tilde{U}_{\mathrm{P}} = U_{\mathrm{P}}\mathcal{K}_{\mathrm{P}}. \tag{8.8}$$

However, if $|\tilde{U}_{\mathrm{P}}| \leqslant 10^{-12}$ the pair is considered negligible and discarded from the shell–pair; this allows an increasingly large proportion of the Gaussian product space to be discarded as the system becomes larger. Within this framework of (reduced) shell–pairs, the contraction of (8.3) may be applied as early as possible in each integral algorithm to yield contracted integrals.

Advocates of the shell–pair approach have noted[281,282] that the number of components within a charge distribution may be minimised by a process of modelling the shell–pair with as few primitive components as possible but which yields a matching electrostatic potential[283]; this could yield further efficiencies and may be considered in future work.

8.2.2 Transformation Matrices

In the computation of ERIs in particular, the horizontal recursion relation (HRR) of Head–Gordon & Pople[284], equivalent to the transfer relation of Rys and co-workers[285], can be used for more efficient treatment of contracted basis functions,

$$(\mathbf{ab}+\mathbf{1}_i| = (\mathbf{a}+\mathbf{1}_i\mathbf{b}| + \mathbf{AB}_i\,(\mathbf{ab}| \tag{8.9}$$

This recursion relation, whilst being relatively simple, can grow significantly in cost when the total angular momentum of the integral becomes large[286,287]. Given that this relation is only a two–index quantity, it is convenient to reformulate it as the application of a transformation matrix[288], allowing the recursion to be efficiently executed as a matrix multiplication,

$$\begin{aligned} (\mathbf{ab}| = &\sum_{e_x=a_x}^{a_x+b_x} \binom{b_x}{e_x} (\mathrm{A}_x - \mathrm{B}_x)^{b_x-e_x} \\ &\times \sum_{e_y=a_y}^{a_y+b_y} \binom{b_y}{e_y} (\mathrm{A}_y - \mathrm{B}_y)^{b_y-e_y} \\ &\times \sum_{e_z=a_z}^{a_z+b_z} \binom{b_z}{e_z} (\mathrm{A}_z - \mathrm{B}_z)^{b_z-e_z}\ (\mathbf{e0}|\,. \end{aligned} \tag{8.10}$$

These relatively expensive matrices may be precomputed for each contracted shell–pair and stored along with the quantities in (8.8), providing an efficient way of executing the HRR of (8.9).

Cartesian Gaussian functions of the form defined in (8.1) can alternatively be expressed as spherical harmonic Gaussian functions of the form

$$\phi_a(\mathbf{r}) = \mathcal{C}_{\ell m}(\theta,\varphi) \sum_{k=1}^{K_a} d_k \exp\left(-\alpha_k\left|\mathbf{r}-\mathbf{A}\right|^2\right), \tag{8.11}$$

where ℓ is the orbital angular momentum, m is the magnetic angular momentum number and $\mathcal{C}_{\ell m}(\theta,\varphi)$ is the $\{\ell m\}^{th}$ *solid harmonic*. It can be shown that the set of functions defined (8.11) is a *subset* of those defined by (8.1). This can be seen by comparing the explicit functional forms of Cartesian and spherical GTOs for orbital angular momenta $0-2$,

$$\begin{aligned}
&\ell = 0 \qquad \phi \sim \exp\left(-\alpha r^2\right) \qquad\qquad \phi \sim \exp\left(-\alpha r^2\right) \\
&\ell = 1 \qquad \phi \sim \begin{cases} x\exp\left(-\alpha r^2\right) \\ y\exp\left(-\alpha r^2\right) \\ z\exp\left(-\alpha r^2\right) \end{cases} \qquad \phi \sim \begin{cases} \mathcal{C}_{1,-1}(\theta,\varphi) r\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{1,\ 0}(\theta,\varphi) r\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{1,-1}(\theta,\varphi) r\exp\left(-\alpha r^2\right) \end{cases} \\
&\ell = 2 \qquad \phi \sim \begin{cases} x^2\exp\left(-\alpha r^2\right) \\ y^2\exp\left(-\alpha r^2\right) \\ z^2\exp\left(-\alpha r^2\right) \\ xy\exp\left(-\alpha r^2\right) \\ xz\exp\left(-\alpha r^2\right) \\ yz\exp\left(-\alpha r^2\right) \end{cases} \qquad \phi \sim \begin{cases} \mathcal{C}_{2,-2}(\theta,\varphi) r^2\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{2,-1}(\theta,\varphi) r^2\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{2,\ 0}(\theta,\varphi) r^2\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{2,\ 1}(\theta,\varphi) r^2\exp\left(-\alpha r^2\right) \\ \mathcal{C}_{2,\ 2}(\theta,\varphi) r^2\exp\left(-\alpha r^2\right) \end{cases}
\end{aligned} \tag{8.12}$$

It can be seen here that, for $\ell = 0$ and $\ell = 1$, the set of Cartesian and spherical Gaussian functions are the same. However, for $\ell = 2$, the Cartesian set comprises six functions whereas the spherical set comprises only five functions. It is found that the difference between the two sets is equivalent to an *s*–type function, which is an artefact of the representation of GTOs by three Cartesian indices as opposed to the two angular momenta quantum numbers as in the spherical representation. The number of additional functions in the Cartesian representation increases further with increasing angular momentum, such that the number of Cartesian GTOs can become substantially greater than the number of spherical GTOs at high angular momentum.

Most integral algorithms, including those discussed in the present chapter, are formulated for Cartesian Gaussian functions since these are more amenable to computational implementation than spherical Gaussian functions. However, it can be advantageous to use a spherical Gaussian basis since the lower number of individual functions present for $\ell \geqslant 2$ reduces the computational costs whilst maintaining the same number and range of exponents. A common approach is to compute batches of integrals in the Cartesian basis and subsequently transform them into the spherical basis. In the present work, if a spherical harmonic Gaussian basis is being used, the Cartesian to spherical transformation can be built into the HRR transformation matrices, eliminating the need for a costly four–index transformation on each shell–quartet and also reducing the size of the HRR matrices to be stored. For standard GTOs, the Cartesian to spherical transformation is that derived by Schlegel and Frisch[289],

$$\mathcal{S}^{\ell m}_{l_x l_y l_z} = \sqrt{\frac{(2l_x)!(2l_y)!(2l_z)!\ell!(\ell-|m|)!}{(2\ell)!l_x!l_y!l_z!(\ell+|m|)!}} \frac{1}{2^\ell \ell!} \sum_{i=0}^{(\ell-|m|)/2} \binom{\ell}{i} \binom{i}{j} \frac{(-1)^i (2\ell-2i)!}{(\ell-|m|-2i)!} \sum_{k=0}^{j} \binom{j}{k} \binom{|m|}{l_x-2k} (-1)^{\mathrm{sgn}(m)(|m|-l_x+2k)/2} \tag{8.13}$$

in which $\ell = l_x + l_y + l_z$ and $j = (l_x + l_y - |m|)/2$; if j has a half–integer value, the respective $\mathcal{S}^{\ell m}_{l_x l_y l_z} = 0$. Given that the complex component of LAOs is simply a phase–factor to the real Gaussian–type orbital, the transformation from Cartesian to spherical Gaussian basis is unchanged from the standard case. The transformations for $m = 0$ are real, whilst for nonzero m the complex transformations may be combined into two real forms, $(\mathcal{S}^{\ell m} + \mathcal{S}^{\ell -m})/\sqrt{2}$ for $m > 0$ and $(\mathcal{S}^{\ell m} - \mathcal{S}^{\ell -m})/\sqrt{-2}$ for $m < 0$.

8.3 One Electron Integrals

Having assembled the shell–pair quantities, it is necessary to compute the one–electron integrals required for the construction of the Fock matrix; these are the overlap, kinetic energy and nuclear–attraction integrals.

8.3.1 Overlap Integrals

The two–centre overlap integral is the simplest to compute and is defined for LAOs as

$$(\mathbf{a}\,\mathbf{b}) = \int \omega_a^*(\mathbf{r})\omega_b(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{8.14}$$

Overlap integrals between Cartesian Gaussian functions are easily computed from shell–pair data using recurrence relations derived by Obara & Saika (OS) from the differential properties of Gaussian functions[266,290], with which the

integrals over GTOs of arbitrary angular momentum can be reduced to those of GTOs with angular momentum zero. Considering two such functions with contraction length one, their overlap integral (8.14) can be written as

$$\begin{aligned}\int \omega_a^*(\mathbf{r})\omega_b(\mathbf{r})\mathrm{d}\mathbf{r} &= \int \exp\left(-\alpha|\mathbf{r}-\mathbf{A}|^2\right)\exp(-\imath\mathbf{k}_a\cdot\mathbf{r})^*\exp\left(-\beta|\mathbf{r}-\mathbf{B}|^2\right)\exp(-\imath\mathbf{k}_b\cdot\mathbf{r})\mathrm{d}\mathbf{r}\\ &= \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\exp\left(-\frac{|\chi_\mathrm{P}|^2}{4\zeta}-\imath\mathbf{P}\cdot\chi_\mathrm{P}\right)\int\exp\left(-\zeta\left|\mathbf{r}-\mathbf{P}-\imath\frac{\chi_\mathrm{P}}{2\zeta}\right|^2\right)\mathrm{d}\mathbf{r}\\ &= \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int\exp\left(-\zeta\left|\mathbf{r}-\tilde{\mathbf{P}}\right|^2\right)\mathrm{d}\mathbf{r}\\ &= \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\left(\frac{\pi}{\zeta}\right)^{\frac{3}{2}} = U_\mathrm{P}\mathcal{K}_\mathrm{P} = \tilde{U}_\mathrm{P}\end{aligned} \tag{8.15}$$

The integral over LAOs with angular momentum zero is thus simply $\tilde{U}_\mathrm{P}$. Given therefore the definitions

$$\mathbf{PA} = \tilde{\mathbf{P}} - \mathbf{A} \qquad [\mathbf{0}\,\mathbf{0}] = \tilde{U}_\mathrm{P} \tag{8.16}$$

and the separability of the integrand by Cartesian axes, overlap integrals of higher angular momentum can be obtained by applying the recursion relation to primitive integrals in each Cartesian axis

$$[\mathbf{a}+\mathbf{1}_i\,\mathbf{b}] = \mathbf{PA}_i\,[\mathbf{a}\,\mathbf{b}] + \frac{a_i}{2\zeta}\,[\mathbf{a}-\mathbf{1}_i\,\mathbf{b}] + \frac{b_i}{2\zeta}\,[\mathbf{a}\,\mathbf{b}-\mathbf{1}_i] \tag{8.17}$$

A primitive overlap integral may then be computed as

$$[\mathbf{a}\,\mathbf{b}] = [a_x\;b_x]\cdot[a_y\;b_y]\cdot[a_z\;b_z] \tag{8.18}$$

which is then contracted according to (8.3) and transformed into a spherical Gaussian basis if required.

8.3.2 Multipole, Differential & Kinetic Energy Integrals

The kinetic energy integrals are slightly more complicated to evaluate than overlap integrals due to the presence of the diamagnetic and paramagnetic terms in the kinetic energy operator, which are not normally considered in the zero–field case. In the Coulomb gauge, the kinetic energy operator is given by half the square of the kinetic–momentum operator $\hat{\boldsymbol{\pi}}$, thus the kinetic energy integral is defined as

$$\left(\mathbf{a}\left|\frac{1}{2}\hat{\boldsymbol{\pi}}^2\right|\mathbf{b}\right) = \frac{1}{2}\int\omega_a^*(\mathbf{r})\hat{\boldsymbol{\pi}}^2\omega_b(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{8.19}$$

In a uniform field $\mathcal{B}$, the square of the kinetic–momentum operator at position $\mathbf{R}$ can be expanded as

$$\begin{aligned}\hat{\boldsymbol{\pi}}^2 &= \left(-\imath\nabla+\frac{1}{2}\mathcal{B}\times\mathbf{R}\right)^2\\ &= -\nabla^2 - \frac{\imath}{2}\nabla\left(\mathcal{B}\times\mathbf{R}\right) - \frac{\imath}{2}\left(\mathcal{B}\times\mathbf{R}\right)\nabla + \frac{1}{4}\left(\mathcal{B}\times\mathbf{R}\right)^2.\end{aligned} \tag{8.20}$$

Considering the x–component of this operator, (8.20) may be further resolved by substituting the corresponding components for the cross–products,

$$\begin{aligned}\hat{\pi}_x^2 = &-\frac{\partial^2}{\partial x^2} - \frac{\imath}{2}\frac{\partial}{\partial x}\left(\mathcal{B}_y\mathbf{R}_z - \mathcal{B}_z\mathbf{R}_y\right)\\ &-\frac{\imath}{2}\left(\mathcal{B}_y\mathbf{R}_z-\mathcal{B}_z\mathbf{R}_y\right)\frac{\partial}{\partial x} + \frac{1}{4}\left(\mathcal{B}_y\mathbf{R}_z-\mathcal{B}_z\mathbf{R}_y\right)^2.\end{aligned} \tag{8.21}$$

Given that the total kinetic energy operator is equal to

$$\frac{1}{2}\hat{\boldsymbol{\pi}}^2 = \frac{1}{2}\hat{\pi}_x^2 + \frac{1}{2}\hat{\pi}_y^2 + \frac{1}{2}\hat{\pi}_z^2, \tag{8.22}$$

the integral is separable into the sum of its components in each of the Cartesian directions and it can be evaluated as such; the operator for the x–component is given by

$$\begin{aligned}\frac{1}{2}\hat{\pi}_x^2 = &-\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\imath}{2}\mathcal{B}_y\frac{\partial}{\partial x}\mathbf{R}_z + \frac{\imath}{2}\mathcal{B}_z\frac{\partial}{\partial x}\mathbf{R}_y\\ &+\frac{1}{8}\mathcal{B}_y^2\mathbf{R}_z^2 + \frac{1}{8}\mathcal{B}_z^2\mathbf{R}_y^2 - \frac{1}{4}\mathcal{B}_y\mathcal{B}_z\mathbf{R}_y\mathbf{R}_z,\end{aligned} \tag{8.23}$$

with equivalent expressions defining the y and z components. All required terms are derived from the corresponding overlap integrals, computed in the standard way described in subsection 8.3.1. From these, multipole and differential integrals up to second order are obtained by application of their respective recursion relations. For simplicity of notation, n^{th}–order multipole and differential operators are respectively denoted as

$$(x - O_x)^n \rightarrow x_o^n \qquad \frac{\partial^n}{\partial x^n} \rightarrow \partial_x^n. \tag{8.24}$$

This leads to the following relation for multipole operators,

$$\left[\mathbf{a}\left|x_o^{n+1}\right|\mathbf{b}\right] = \mathbf{BO}_x\left[\mathbf{a}\left|x_o^n\right|\mathbf{b}\right] + \left[\mathbf{a}\left|x_o^n\right|\mathbf{b} + \mathbf{1}_x\right] \tag{8.25}$$

and the corresponding recursion relation for differential operators,

$$\left[\mathbf{a}\left|\partial_x^{n+1}\right|\mathbf{b}\right] = b_x\left[\mathbf{a}\left|\partial_x^n\right|\mathbf{b} - \mathbf{1}_x\right] - \imath\mathbf{k}_{b,x}\left[\mathbf{a}\left|\partial_x^n\right|\mathbf{b}\right] - 2\beta\left[\mathbf{a}\left|\partial_x^n\right|\mathbf{b} + \mathbf{1}_x\right], \tag{8.26}$$

where $\mathbf{O}$ is the gauge–origin, $\mathbf{k}_b$ the London phase–factor at $\mathbf{B}$, and

$$\left[\mathbf{a}\left|x_o^0\right|\mathbf{b}\right] = \left[\mathbf{a}\left|\partial_x^0\right|\mathbf{b}\right] = [\mathbf{a}\,\mathbf{b}]. \tag{8.27}$$

Higher order multipole and differential integrals can be obtained by repeated application of the above recurrence relations. The final expression for the x–component of the kinetic energy integral is given by

$$\begin{aligned}\left[\mathbf{a}\left|\frac{1}{2}\hat{\pi}_x^2\right|\mathbf{b}\right] = &-\frac{1}{2}\left[a_x\left|\partial_x^2\right|b_x\right]\cdot[a_y\,b_y]\cdot[a_z\,b_z] + \frac{\imath}{2}\mathcal{B}_z\left[a_x\left|\partial_x\right|b_x\right]\cdot[a_y\left|y_o\right|b_y]\cdot[a_z\,b_z] \\ &-\frac{\imath}{2}\mathcal{B}_y\left[a_x\left|\partial_x\right|b_x\right]\cdot[a_y\,b_y]\cdot[a_z\left|z_o\right|b_z] + \frac{1}{8}\mathcal{B}_y^2\left[a_x\,b_x\right]\cdot[a_y\,b_y]\cdot\left[a_z\left|z_o^2\right|b_z\right] \\ &+\frac{1}{8}\mathcal{B}_z^2\left[a_x\,b_x\right]\cdot\left[a_y\left|y_o^2\right|b_y\right]\cdot[a_z\,b_z] - \frac{1}{4}\mathcal{B}_y\mathcal{B}_z\left[a_x\,b_x\right]\cdot[a_y\left|y_o\right|b_y]\cdot[a_z\left|z_o\right|b_z],\end{aligned} \tag{8.28}$$

with equivalent expressions for the y and z–components. Hence from (8.22), the primitive kinetic energy integral is given by

$$\left[\mathbf{a}\left|\frac{1}{2}\hat{\pi}^2\right|\mathbf{b}\right] = \left[\mathbf{a}\left|\frac{1}{2}\hat{\pi}_x^2\right|\mathbf{b}\right] + \left[\mathbf{a}\left|\frac{1}{2}\hat{\pi}_y^2\right|\mathbf{b}\right] + \left[\mathbf{a}\left|\frac{1}{2}\hat{\pi}_z^2\right|\mathbf{b}\right]. \tag{8.29}$$

In the same way as for the overlap integrals, kinetic–energy integrals over LAOs are computed in the primitive Cartesian Gaussian basis for a given shell–pair, and subsequently contracted according to (8.3) and transformed into the spherical Gaussian basis if required.

8.3.3 Nuclear Attraction Integrals

In contrast to the aforementioned one–electron integrals, the nuclear attraction integral is not separable into Cartesian components due to the Coulomb operator that defines the integral,

$$\left(\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right) = \int\frac{\omega_a^*(\mathbf{r})\omega_b(\mathbf{r})}{|\mathbf{r} - \mathbf{C}|}\mathrm{d}\mathbf{r}, \tag{8.30}$$

where $\mathbf{C}$ is the position of an atomic nucleus with unit charge. The most common approach to computing such integrals is to reduce them by applying the Gaussian product theorem and substituting the Coulomb operator with its Laplace transform

$$\frac{1}{\mathbf{r}} = \frac{2}{\sqrt{\pi}}\int_0^\infty \exp\left(-u^2|\mathbf{r}|^2\right)\mathrm{d}u, \tag{8.31}$$

to yield a one–dimensional integral that can be approximated numerically relatively inexpensively, using the molecular incomplete gamma function. Given two LAOs of angular momentum zero and contraction length of one, the transformation of the Coulomb operator can be derived as

$$\begin{aligned}\int\frac{\omega_a^*(\mathbf{r})\omega_b(\mathbf{r})}{|\mathbf{r} - \mathbf{C}|}\mathrm{d}\mathbf{r} &= \int\frac{\exp\left(-\alpha|\mathbf{r} - \mathbf{A}|^2\right)\exp(-\imath\mathbf{k}_a\cdot\mathbf{r})^*\exp\left(-\beta|\mathbf{r} - \mathbf{B}|^2\right)\exp(-\imath\mathbf{k}_b\cdot\mathbf{r})}{|\mathbf{r} - \mathbf{C}|}\mathrm{d}\mathbf{r} \\ &= \exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int\exp\left(-\zeta\left|\mathbf{r} - \tilde{\mathbf{P}}\right|^2\right)\frac{1}{|\mathbf{r} - \mathbf{C}|}\mathrm{d}\mathbf{r} \\ &= \frac{2}{\sqrt{\pi}}\exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int\exp\left(-\zeta\left|\mathbf{r} - \tilde{\mathbf{P}}\right|^2\right)\left[\int_0^\infty\exp\left(-u^2|\mathbf{r} - \mathbf{C}|^2\right)\mathrm{d}u\right]\mathrm{d}\mathbf{r}.\end{aligned} \tag{8.32}$$

A subsequent change in the order of integration allows the Gaussian product theorem to be applied once more as

$$\begin{aligned}\int \frac{\omega_a^*(\mathbf{r})\omega_b(\mathbf{r})}{|\mathbf{r}-\mathbf{C}|}\mathrm{d}\mathbf{r} &= \frac{2}{\sqrt{\pi}}\exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int \exp\left(-\zeta\left|\mathbf{r}-\tilde{\mathbf{P}}\right|^2\right)\left[\int_0^\infty \exp\left(-u^2|\mathbf{r}-\mathbf{C}|^2\right)\mathrm{d}u\right]\mathrm{d}\mathbf{r} \\ &= \frac{2}{\sqrt{\pi}}\exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int_0^\infty\left[\int \exp\left(-(\zeta+u^2)\left|\mathbf{r}-\frac{\zeta\tilde{\mathbf{P}}+u^2\mathbf{C}}{\zeta+u^2}\right|^2\right)\mathrm{d}\mathbf{r}\right]\exp\left(-\frac{\zeta u^2}{\zeta+u^2}|\tilde{\mathbf{P}}-\mathbf{C}|^2\right)\mathrm{d}u \\ &= \frac{2}{\pi}\exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\mathcal{K}_\mathrm{P}\int_0^\infty\left(\frac{1}{\zeta+u^2}\right)^{\frac{3}{2}}\exp\left(-\frac{\zeta u^2}{\zeta+u^2}|\tilde{\mathbf{P}}-\mathbf{C}|^2\right)\mathrm{d}u.\end{aligned} \tag{8.33}$$

This intermediate is further transformed by applying the substitution of variables

$$t^2 = u^2(\zeta+u^2) \qquad \Longrightarrow \qquad \mathrm{d}u = \frac{\left(\zeta+u^2\right)^{\frac{3}{2}}}{\zeta}\mathrm{d}t, \tag{8.34}$$

which, substituted into (8.33), yields the final result containing the one–dimensional integral

$$\int \frac{\omega_a^*(\mathbf{r})\omega_b(\mathbf{r})}{|\mathbf{r}-\mathbf{C}|}\mathrm{d}\mathbf{r} = \frac{2\pi}{\zeta}\mathcal{K}_\mathrm{P}\exp\left(-\frac{\alpha\beta}{\zeta}|\mathbf{AB}|^2\right)\int_0^1 \exp\left(-\zeta|\tilde{\mathbf{P}}-\mathbf{C}|^2t^2\right)\mathrm{d}t. \tag{8.35}$$

The derivation of this transformation is extensively detailed in the literature, for example in Refs. 281,291,292 and can be shown in several different ways.

The molecular incomplete gamma function, also known as the Boys Function, is formally defined as[293]

$$\mathrm{F}_\mathrm{m}(z) = \int_0^1 t^{2\mathrm{m}}\exp\left(-zt^2\right)\mathrm{d}t, \tag{8.36}$$

thus is a transcendental function related to the error function `erf` by

$$\mathrm{F}_0(z) = \sqrt{\frac{\pi}{4z}}\mathrm{erf}\left(\sqrt{z}\right) \tag{8.37}$$

and identified as a scaled form of Kummer's confluent hypergeometric function $M(a,b,z)$[294],

$$\mathrm{F}_\mathrm{m}(z) = \frac{1}{2\mathrm{m}+1}M\left(\mathrm{m}+\frac{1}{2},\mathrm{m}+\frac{3}{2},-z\right) \quad m > -\frac{1}{2}. \tag{8.38}$$

Differentiation of the Boys Function with respect to the argument yields

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{F}_\mathrm{m}(z) = -\mathrm{F}_{\mathrm{m}+1}(z), \tag{8.39}$$

from which the following recurrence relation can be derived using integration by parts,

$$\begin{aligned}\mathrm{F}_\mathrm{m}(z) &= \left[\frac{t^{2\mathrm{m}+1}}{2\mathrm{m}+1}\exp\left(-zt^2\right)\right]_{t=0}^{t=1} - \int_0^1 \frac{-2zt^{2\mathrm{m}+2}}{2\mathrm{m}+1}\exp\left(-zt^2\right)\mathrm{d}t \\ &= \frac{1}{2\mathrm{m}+1}\left\{\exp(-z) + 2z\int_0^1 t^{2(\mathrm{m}+1)}\exp\left(-zt^2\right)\mathrm{d}t\right\} \\ &= \frac{1}{2\mathrm{m}+1}\left\{\exp(-z) + 2z\mathrm{F}_{\mathrm{m}+1}(z)\right\}.\end{aligned} \tag{8.40}$$

Repeated application of this recursion relation yields a series expansion that may be used to approximate the Boys Function[291]

$$\mathrm{F}_\mathrm{m}(z) = \exp(-z)\sum_{i=0}^{\infty}\frac{(2\mathrm{m}-1)!!(2z)^i}{(2\mathrm{m}+2i+1)!!}. \tag{8.41}$$

The expansion in (8.41) provides a numerically stable means of approximating $\mathrm{F}_\mathrm{m}(z)$ for a relatively wide range of arguments, z. However, the series converges rapidly and thus the approximation is most suited for smaller arguments $|z| \leqslant 25.0$. At larger z, corresponding to well–separated shell pairs that are increasingly prevalent in large systems, the Boys Function may be efficiently computed by the much simpler asymptotic approximation,

$$\mathrm{F}_\mathrm{m}(z) \approx \frac{(2\mathrm{m}-1)!!}{2^{\mathrm{m}+1}}\sqrt{\frac{\pi}{z^{2\mathrm{m}+1}}}. \tag{8.42}$$

For convenience, the following intermediate function is defined that combines the Boys Function with a pre–factor by which it is always multiplied

$$\mathrm{G_m}(z) = \sqrt{\frac{2}{\pi}}\mathrm{F_m}(z). \tag{8.43}$$

The principle difference in approximating the Boys Function for integrals over LAOs as rather than standard GTOs is that the argument z is generally complex with LAOs, whilst it is always real with GTOs. With a complex argument, the standard methods of approximation may be less stable numerically and become unreliable, thus a careful approach is required. This problem has been extensively studied in Refs. 295–297 and a multitude of methods examined for different ranges of z.

In the present work, (8.41) & (8.42) have been tested along with several from Ref. 296 against a range of complex argument z. There were no difficulties with numerical instability observed in the implementation of (8.41) & (8.42) and no improvement was observed by using alternative methods. At present, a combination of (8.41) & (8.42) appear sufficient to provide a reliable method of approximating the Boys Function for complex argument.

With a stable method of approximating the Boys Function available, shell–pair data makes it simple to generalise the OS recursion relation for nuclear attraction integrals over GTOs[266] to those over LAOs. The OS recursion relation requires a set of auxiliary integrals, derived in detail in Ref. 266, to enable the incrementation of angular momentum for each function. For each primitive integral, the shell–pair data and nuclear position are used to calculate the parameters,

$$\mathbf{PC} = \tilde{\mathbf{P}} - \mathbf{C} \qquad R^2 = |\mathbf{PC}|^2 \qquad Z_{\mathrm{PC}} = \zeta R^2, \tag{8.44}$$

from which the auxiliary integrals are constructed as

$$\left[\mathbf{0}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{0}\right]^{(\mathrm{m})} = \begin{cases} \tilde{U}_{\mathrm{P}}\,(2\zeta)^{\mathrm{m}+\frac{1}{2}}\,\mathrm{G_m}\,(Z_{\mathrm{PC}}) & |Z_{\mathrm{PC}}| \leqslant 25.0 \\ \tilde{U}_{\mathrm{P}}\,\frac{(2\mathrm{m}-1)!!}{R^{2\mathrm{m}+1}} & |Z_{\mathrm{PC}}| > 25.0 \end{cases} \tag{8.45}$$

The OS recursion relation for the nuclear attraction integral has the form

$$\begin{aligned} \left[\mathbf{a}+\mathbf{1}_i\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m})} &= \mathbf{PA}_i\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m})} - \mathbf{PC}_i\left(\frac{1}{2\zeta}\right)\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m}+1)} \\ &+ a_i\left(\frac{1}{2\zeta}\right)\left\{\left[\mathbf{a}-\mathbf{1}_i\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m})} - \left(\frac{1}{2\zeta}\right)\left[\mathbf{a}-\mathbf{1}_i\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m}+1)}\right\} \\ &+ b_i\left(\frac{1}{2\zeta}\right)\left\{\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}-\mathbf{1}_i\right]^{(\mathrm{m})} - \left(\frac{1}{2\zeta}\right)\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}-\mathbf{1}_i\right]^{(\mathrm{m}+1)}\right\} \\ \left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}+\mathbf{1}_i\right]^{(\mathrm{m})} &= \mathbf{PB}_i\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m})} - \mathbf{PC}_i\left(\frac{1}{2\zeta}\right)\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m}+1)} \\ &+ a_i\left(\frac{1}{2\zeta}\right)\left\{\left[\mathbf{a}-\mathbf{1}_i\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m})} - \left(\frac{1}{2\zeta}\right)\left[\mathbf{a}-\mathbf{1}_i\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}\right]^{(\mathrm{m}+1)}\right\} \\ &+ b_i\left(\frac{1}{2\zeta}\right)\left\{\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}-\mathbf{1}_i\right]^{(\mathrm{m})} - \left(\frac{1}{2\zeta}\right)\left[\mathbf{a}\left|\frac{1}{\mathbf{r}_C}\right|\mathbf{b}-\mathbf{1}_i\right]^{(\mathrm{m}+1)}\right\} \end{aligned} \tag{8.46}$$

where the superscript index m denotes an m^{th}–order auxiliary integral; final integrals have $\mathrm{m} = 0$. Once these have been computed in the primitive Cartesian Gaussian basis for each shell–pair and nucleus, the integral batch is contracted according to (8.3) and transformed into the spherical Gaussian basis if necessary.

8.4 Two Electron Integrals

The two–electron integrals present a much greater computational task than do the one–electron integrals; they are both individually more complex to evaluate than, and outnumber by orders of magnitude, their one–electron counterparts. In addition, ERIs over LAOs have a lower order of permutational symmetry than those over standard GTOs, with twice as many needing to be computed. Their efficient evaluation is therefore of much importance. We consider LAO–ERIs of the form

$$(\mathbf{ab}|\mathbf{cd}) = \iint \frac{\omega_a^*(\mathbf{r}_1)\omega_b(\mathbf{r}_1)\omega_c^*(\mathbf{r}_2)\omega_d(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2, \tag{8.47}$$

where the presence of the Coulomb operator again requires a transformation similar to that for nuclear attraction integrals, detailed in the literature[281,291,292]. A common approach is to introduce the molecular incomplete gamma function and construct a set of auxiliary intermediates, analogous to that described in subsection 8.3.3, then apply recursion relations to generate the integrals from these auxiliaries. A long–established alternative to this is to generate the *integrands* from unity by recursion and evaluate the integral using a Gaussian quadrature with the Rys weight function[271,272], discussed in detail in subsection 8.4.3.

Several algorithms by which these ERIs may be computed are set–out here, each having their own advantages and disadvantages. They do however all require the same set of intermediate quantities for each shell–quartet[281],

$$\mathbf{PQ} = \tilde{\mathbf{P}} - \tilde{\mathbf{Q}} \qquad R^2 = |\mathbf{PQ}|^2 \qquad U_{\mathrm{PQ}} = \tilde{U}_{\mathrm{P}}\tilde{U}_{\mathrm{Q}} \qquad Z_{\mathrm{PQ}} = \vartheta R^2 \qquad 2\vartheta = \left\{\frac{1}{2\zeta} + \frac{1}{2\eta}\right\}^{-1}, \tag{8.48}$$

all of which are readily constructed from shell–pair data, where η, $\tilde{U}_{\mathrm{Q}}$ and $\tilde{\mathbf{Q}}$ are the second shell–pair equivalent of ζ, $\tilde{U}_{\mathrm{P}}$ and $\tilde{\mathbf{P}}$ respectively.

8.4.1 The McMurchie–Davidson Algorithm

The algorithm that is most established for practical use in calculating ERIs over LAOs is the MD algorithm[273], in which charge distributions are expanded in *Hermite Gaussian functions*, defined in the absence of a complex phase factor as

$$\Lambda_{\mathbf{n}}(\zeta, \mathbf{P}) = \frac{\partial^{n_x}}{\partial P_x^{n_x}} \frac{\partial^{n_y}}{\partial P_y^{n_y}} \frac{\partial^{n_z}}{\partial P_z^{n_z}} \exp\left(-\zeta|\mathbf{r} - \mathbf{P}|^2\right), \tag{8.49}$$

where $\mathbf{P}$ is the Gaussian product centre as defined in (8.5) and which were shown by MD to satisfy the recurrence relation

$$(\mathbf{r}_i - \mathbf{A}_i)\Lambda_{\mathbf{n}}(\zeta, \mathbf{P}) = n_i\Lambda_{\mathbf{n}-\mathbf{1}_i}(\zeta, \mathbf{P}) + \mathbf{PA}_i\Lambda_{\mathbf{n}}(\zeta, \mathbf{P}) + \frac{1}{2\zeta}\Lambda_{\mathbf{n}+\mathbf{1}_i}(\zeta, \mathbf{P}). \tag{8.50}$$

Writing the charge distribution as the sum of derivatives of Hermite Gaussian functions with respect to $\mathbf{P}$ allows these terms to be taken outside of the integral over electron coordinates as

$$\begin{aligned}\int\int \frac{\Lambda_{tuv}(\mathbf{P}, \mathbf{r}_1)\Lambda_{\tau\upsilon\nu}(\mathbf{Q}, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 &= \frac{\partial^t}{\partial P_x^t}\frac{\partial^u}{\partial P_y^u}\frac{\partial^v}{\partial P_z^v}\frac{\partial^\tau}{\partial Q_x^\tau}\frac{\partial^\upsilon}{\partial Q_y^\upsilon}\frac{\partial^\nu}{\partial Q_z^\nu}\int\int \frac{\exp\left(-\zeta|\mathbf{r}_1 - \mathbf{P}|^2\right)\exp\left(-\eta|\mathbf{r}_2 - \mathbf{Q}|^2\right)}{|\mathbf{r}_1 - \mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\ &= \frac{\partial^{t+\tau}}{\partial P_x^{t+\tau}}\frac{\partial^{u+\upsilon}}{\partial P_y^{u+\upsilon}}\frac{\partial^{v+\nu}}{\partial P_z^{v+\nu}}(-1)^{\tau+\upsilon+\nu}\int\int \frac{\exp\left(-\zeta|\mathbf{r}_1 - \mathbf{P}|^2\right)\exp\left(-\eta|\mathbf{r}_2 - \mathbf{Q}|^2\right)}{|\mathbf{r}_1 - \mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2,\end{aligned} \tag{8.51}$$

where the reduction to differentiation over only one index follows from the property

$$\frac{\partial^n}{\partial P_x^n}\mathrm{F}_0(Z_{\mathrm{PQ}}) = -\frac{\partial^n}{\partial Q_x^n}\mathrm{F}_0(Z_{\mathrm{PQ}}). \tag{8.52}$$

To calculate the integrals over Hermite Gaussian functions, MD proposed the one–centre recursion relation over the set of auxiliary integrals, constructed in (8.54), with subsequent transformation to the required two–centre integral as

$$[\mathbf{r} + \mathbf{1}_i]^{(\mathrm{m})} = \mathbf{PQ}_i\,[\mathbf{r}]^{(\mathrm{m}+1)} + r_i\,[\mathbf{r} - \mathbf{1}_i]^{(\mathrm{m}+1)} \qquad [\mathbf{p}|\mathbf{q}] = (-1)^{\mathbf{q}}\,[\mathbf{p} + \mathbf{q}]^{(0)}\,. \tag{8.53}$$

Integrals computed over the Hermite Gaussian functions are then transformed back into the Cartesian basis using the recurrence relation in (8.56) and subsequently into the spherical basis if required.

This approach is effective in the absence of an external field, as the use of Hermite functions allows the four centre integral to be reduced to just one centre, resulting in a recurrence relation for incrementing angular momentum with only two terms. The result of this is that the recursion step in the MD algorithm becomes more advantageous with higher total angular momentum, however the MD algorithm is disadvantaged by the substantial computational cost of transforming the Hermite integrals back to the Cartesian basis.

The expansion of AOs with a complex phase factor in a Hermite Gaussian basis has been the most widely practised approach to evaluation of the necessary integrals; Colle *et. al.* presented general formulae for integrals over Hermite Gaussian functions with complex phase factors in Refs. 298,299, noting that transforming the integrals to the Cartesian basis would be a trivial extension. Further progress has been made by Tachikawa and co–workers[274,275] and by Tellgren *et. al*[172,256] to generalise the MD algorithm for the evaluation of integrals over Gaussian functions with a complex phase–factor.

In the interests of comparison, the generalised MD algorithm was implemented as part of the present study, broadly following the scheme detailed in Ref. 172, but adapted to the context of the shell–pair scheme. The zeroth–order Hermite integrals are computed from the molecular incomplete gamma function as

$$[\mathbf{0}]^{(\mathrm{m})} = \begin{cases} U_{\mathrm{PQ}}(-1)^{\mathrm{m}}\,(2\vartheta)^{\mathrm{m}+\frac{1}{2}}\,\mathrm{G}_{\mathrm{m}}\left(Z_{\mathrm{PQ}}\right) & |Z_{\mathrm{PQ}}| \leqslant 25.0 \\ U_{\mathrm{PQ}}(-1)^{\mathrm{m}}\,\frac{(2\mathrm{m}-1)!!}{R^{2\mathrm{m}+1}} & |Z_{\mathrm{PQ}}| > 25.0 \end{cases} \tag{8.54}$$

from which two–centre Hermite integrals of higher angular momentum are calculated recursively as

$$\begin{aligned} [\mathbf{p}+\mathbf{1}_i|\mathbf{q}]^{(\mathrm{m})} &= -i\chi_{\mathrm{P},i}\,[\mathbf{p}|\mathbf{q}]^{(\mathrm{m})} + \mathbf{PQ}_i\,[\mathbf{p}|\mathbf{q}]^{(\mathrm{m}+1)} + p_i\,[\mathbf{p}-\mathbf{1}_i|\mathbf{q}]^{(\mathrm{m}+1)} - q_i\,[\mathbf{p}|\mathbf{q}-\mathbf{1}_i]^{(\mathrm{m}+1)} \\ [\mathbf{p}|\mathbf{q}+\mathbf{1}_i]^{(\mathrm{m})} &= -i\chi_{\mathrm{Q},i}\,[\mathbf{p}|\mathbf{q}]^{(\mathrm{m})} - \mathbf{PQ}_i\,[\mathbf{p}|\mathbf{q}]^{(\mathrm{m}+1)} - p_i\,[\mathbf{p}-\mathbf{1}_i|\mathbf{q}]^{(\mathrm{m}+1)} + q_i\,[\mathbf{p}|\mathbf{q}-\mathbf{1}_i]^{(\mathrm{m}+1)}\,. \end{aligned} \tag{8.55}$$

When considering non–zero external field, a consequence of the complex phase factor in LAOs is that the four centre integral can no longer be simplified to a one centre integral over Hermite Gaussian functions using (8.52); instead it can only be reduced to a two centre integral, with the recursion relation used to increment angular momentum comprising double the number of terms[172]. Thus the principle advantage of this algorithm is lost, whilst the costly transformation from Hermite to Cartesian Gaussians remains necessary and is given by

$$\begin{aligned} [\mathbf{a}+\mathbf{1}_i\,\mathbf{b}\,\mathbf{p}| &= p_i\,[\mathbf{a}\,\mathbf{b}\,\mathbf{p}-\mathbf{1}_i| + \mathbf{PA}_i\,[\mathbf{a}\,\mathbf{b}\,\mathbf{p}| + \left(\frac{1}{2\zeta}\right)[\mathbf{a}\,\mathbf{b}\,\mathbf{p}+\mathbf{1}_i| \\ |\mathbf{q}\,\mathbf{c}+\mathbf{1}_i\,\mathbf{d}] &= q_i\,|\mathbf{q}-\mathbf{1}_i\,\mathbf{c}\,\mathbf{d}] + \mathbf{QC}_i\,|\mathbf{q}\,\mathbf{c}\,\mathbf{d}] + \left(\frac{1}{2\eta}\right)|\mathbf{q}+\mathbf{1}_i\,\mathbf{c}\,\mathbf{d}]\,. \end{aligned} \tag{8.56}$$

These transformations are unchanged from those in the zero–field algorithm, allowing any techniques developed to optimise this step to be applied to integrals over LAOs. In particular, the Hermite to Cartesian recursion of (8.56) and the contraction stage of (8.3) be applied to one index of the Hermite integral at a time, improving efficiency as (8.56) can then be applied to the other index with the integral already half contracted.

8.4.2 The Head–Gordon Pople Algorithm

In view of the apparent shortcomings of the MD algorithm when applied to integrals over LAOs, this work has sought to identify and implement alternative methods that will allow the integrals to be computed with greater efficiency.

The first of these to be considered is an adaptation of the HGP algorithm, itself a modification of the OS scheme[281,284]. In this approach, the zeroth–order auxiliary integrals are constructed in much the same way as for the nuclear attraction integrals, from the molecular incomplete gamma function scaled with the requisite pre–factors as

$$[\mathbf{0}]^{(\mathrm{m})} = \begin{cases} U_{\mathrm{PQ}}\,(2\vartheta)^{\mathrm{m}+\frac{1}{2}}\,\mathrm{G}_{\mathrm{m}}\left(Z_{\mathrm{PQ}}\right) & |Z_{\mathrm{PQ}}| \leqslant 25.0 \\ U_{\mathrm{PQ}}\,\frac{(2\mathrm{m}-1)!!}{R^{2\mathrm{m}+1}} & |Z_{\mathrm{PQ}}| > 25.0 \end{cases} \tag{8.57}$$

From these auxiliary integrals, angular momentum can be incremented at each of the indices by applying the eight–term OS recursion relation[266]. In the HGP algorithm, this is significantly simplified by building up angular momentum at only two indices, reducing the number of terms present in the recursion relation from eight to five, given by

$$\begin{aligned} [\mathbf{e}+\mathbf{1}_i\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m})} &= \mathbf{PA}_i\,[\mathbf{e}\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m})} - \mathbf{PQ}_i\left(\frac{1}{2\zeta}\right)[\mathbf{e}\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m}+1)} \\ &\quad + e_i\left(\frac{1}{2\zeta}\right)\left\{[\mathbf{e}-\mathbf{1}_i\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m})} - \left(\frac{1}{2\zeta}\right)[\mathbf{e}-\mathbf{1}_i\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m}+1)}\right\} \\ &\quad + f_i\left(\frac{1}{2\zeta}\right)\left(\frac{1}{2\eta}\right)[\mathbf{e}\mathbf{0}|\mathbf{f}-\mathbf{1}_i\mathbf{0}]^{(\mathrm{m}+1)} \\ [\mathbf{e}\mathbf{0}|\mathbf{f}+\mathbf{1}_i\mathbf{0}]^{(\mathrm{m})} &= \mathbf{QC}_i\,[\mathbf{e}\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m})} + \mathbf{PQ}_i\left(\frac{1}{2\eta}\right)[\mathbf{e}\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m}+1)} \\ &\quad + f_i\left(\frac{1}{2\eta}\right)\left\{[\mathbf{e}\mathbf{0}|\mathbf{f}-\mathbf{1}_i\mathbf{0}]^{(\mathrm{m})} - \left(\frac{1}{2\eta}\right)[\mathbf{e}\mathbf{0}|\mathbf{f}-\mathbf{1}_i\mathbf{0}]^{(\mathrm{m}+1)}\right\} \\ &\quad + e_i\left(\frac{1}{2\eta}\right)\left(\frac{1}{2\zeta}\right)[\mathbf{e}-\mathbf{1}_i\mathbf{0}|\mathbf{f}\mathbf{0}]^{(\mathrm{m}+1)}\,, \end{aligned} \tag{8.58}$$

where $L_e = \{L_a, L_a + L_b\}$ and $L_f = \{L_c, L_c + L_d\}$. A further significant advantage of the HGP scheme is that the integrals may be contracted at this stage, before applying the HRR of subsection 8.2.2, yielding the contracted intermediate

$$(\mathbf{e0}|\mathbf{f0}) = \sum_{K_a}\sum_{K_b}\sum_{K_c}\sum_{K_d}[\mathbf{e0}|\mathbf{f0}]^{(0)}. \tag{8.59}$$

With the pre–computed transformation matrices of (8.10), the HRR (in combination with spherical transformation if necessary) may be simply applied as

$$\begin{aligned}(\mathbf{e0}|\mathbf{cd}) &= \sum_{\mathbf{f}}(\mathbf{e0}|\mathbf{f0})\,(\mathbf{f}|\mathbf{cd})\\ (\mathbf{ab}|\mathbf{cd}) &= \sum_{\mathbf{e}}(\mathbf{ab}|\mathbf{e})\,(\mathbf{e0}|\mathbf{cd})\end{aligned} \tag{8.60}$$

where $(\mathbf{ab}|\mathbf{e})$ and $(\mathbf{f}|\mathbf{cd})$ are the transformation matrices for the each shell–pair respectively, thus providing a significant increase in efficiency when the contraction length of the shell–pairs is high. The matrix multiplication approach to the HRR employed in this work enhances these gains by making use of optimised math libraries.

8.4.3 The Rys Polynomial Approach

As a second alternative to the MD scheme, we also consider the generalisation of Rys quadrature for the calculation of LAO-ERIs. According to the Gauss quadrature theory, a definite integral of a polynomial $P(x)$ of order up to $2N-1$ weighted by a function $W(x)$ can be computed exactly by summing over N weights w_i multiplied by $P(x)$ evaluated at N roots x_i of a polynomial $p_N(x)$[278]:

$$\int_a^b P(x)W(x)\mathrm{d}x = \sum_{i=1}^{N} P(x_i)w_i. \tag{8.61}$$

The algorithm to obtain the necessary roots and weights is described in subsubsection 8.4.3.1. In order to evaluate an ERI using (8.61), we have to identify the polynomial and weight function components of its integrand.

Following a Laplace transformation of the Coulomb operator, given in (8.31), the integral can be written as

$$(\mathbf{ab}|\mathbf{cd}) \propto U_{\mathrm{PQ}}\int_0^1 \exp\left(-t^2 Z_{\mathrm{PQ}}\right)\mathcal{I}_x(t)\mathcal{I}_y(t)\mathcal{I}_z(t)\mathrm{d}t, \tag{8.62}$$

where $\mathcal{I}_x(t)$, $\mathcal{I}_y(t)$, $\mathcal{I}_z(t)$ are two–dimensional integrals over Cartesian components which depend on the angular momenta and exponents of the quartet. It can be shown that these 2D integrals are polynomials in t^2. Moreover, after a substitution $x = t^2$ we can identify the weight function of (8.61) as the exponential part of the integrand in (8.62), the so called Rys weight function[272], $W(x) = \exp(-xZ_{\mathrm{PQ}})/\sqrt{x}$. Hence, the ERI for any shell quartet can be evaluated using the Gauss quadrature technique provided that a set of roots and corresponding weights can be obtained for any value of the argument Z_{PQ}.

In this work, we adapt the algorithm to generate real roots and weights of Golub and Welsh[300] and of Flocke[278] for complex Gauss quadrature needed to compute ERIs over LAOs.

8.4.3.1 The Gauss Quadrature Rules

For a fixed weight function $W(x) \geqslant 0$ on $x \in [a, b]$, it is possible to define a sequence of real polynomials $p_0(x)$, $p_1(x)$, …, $p_n(x)$ of the form $p_n(x) = k_n\prod_{j=1}^{n}(x - x_j)$, $k_n > 0$, which are orthogonal with respect to the weight function,

$$\langle p_m(x)|p_n(x)\rangle_w \equiv \int_a^b p_m(x)p_n(x)W(x)\mathrm{d}x = 0, \quad m \neq n. \tag{8.63}$$

We note that a system of polynomials orthogonal with respect to the Rys weight function are called the Rys polynomials[271]. Any set of real orthogonal polynomials $\{p_n(x)\}_{n=1}^{N}$ can be constructed using a three–term recurrence relation,

$$p_n(x) = (a_n x - b_n)p_{n-1}(x) - c_n p_{n-2}(x) \quad n > 1 \tag{8.64}$$

where $p_{-1}(x) \equiv 0$ and $p_0(x) \equiv 1$. The two conditions that p_n is orthogonal to p_{n-1} and p_{n-2} do not determine the three coefficients a_n, b_n, and c_n uniquely. Different choices of the third condition are discussed in literature resulting

in monic polynomials[278] (the coefficient of the largest power of x is equal to 1) or orthonormal polynomials used in this work and in Ref. 301.

In order to find N roots $\{x_i\}$ of $p_N(x)$ required in (8.61), the recurrence relation of (8.64) can be written in a matrix form as

$$x\mathbf{p}(x) = T\mathbf{p}(x) + p_N(x)\mathbf{e}_N, \tag{8.65}$$

where $\mathbf{p}(x) = [p_0(x), p_1(x), \ldots, p_{N-1}(x)]^T$, $\mathbf{e}_N = [0, 0, \ldots, 1]^T$ and T is the tridiagonal matrix

$$T = \begin{bmatrix} b_1/a_1 & 1/a_1 & & & \\ c_2/a_2 & b_2/a_2 & 1/a_2 & & \\ & c_3/a_3 & b_3/a_3 & \ldots & \\ & & \ldots & \ldots & 1/a_{N-1} \\ & & & c_N/b_N & b_N/a_N \end{bmatrix}, \tag{8.66}$$

which becomes symmetric in case of an orthonormal polynomial set $\{q_n(x)\}_{n=1}^N$ with $q_0(x) \equiv (\int_a^b W(x)\mathrm{d}x)^{-1/2}$ chosen to satisfy $\langle q_0(x)|q_0(x)\rangle_w = 1$ (see Refs. 301,302 for a more detailed discussion). A transition to recursion coefficients generating an orthonormal sequence of polynomials corresponds to a similarity transformation to a Jacobi matrix $J = DTD^{-1}$, where

$$J = \begin{bmatrix} \alpha_0 & \beta_1 & & & \\ \beta_1 & \alpha_1 & \beta_2 & & \\ & \beta_2 & \alpha_2 & \ldots & \\ & & \ldots & \ldots & \beta_{N-1} \\ & & & \beta_{N-1} & \alpha_{N-1} \end{bmatrix}, \tag{8.67}$$

with $\alpha_n = b_n/a_n$ and $\beta_n = (-a_n a_{n+1}/c_{n+1})^{-1/2}$.

In analogy to (8.65), we can write a matrix equation $x\mathbf{q}(x) = J\mathbf{q}(x) + q_N(x)\mathbf{e}_N$. Solving the characteristic equation $x_i\mathbf{q}(x_i) = J\mathbf{q}(x_i)$ leads to $q_N(x_i) = 0$, hence the eigenvalues of matrix J (and T) are the required roots.

The matrix J is symmetric and real (i.e. Hermitian) so its set of eigenvectors $\{\mathbf{q}(x_i)\}_{i=1}^N$ associated with eigenvalues x_i is orthogonal with respect to the standard inner product

$$\langle \mathbf{q}(x_i)|\mathbf{q}(x_j)\rangle \equiv \sum_{n=1}^{N} \mathbf{q}_n(x_i)\overline{\mathbf{q}_n(x_j)} = 0, \quad i \neq j. \tag{8.68}$$

However, these eigenvectors are not normalised by virtue of the method with which they are constructed. For the corresponding weights w_i, as a consequence of the Christoffel–Darboux identity, we have[300,302]: $w_i \sum_{n=1}^N \mathbf{q}_n^2(x_i) = 1$.

We can find an orthonormal set of eigenvectors $\{\mathbf{u}_i\}_{i=1}^N$, forming columns of a unitary matrix U which satisfies the identity $U^\dagger U = E$ in general and $U^T U = E$ in the case of real eigenvectors. The combination of this and the Christoffel–Darboux identity allows us to write an explicit expression for the weights w_i as

$$w_i = \frac{\mathbf{u}_{j,i}^2}{\mathbf{q}_j^2(x_i)} \qquad j = 1, \ldots, N. \tag{8.69}$$

Finally, we can use the fact that the first component of any eigenvector $\mathbf{q}(x_i)$ is a constant

$$\mathbf{q}_1(x_i) = q_0 \equiv \left(\int_a^b W(x)\mathrm{d}x\right)^{-1/2} \tag{8.70}$$

and obtain $w_i = u_{1,i}^2 \int_a^b W(x)\mathrm{d}x$. The problem of evaluating the ERIs can therefore be reduced to that of obtaining the recurrence coefficients a_n, b_n, and c_n.

Unfortunately, the Rys polynomials do not belong to any classical sets of orthogonal polynomials with known recurrence coefficients. Several alternative Rys quadrature methods are reviewed in Ref. 278, highlighting the efficiency and numerical stability of the modified moment method[303] for real positive values of the argument Z_{PQ}. Here we adopt this method for complex values of Z_{PQ}, which occur for integrals over LAOs in the presence of homogeneous magnetic field. The justification of this approach, based on the work of Saylor and Smolarski[301], is outlined briefly below.

8.4.3.2 The Complex Rys Quadrature

In general, there is no three–term recurrence relation that may be applied to generate a set of complex polynomials orthogonal with respect to the standard weighted inner product,

$$\langle p_m(x)|p_n(x)\rangle_w \equiv \int_\gamma p_m(x)\overline{p_n(x)}W(x)\mathrm{d}x = 0, \quad m \neq n, \tag{8.71}$$

where $\overline{p_n(x)}$ indicates complex conjugation and γ is an arc in the complex plane[301]. However, we can construct a set of *formally orthogonal* (sometimes referred to as *conjugate orthogonal*) polynomials with respect to a bilinear form,

$$[p_m(x)|p_n(x)]_w \equiv \int_\gamma p_m(x)p_n(x)W(x)\mathrm{d}x = 0, \quad m \neq n. \tag{8.72}$$

As a result, we obtain a tridiagonal complex symmetric Jacobi matrix. The required roots can be calculated as eigenvalues of this matrix, though they are complex[301]. We emphasise that the eigenvectors of the non–Hermitian Jacobi matrix are *not* orthogonal with respect to the standard inner product of (8.68), but they are orthogonal with respect to a bilinear form,

$$[\mathbf{v}_i|\mathbf{v}_j] \equiv \sum_k \mathbf{v}_{k,i}\mathbf{v}_{k,j} = 0, \qquad i \neq j. \tag{8.73}$$

The normalisation of the eigenvectors of the Jacobi matrix is as crucial for calculation of the weights of the complex Gauss quadrature as in the real case. We can find a formally orthonormal set of eigenvectors $\{\mathbf{v}_i\}_{i=1}^N$, which form columns of an orthogonal matrix[304] V that satisfies $V^TV = E$, exactly as in the real case. However, for complex eigenvectors this identity implies normalisation with respect to the bilinear form, $[\mathbf{v}_i|\mathbf{v}_i] = 1$.

We note that standard eigenvalue solvers such as subroutine zgeev from the LAPACK library[305] return eigenvectors that are individually normalised using the standard inner product, the largest component of which are chosen to be real. However, these eigenvectors cannot be made orthonormal with respect to this inner product so they cannot be used in (8.69) to find the required weights. We therefore perform an additional normalisation based on the bilinear form above. The whole set then satisfies the condition $V^TV = E$, as in the real case, thus can be used to compute the complex Gauss quadrature weights using standard double precision arithmetic.

For $N < 12$, the recurrence coefficients needed to construct the complex Jacobi matrix can be obtained using the moment method (also known as the Chebyshev algorithm) following the work of Golub and Welsh[300]. We have observed that this method, based on integrating x^n ($n = 1, \ldots, N$) over the Rys weight, becomes numerically unstable around $N = 12$ both in case of a real and complex argument Z_{PQ}.

One approach to accommodate the changes in electronic structure induced by the application of a magnetic field is to include a greater number of basis functions of high angular momenta. As a result, we choose to use the modified moment method as outlined by Flocke for real Rys quadrature[278], which remains numerically stable well beyond $N = 12$. The method uses multiple sets of auxiliary polynomials with known coefficients of their three-term recurrence relation. The choice of the set depends on the size of the argument Z_{PQ}. We have found that we can cover the whole interval relevant for ERIs arising in typical basis sets and magnetic fields up to one atomic unit using the shifted Jacobi polynomials for $|Z_{\mathrm{PQ}}| < 30$ and generalised T-scaled Laguerre polynomials for $|Z_{\mathrm{PQ}}| > 30$. This approach can be used for complex numbers without modifications.

We have found that the absolute value of the complex argument can be considered instead of its positive real component when deciding which set of auxiliary polynomials to use. In case of very large $|Z_{\mathrm{PQ}}|$, it becomes computationally advantageous to use the classical Hermite polynomials[272], where the roots and weights are precomputed and rescaled by $\sqrt{|Z_{\mathrm{PQ}}|}$, as this approximation becomes increasingly accurate in the limit of large $|Z_{\mathrm{PQ}}|$. The calculation of the first component of eigenvectors q_0 required to obtain the complex weights is described in subsection 8.3.3; it is simply the zeroth order molecular incomplete gamma function.

Finally, we compare the behaviour of the standard and modified moment methods close to singularities of the elements of the Jacobi matrix as a function of Z_{PQ} in the complex plane, as indicated by Reynolds and Shiozaki[261] for $N = 2$. We observe that the relative error of a test integral utilising the relevant roots and weights does not exceed 10^{-10} when we do not approach the singularity Z^s_{PQ} closer than $|Z_{\mathrm{PQ}} - Z^s_{\mathrm{PQ}}| \sim 10^{-9}$. In practical calculations and testing carried out during this work with field strengths less than 3 atomic units, we have encountered no practical issues associated with such singularities. In the event that such singularities are encountered, the integral batch may be recomputed with a larger number of roots and weights, or using the HGP or MD algorithms.

8.4.3.3 Vertical Recursion Relation

In the Rys polynomial scheme, the zeroth order terms are not computed from the scaled molecular incomplete gamma function, but from the standard Gaussian pre–factors and the Rys quadrature weights w_λ as

$$\mathcal{I}_x(0,0;\lambda) = 1.0 \qquad \mathcal{I}_y(0,0;\lambda) = 1.0 \qquad \mathcal{I}_z(0,0;\lambda) = U_{\mathrm{PQ}}\sqrt{\frac{4\vartheta}{\pi}}w_\lambda \tag{8.74}$$

where the integrand is resolved into the three Cartesian components. For higher order integrals, angular momentum can be incremented using the following recursion relations which are analogous to the HGP vertical recursion relation as shown by Lindh *et. al.*[306]

$$\begin{aligned}
\mathcal{I}_i(e+1,f;\lambda) &= \left\{\mathbf{PA}_i - \frac{\eta t_\lambda^2}{\zeta+\eta}\mathbf{PQ}_i\right\}\mathcal{I}_i(e,f;\lambda) + \frac{e}{2\zeta}\left\{1-\frac{\eta t_\lambda^2}{\zeta+\eta}\right\}\mathcal{I}_i(e-1,f;\lambda) + \frac{f t_\lambda^2}{2(\zeta+\eta)}\mathcal{I}_i(e,f-1;\lambda) \\
\mathcal{I}_i(e,f+1;\lambda) &= \left\{\mathbf{QC}_i + \frac{\zeta t_\lambda^2}{\zeta+\eta}\mathbf{PQ}_i\right\}\mathcal{I}_i(e,f;\lambda) + \frac{f}{2\eta}\left\{1-\frac{\zeta t_\lambda^2}{\zeta+\eta}\right\}\mathcal{I}_i(e,f-1;\lambda) + \frac{e t_\lambda^2}{2(\zeta+\eta)}\mathcal{I}_i(e-1,f;\lambda)
\end{aligned} \tag{8.75}$$

where t_λ^2 are the Rys roots, described in (8.62). The resolution of the integrand into Cartesian components allows angular momentum to be incremented separately in each direction, resulting in a vertical recursion relation that scales much more favourably with angular momentum than either the MD or HGP algorithms.

8.4.3.4 Reduced Multiplication Scheme

Integrals suitable for HRR transformation are obtained by multiplying the relevant x, y and z components of the 2D integrals and summing over the Rys polynomial nodes,

$$[\mathbf{e0}|\mathbf{f0}] = \sum_{\lambda=1}^{N} \mathcal{I}_x(e_x,f_x;\lambda)\,\mathcal{I}_y(e_y,f_y;\lambda)\,\mathcal{I}_z(e_z,f_z;\lambda) \tag{8.76}$$

and subsequently summing over the primitive components of the contracted integral using (8.59). This summation step is generally the computational bottleneck of the Rys quadrature approach, scaling less favourably with angular momentum than the comparatively inexpensive VRR. Lindh *et. al.* developed the reduced multiplication scheme[306]; a technique effective at improving the efficiency of this step by maximising the re–use of intermediates and avoiding redundant multiplications.

In the construction of $[\mathbf{e0}|\mathbf{f0}]$ integrals, it was noted that each combination of x and y components was frequently combined with multiple z components, thus creation of an xy intermediate to be combined with many z components in summation over Rys quadrature nodes would reduce the number of individual multiplications required by the number of nodes for each re–use of the intermediate. This is demonstrated by example in Figure 8.1.

Aside from the diligent re–use of intermediates, the other main facet of the reduced multiplication scheme is the elimination of superfluous multiplications by unity, discarding $\mathcal{I}_x(0,0;\lambda)$ and $\mathcal{I}_y(0,0;\lambda)$ from summations where these occur; $\mathcal{I}_z(0,0;\lambda)$ cannot be discarded as it carries the Rys weights and other pre–factors. We have employed the reduced multiplication in the generalised Rys quadrature implementation used in this work as it is a useful enhancement to the efficiency of the overall algorithm.

$$\sum_\lambda \mathcal{I}_x(0,1;\lambda) * \mathcal{I}_y(1,0;\lambda) * \begin{cases} \mathcal{I}_z(0,0;\lambda) \longrightarrow [p_y\ s|p_x\ s] \\ \mathcal{I}_z(0,1;\lambda) \longrightarrow [p_y\ s|d_{xz}s] \\ \mathcal{I}_z(1,0;\lambda) \longrightarrow [d_{yz}s|p_x\ s] \\ \mathcal{I}_z(1,1;\lambda) \longrightarrow [d_{yz}s|d_{xz}s] \end{cases}$$

Figure 8.1: An illustration of how the reduced multiplication scheme can be effective at reducing the amount of computation required. In this example, the 2D integrals in the x and y axes are premultiplied for all λ to form an xy intermediate; this can be combined with four different z components and summed over λ to form four different integrals.

8.4.4 Cauchy–Schwarz Screening

In addition to the screening of negligible shell–pairs described in subsection 8.2.1, the Cauchy–Schwarz inequality is exploited to identify and avoid computing negligible batches of integrals,

$$|(\mathbf{ab}|\mathbf{cd})| \leqslant \sqrt{|(\mathbf{ab}|\mathbf{ba})|} \cdot \sqrt{|(\mathbf{cd}|\mathbf{dc})|}. \tag{8.77}$$

In this work, the threshold for screening at this level is equal to that for screening individual shell–pairs, 10^{-12}.

8.5 Assessing The Efficiency Of LAO-ERI Algorithms

The three ERI algorithms described in Section 8.4, along with the one–electron integrals of Section 8.3, have been implemented in the QUEST rapid development platform[63]. This program is written predominantly in the PYTHON language, exploiting just-in-time compilation techniques using the NUMBA compiler[119,120].

To explore the relative efficiency of the algorithms we consider two example systems, O_2 and C_4H_4. We present relative timings of the integral calculation, broken down by quartet angular momentum, for these systems in both the primitive and contracted form of Dunning's aug-cc-pCVTZ basis set[117] in the spherical harmonic representation. This choice of basis set represents a size typical of that used in production calculations and allows for an assessment of the performance of the integral algorithms over a range of angular momenta.

In Figure 8.2, integral timings are presented for calculations on the small paramagnetic molecule O_2. These calculations were configured to simulate a uniform external magnetic field of 1 atomic unit, perpendicular to the bonding axis, on the electronic state corresponding to lowest energy component of the triplet ground state of O_2 at zero–field. The left–hand panel of Figure 8.2 shows the timings pertaining to the primitive basis set, whilst the right–hand panel shows the corresponding data for the contracted basis set.

In the presence of a strong magnetic field, the electron density of the system is compressed in an anisotropic manner, with the compression perpendicular to the field being more pronounced than that in the direction of the applied field, as described for example in Refs. 257,263,264. These changes are however small relative to the changes in density upon formation of a typical covalent bond and so a simple approach to introduce sufficient flexibility into the basis set to adequately accommodate these effects is to un–contract the basis functions. The development of basis sets tailored to finite field calculations has not yet been pursued, however for such sets, contracted functions could be developed to enhance computational efficiency. In Figure 8.2 we also therefore present timings for the aug-cc-pCVTZ basis set in its contracted form, where the contractions used are those defined for zero–field calculations.

In this work, we have only considered fields of up to ~ 1 atomic unit in strength, equivalent to 235′000 Tesla. For the purposes of checking the numerical stability of the implemented code, some additional calculations were undertaken with fields of up to ~ 3 atomic units in strength. We note however that, at field strengths greater than a few atomic units, the use of anisotropic basis sets may become necessary[307,308], along with a more careful consideration of non Born–Oppenheimer corrections[309].

The timings for the primitive basis for O_2 in the left hand panels of Figure 8.2 are classified by the total angular momentum of the shell quartet, $L_{\text{bra}} + L_{\text{ket}}$. In the upper panel, the average CPU time per integral in each class is presented; this gives a system independent measure of the performance of each algorithm for the different classes of integral. As would be expected from subsection 8.4.1, the MD algorithm exhibits the least favourable performance overall. The time per integral for the MD algorithm is little different from that of the HGP algorithm for very low angular momenta of $\lesssim 2$, perhaps having a marginal advantage over HGP for these classes, however the time per integral for MD rapidly begins to exceed that of HGP for classes of increasing angular momenta. This is a result of the recursion relation of (8.55) having twice the number of terms and indices as its zero–field counterpart, effectively eliminating a principle advantage of the MD algorithm whilst retaining the principle disadvantage in the form of (8.56), leading to the relatively poor scaling with angular momentum observed here.

The HGP algorithm delivers a superior performance to the MD algorithm for integrals of angular momenta above the range $\lesssim 2$. This is unsurprising because, whilst the VRR of (8.58) has one term more than the MD recursion relation, the HRR of (8.9) is much simpler than the transformation stage of the MD algorithm, (8.56). In essence, the HGP algorithm for LAO-ERIs is little different to that over standard GTOs, thus retains many of its well documented comparative advantages[281,310].

A shortcoming that both the MD and HGP algorithms share is that their efficiency begins to deteriorate when the total angular momentum of the integral increases further, exceeding ~ 5 for primitive functions. It is clear from Figure 8.2 however that the Rys quadrature scheme performs significantly better than either the MD or HGP

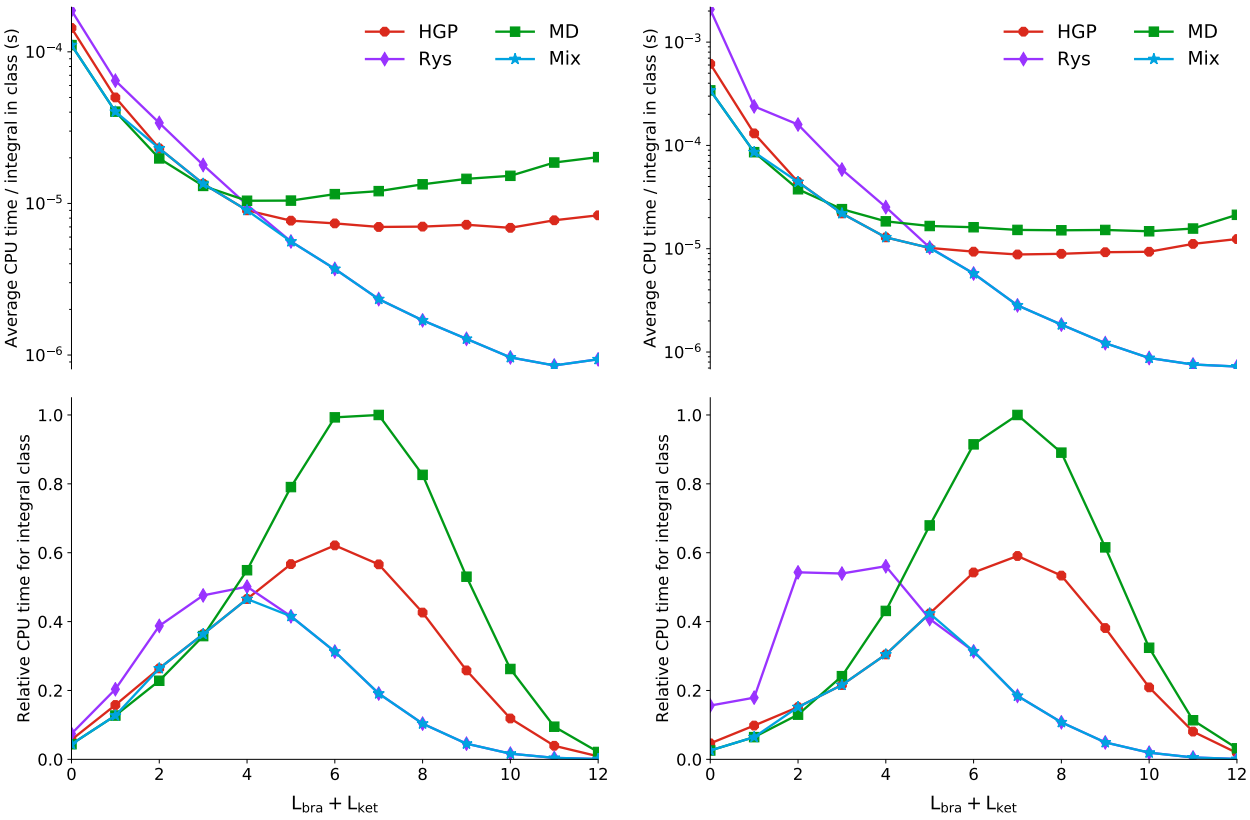

Figure 8.2: Relative timings for LAO-ERIs in a calculation on the O_2 molecule in the aug-cc-pCVTZ basis, classified according to the total angular momenta $L_{bra} + L_{ket}$ for each shell quartet. The left hand panels show the results for the primitive basis, the right hand panels for the contracted basis set.

algorithms for integrals with angular momentum $\gtrsim 5$, becoming increasingly advantageous with higher total angular momentum. This trend can be understood by comparing the HGP VRR of (8.58) and the Rys VRR of (8.75); in the Rys scheme recursion is applied to the integrand polynomial, which is separable by Cartesian components allowing angular momentum to be incremented in each component independently and resulting is a much lower scaling with angular momentum than the HGP VRR, where the intermediate auxiliary integrals are non–separable. The scaling of the Rys summation step of (8.76), the steepest scaling part of the Rys algorithm, is minimised by application of the reduced multiplication scheme as described in subsubsection 8.4.3.4.

In contrast it can also be seen in Figure 8.2 that the Rys algorithm is the least efficient for computing ERIs of very low total angular momentum, being in this case inferior to both MD and HGP. This is because the zeroth–order terms of the MD and HGP algorithms are simply scaled molecular incomplete gamma functions and are relatively inexpensive to compute, whereas the zeroth–order terms of the Rys algorithm requires the more computationally expensive calculation of quadrature nodes and weights. For integrals of low angular momentum, computing the zeroth–order term is generally the dominant step, thus the pre–factor is greater for the Rys quadrature than for the MD or HGP algorithms, impairing the performance of the Rys scheme at low angular momenta.

In the lower panels of Figure 8.2, the relative CPU time for each integral class in the calculation is presented; this plot reflects not only the angular momenta of integrals present but also their relative distribution as determined by the particular choice of basis and the system under study. It is again clear that the MD approach is the least efficient, with the largest amount of time spent on integrals with $L_{bra} + L_{ket} = 7$, the HGP algorithm delivers a considerable improvement for $L_{bra} + L_{ket} > 4$, and is comparable with MD for lower angular momenta integrals. For the HGP approach, the most time is spent on integrals with $L_{bra} + L_{ket} = 6$. Most striking is the significant improvement offered by the use of Rys quadrature for $L_{bra} + L_{ket} > 4$, with the largest amount of time being spent on integrals with $L_{bra} + L_{ket} = 4$ for this approach. Furthermore, as would be anticipated from the above discussion, for $L_{bra} + L_{ket} < 4$ the Rys approach is noticeably less efficient than either the MD or HGP schemes.

In the right hand panels of Figure 8.2, the corresponding analysis is presented for the contracted aug-cc-pCVTZ basis. Qualitatively the plots are similar, however, there are some important differences. Consistent with the discussion in subsection 8.4.2, the HGP algorithm is particularly efficient for contracted basis sets since its HRR may be applied to contracted integrals. This effect can be discerned in the upper right plot of Figure 8.2, noting that the HGP algorithm

becomes more efficient than MD earlier at $L_{\text{bra}} + L_{\text{ket}} = 2$. On the other hand, the efficiency of the Rys quadrature calculations is adversely effected by contraction, as it necessitates multiple sets of quadrature roots and weights to be computed per integral batch; this effect can be discerned from the upper right plot, noting a later crossover with HGP at $L_{\text{bra}} + L_{\text{ket}} = 5$, beyond which the Rys quadrature again becomes the most efficient technique.

A similar analysis has been conducted for the cyclobutadiene molecule, C_4H_4, at the same geometry as used in Ref. 172 with a uniform magnetic field of strength 1 atomic unit perpendicular to the molecular plane. The results pertaining to the C_4H_4 molecule are presented in Figure 8.3. The upper panels showing the system independent plots

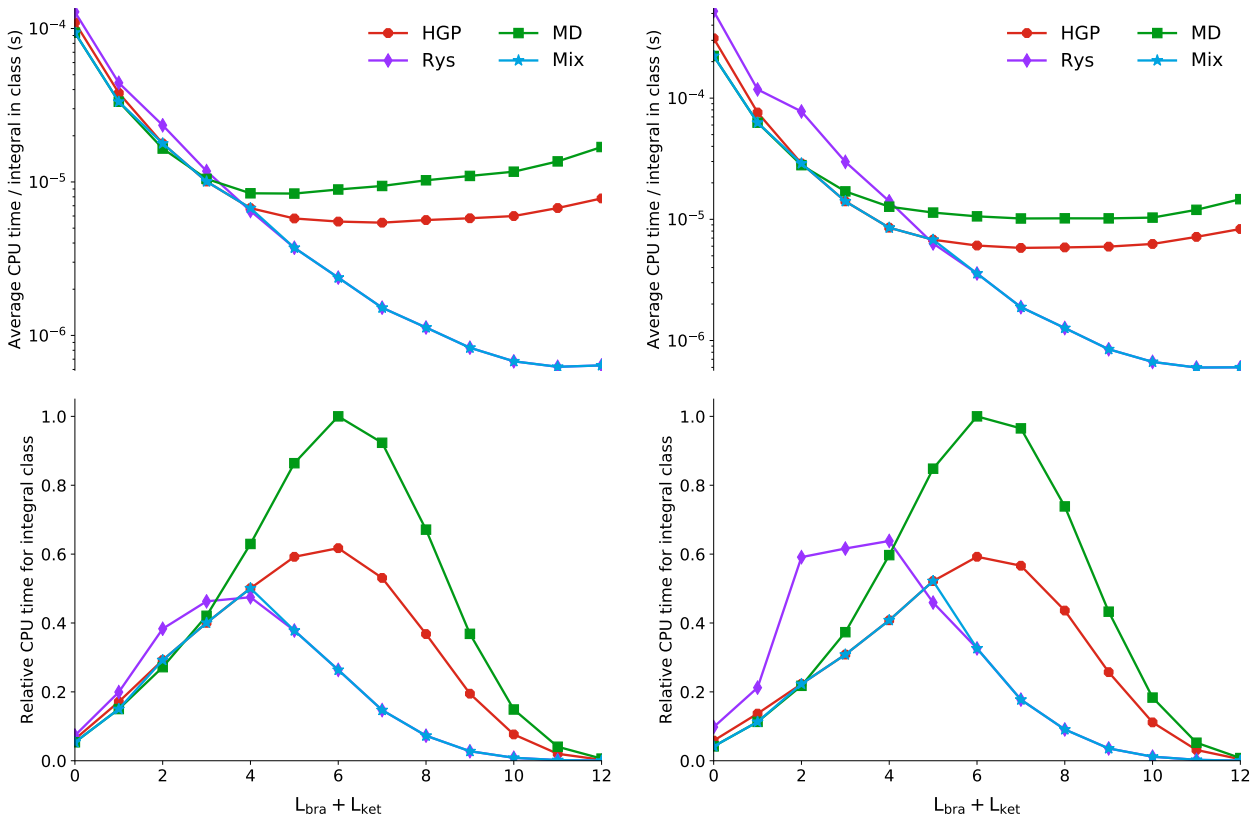

Figure 8.3: Relative timings for LAO-ERIs in a calculation on the C_4H_4 molecule in the aug-cc-pCVTZ basis, classified according to the total angular momenta $L_{\text{bra}} + L_{\text{ket}}$ for each shell quartet. The left hand panels show the results for the primitive basis, the right hand panels for the contracted basis set.

are remarkably similar to those for O_2, confirming the use of this measure for assessing the efficiency of the different algorithms. In particular, the crossover values at which different methods become the most efficient are remarkably transferrable. The lower panels are also very similar, reflecting similar types of functions contributing to the basis set. We have confirmed that increasing the cardinal number of the Dunning-type basis set does not significantly affect conclusions on the relative efficiency of the approaches considered and that the system independent plots (upper panels) remain broadly system independent.

Given that the crossovers in performance between the algorithms are remarkably transferrable, we have considered a simple mixed approach to achieve the optimal efficiency in evaluating the LAO-ERIs. In all constituent plots of Figures 8.2 and 8.3, the results of the mixed approach are plotted with line style $\star$–$\star$. In our integral evaluation code, we have introduced a simple function to select the most appropriate algorithm based on the overall contraction length and the value of $L_{\text{bra}} + L_{\text{ket}}$. After testing over a range of systems we select the MD, HGP and Rys algorithms based on these values according to Table 8.1. This approach can be highly effective in reducing overall computation time,

	MD	HGP	Rys
Primitive	0–1	2–4	5+
Contracted	0–1	2–5	6+

Table 8.1: Ranges of $L_{\text{bra}} + L_{\text{ket}}$ in which each of the considered integral algorithms may be considered optimal. These values are used for the construction of a mixed scheme (see text for details).

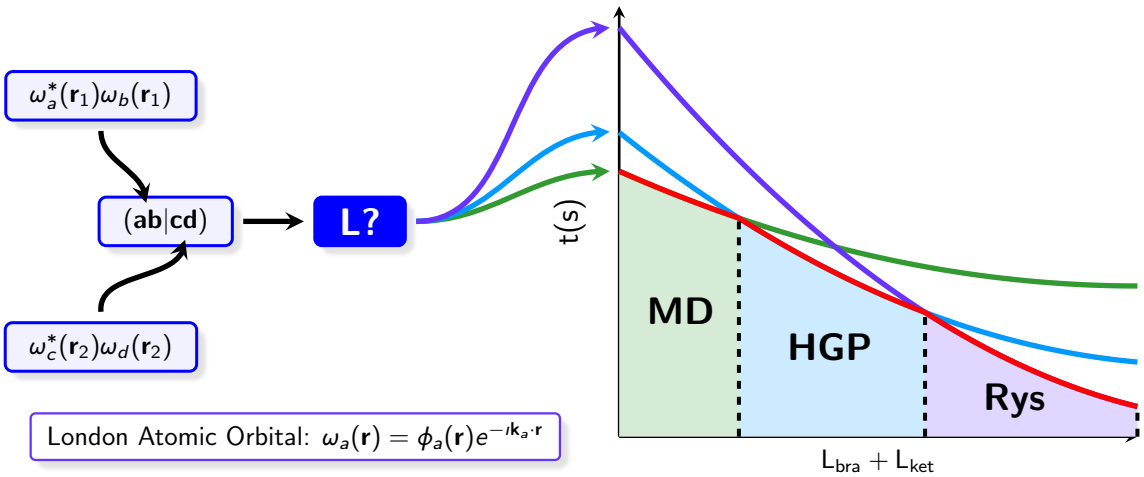

Figure 8.4: A graphical representation of the mixed scheme for the efficient evaluation of LAO integrals, in which $L_{bra} + L_{ket}$ is used to select the optimal integral algorithms for a given shell quartet.

compared with the exclusive use of any one of the algorithms, often with savings between 10% and 30%. The largest gains using the mixed approach are apparent when using contracted functions (see e.g. the lower right panels of Figures 8.2 and 8.3). It is interesting to note that all three approaches considered contribute to the mixed scheme, with MD being most efficient for the lowest angular momenta integrals, Rys quadrature the most efficient for the high angular moment integrals and HGP being optimal in the intermediate region. We expect the simple heuristics used in the mixed approach here to be transferrable to many common basis sets where contraction is most significant in the lower angular momenta core functions. However, some re–calibration may of course be necessary for basis sets containing functions with significantly greater contraction lengths than tested in this work, such as the atomic natural orbital (ANO) series of basis sets by Almlöf & Taylor[311].

8.6 Conclusions

The implementation of integral schemes for non-perturbative electronic structure calculations in the presence of strong magnetic fields using LAOs has been presented. The impact of the use of LAOs on the complexity of three LAO-ERI algorithms was discussed in detail. For the MD method, the introduction of LAOs leads to a significant increase in complexity and has a negative impact on its overall efficiency, particularly for integrals over higher angular momenta basis functions. However, the method still retains efficiency for integrals involving only the lowest angular momenta functions. For the HGP method, the introduction of LAOs does not significantly complicate the underlying algorithm and the efficiency of the HGP scheme for contracted basis functions remains a significant advantage in the generalised form.

For the Rys quadrature method, a careful approach is required to extend the method to the complex case required for LAO integrals. In contrast with previous implementations, we use Flocke's approach[278] rather than an interpolation scheme for the roots and weights, avoiding the need to construct and store large interpolation tables[261]. We also find that the computation of the roots and weights on the fly can be carried out with standard double precision arithmetic. The reduced multiplication scheme of Lindh[306] was adopted in this context to further improve the efficiency of the summation step. One potential complicating factor for the application of Rys quadrature in this context is the possibility of encountering singularities in the complex plane[261,301]. Whilst we have confirmed that these are still present with use of the alternative algorithms presented here, we have not encountered difficulties due to these in any practical calculations, including tests with molecules at a wide range of geometries with fields up to 3 atomic units in strength. In our implementation, if such problems are detected the evaluation of the integral batch can be repeated automatically with a larger number of roots and weights or with one of the alternative LAO-ERI algorithms.

A simple mixed approach was designed to exploit the different advantages of the LAO-ERI algorithms. Specifically, Rys quadrature is used for high angular momenta integrals, the HGP algorithm is used for intermediate angular momenta integrals and particularly where contracted functions are employed and the MD method is employed for the lowest angular momenta integrals. When combined, the mixed approach can achieve substantial efficiencies in overall computation compared with the exclusive use of any one integral algorithm. This provides a robust approach for calculating LAO molecular integrals over a very broad range of fields in an efficient manner, enabling the efficient implementation of non-perturbative treatments of strong magnetic fields in electronic structure calculations. In future

work, we will consider the evaluation of geometrical derivatives of these integrals in a similar fashion to enable the efficient calculation of molecular gradients and optimised geometries in the presence of strong magnetic fields.

Perspective

In this work, a systematic approach to studying the exchange–correlation energy in density functional theory has been outlined, a number of methods for its approximation have been investigated and the requisite apparatus for the generalisation of this approach to systems in magnetic fields has been developed.

In Chapter 1, some of the basic principles underlying all other aspects of this work were set–out; the fundamental problem itself in Section 1.1 and its simplest translation into the language of quantum chemistry in Section 1.2 as Hartree–Fock theory. Chapter 2 described how the Hartree–Fock model can be built–on and improved in a systematic manner by mixing the Hartree–Fock wavefunction with the manifold of its excited states, either as a straightforward linear combination in Section 2.2 or determined using perturbation theory in Section 2.3 or by coupled–cluster theory in Section 2.4. The implementation of perturbation theory and coupled–cluster theory in the quantum chemical rapid development code QUEST was here set–out for the ground–state energy and extended to analytical derivatives in Section 2.5. In Chapter 3, the alternative approach of density functional theory was set–out in detail, tracing its development from the initial theory by Hohenberg & Kohn in Section 3.2 through to the mathematically–rigorous convex–conjugate theory of Lieb in Section 3.4. An overview was given of its Kohn–Sham formulation, the form in which it is applied for practical purposes, in Section 3.5 and the adiabatic connection approach to analysing density functionals was described in Section 3.6. An alternative universal density functional and variation principle was introduced in Chapter 4 and shown to be closely linked to Lieb's theory, yielding the same solutions but by application of different variation principles.

Chapter 5 tied together the theory set out in Chapters 2 & 3, linking in particular the theory in Sections 2.4 & 2.5 with that of Section 3.4. It was described in Section 5.2 how the accuracy and reliability of the coupled–cluster method can be combined with Lieb's mathematically rigorous formulation of density functional theory to compute an energy density with known accuracy, which may be regarded as a target to be modelled in the development of improved approximations for exchange & correlation in Kohn–Sham density functional theory. This approach was applied for a series of atoms and the results presented in Section 5.3, in which the distribution of the energy density and a local form of the adiabatic connection are examined & discussed; an analysis extended to the Hydrogen diatomic molecule in Section 5.4. The distribution of the energy densities in atoms, the Nitrogen diatomic molecule and the Hydrogen Fluoride diatomic molecule is compared to the distribution of the Laplacian of the density in Section 5.5, where the coincidence of features in both is discussed. The energy density and local adiabatic connection are resolved into their spin–parallel and spin–antiparallel contributions in Section 5.6 and the differences in behaviour between the two discussed.

Chapter 6 follows directly from the discussion in Chapter 5 by investigating approaches with which the energy density could be modelled. The first of these, in Section 6.2, is the construction of a model from real–space arguments, parameterised using reference data and tested on a collection of atoms. The second of these, discussed in Section 6.3, sought to construct a model based on interpolating the adiabatic connection at each point in space, for which parameters obtained using both reference data and approximations were considered and which was tested for a range of atomic, ionic & small molecular systems.

This approach to the study of correlation in density functional theory shows promise, since the elucidation of *ab initio*–quality correlation energy densities has led to interesting observations. From the distributions of the correlation energy density, it appears not to be highly localised around atomic nuclei as, for example, the charge & kinetic energy densities are. It instead exhibits a much slower asymptotic decay with distance from the nucleus and does so in a manner that is not necessarily monotonic, instead generally containing features which observations suggest may be delimited by nodes in the Laplacian of the density.

Furthermore, the local adiabatic connection appears to be strongly indicative of how the correlation at each point in space may be characterised; a more linear adiabatic connection indicating the dominance of dynamical correlation or an adiabatic connection with greater curvature indicating a significant presence of nondynamical correlation. In most of the systems examined, the adiabatic connection was near–linear at most points in space and exhibited more considerable curvature in only limited regions of space. Such systems tend to be adequately treated by existing approximations of correlation used in Kohn–Sham calculations.

A well–known example of a system that Kohn–Sham theory is poor at modelling is that of the dissociating Hydrogen molecule. This may be understood in the context of the local adiabatic connection, examined in the Hydrogen molecule as its bond length is extended and the curvature of which becomes increasingly pronounced as the molecule

approaches the limit of dissociation. This behaviour is observed at every point in the system, indicating that the Kohn–Sham model becomes an increasingly poor representation of the entire system.

In addition to this, the spin resolution of the local adiabatic connection suggests that the character of the spin–parallel and spin–antiparallel correlation contributions may not be the same at all points in space. Indeed, limited observations thus far appear to show greater curvature in the spin–parallel local adiabatic connection than its spin–antiparallel equivalent, a pattern that could be rationalised somewhat by considering the spin–parallel contribution to correlation as a higher–order effect of exchange.

The initial results obtained for models of the correlation energy density present an interesting picture. The real–space model yielded approximations to the correlation energy density in atoms that appeared to correctly locate its features, however these approximate correlation energy densities exhibited an excessive rate of asymptotic decay and the functional itself yielded poor approximations to the overall correlation energy of the system. Attempts to improve the functional by varying one of its parameters did improve the estimate for the correlation energy however this was at the cost of completely distorting the correlation energy density.

The other approach considered was that of interpolating the local adiabatic connection. The results of this were encouraging in that the accuracy of local interpolation models was generally superior to that of the equivalent global interpolation models and furthermore that it appears certain approximations to some of the parameters used to construct the local interpolation model could be introduced without a significant deterioration in the accuracy of the resulting correlation energies and energy densities. The dissociation curves of the Hydrogen molecule constructed with this approach indicates the electronic structure of the Hydrogen molecule in the dissociation limit is correctly modelled by some of the local interpolation schemes, but that they perform poorly where the bond length is only moderately extended in the Hydrogen molecule.

A logical starting point for future work would be to construct a correlation functional that is a synthesis of the real–space model and a local interpolation model. They seem somewhat complementary, with the real–space model showing promise in modelling the correlation energy density in higher density regions near the nuclei whilst the interpolation models appear well suited to modelling the low density & asymptotic regions of electronic systems but unsuited to regions of greater charge density around the nuclei. This would suggest a synthesis of the two methods, in which the real–space model was deployed for the region around the nuclei and was gradually replaced by an interpolation model approaching the asymptotic region.

An additional direction for further investigation, signalled in this work, is to consider generalised models for the system that account for the effects an external magnetic field. The treatment of an external magnetic field in the theoretical framework of Section 1.1 is briefly discussed in Chapter 7. It is noted however that, whilst this may be straightforward enough in theory, properly treating magnetic field effects in practical calculations presents new technical challenges that are otherwise absent. An efficient framework to address these matters is presented in detail in Chapter 8 and implemented in QUEST, with the purpose of enabling the study of electronic systems in magnetic fields to be applied to a wider range of molecules and possibly to the calculation and modelling of energy densities in the presence of a magnetic field.

Bibliography

[1] Schrödinger, E. 'Quantisierung als eigenwertproblem.' *Annalen der Physik*, **384**(4), 361–376, 1926.

[2] Born, M. and R. Oppenheimer. 'Zur quantentheorie der molekeln.' *Annalen der Physik*, **389**(20), 457–484, 1927.

[3] Heisenberg, W. 'Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik.' *Zeitschrift für Physik*, **43**(3-4), 172–198, 1927.

[4] Pauli, W. 'The connection between spin and statistics.' *Physical Review*, **58**(8), 716–722, 1940.

[5] Bose, S. 'Plancks gesetz und lichtquantenhypothese.' *Zeitschrift für Physik*, **26**(1), 178–181, 1924.

[6] Einstein, A. 'Quantentheorie des einatomigen idealen gases.' In 'Albert Einstein: Akademie-Vorträge,' pages 237–244. Wiley-VCH Verlag GmbH & Co. KGaA, 2006.

[7] Fermi, E. 'Zur quantelung des idealen einatomigen gases.' *Zeitschrift für Physik*, **36**(11-12), 902–912, 1926.

[8] Dirac, P. A. M. 'Note on exchange phenomena in the Thomas atom.' *Mathematical Proceedings of the Cambridge Philosophical Society*, **26**(03), 376, 1930.

[9] Pauli, W. 'Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren.' *Zeitschrift für Physik*, **31**(1), 765–783, 1925.

[10] Dirac, P. A. M. 'On the theory of quantum mechanics.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **112**(762), 661–677, 1926.

[11] Hartree, D. R. 'The wave mechanics of an atom with a non-Coulomb central field. part I. theory and methods.' *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**(01), 89, 1928.

[12] Slater, J. C. 'The theory of complex spectra.' *Physical Review*, **34**(10), 1293–1322, 1929.

[13] Hartree, D. R. 'The distribution of charge and current in an atom consisting of many electrons obeying Dirac's equations.' *Mathematical Proceedings of the Cambridge Philosophical Society*, **25**(02), 225, 1929.

[14] Fock, V. 'Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems.' *Zeitschrift für Physik*, **61**(1-2), 126–148, 1930.

[15] Kobus, J. 'Diatomic molecules: Exact solutions of HF equations.' In 'Advances in Quantum Chemistry,' pages 1–14. Elsevier BV, 1997.

[16] Roothaan, C. C. J. 'New developments in molecular orbital theory.' *Reviews of Modern Physics*, **23**(2), 69–89, 1951.

[17] Hall, G. G. 'The molecular orbital theory of chemical valency. VIII. a method of calculating ionization potentials.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **205**(1083), 541–552, 1951.

[18] Pople, J. A. and R. K. Nesbet. 'Self-consistent orbitals for radicals.' *The Journal of Chemical Physics*, **22**(3), 571, 1954.

[19] Roothaan, C. C. J. 'Self-consistent field theory for open shells of electronic systems.' *Reviews of Modern Physics*, **32**(2), 179–185, 1960.

[20] Löwdin, P.-O. 'Exchange, correlation, and spin effects in molecular and solid-state theory.' *Reviews of Modern Physics*, **34**(1), 80–87, 1962.

[21] Brillouin, L. 'Les problèmes de perturbations et les champs self-consistents.' *Journal de Physique et le Radium*, **3**(9), 373–389, 1932.

[22] Nesbet, R. K. 'Configuration interaction in orbital theories.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **230**(1182), 312–321, 1955.

[23] Handy, N. C. 'Multi-root configuration interaction calculations.' *Chemical Physics Letters*, **74**(2), 280–283, 1980.

[24] Knowles, P. and N. Handy. 'A new determinant-based full configuration interaction method.' *Chemical Physics Letters*, **111**(4-5), 315–321, 1984.

[25] Olsen, J., B. O. Roos, P. Jørgensen, and H. J. A. Jensen. 'Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces.' *The Journal of Chemical Physics*, **89**(4), 2185, 1988.

[26] Duch, W. and G. H. F. Diercksen. 'Size-extensivity corrections in configuration interaction methods.' *The Journal of Chemical Physics*, **101**(4), 3018, 1994.

[27] Langhoff, S. R. and E. R. Davidson. 'Configuration interaction calculations on the nitrogen molecule.' *International Journal of Quantum Chemistry*, **8**(1), 61–72, 1974.

[28] Rayleigh, J. W. S. *Theory Of Sound*, volume 1 of *Theoretical Chemistry*. MacMillan London, 2 edition, 1894.

[29] Møller, C. and M. S. Plesset. 'Note on an approximation treatment for many-electron systems.' *Physical Review*, **46**(7), 618–622, 1934.

[30] Leininger, M. L., W. D. Allen, H. F. Schaefer, and C. D. Sherrill. 'Is Møller–Plesset perturbation theory a convergent *ab initio* method?' *The Journal of Chemical Physics*, **112**(21), 9213, 2000.

[31] 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*, **45**(11), 4256, 1966.

[32] Čížek, J. and J. Paldus. 'Correlation problems in atomic and molecular systems III. rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methods.' *International Journal of Quantum Chemistry*, **5**(4), 359–379, 1971.

[33] Paldus, J. 'Correlation problems in atomic and molecular systems. V. spin-adapted coupled cluster many-electron theory.' *The Journal of Chemical Physics*, **67**(1), 303–318, 1977.

[34] Paldus, J., J. Čížek, M. Saute, and A. Laforgue. 'Correlation problems in atomic and molecular systems. VI. coupled-cluster approach to open-shell systems.' *Physical Review A*, **17**(3), 805–815, 1978.

[35] Bartlett, R. J. 'Many-body perturbation theory and coupled cluster theory for electron correlation in molecules.' *Annual Review of Physical Chemistry*, **32**(1), 359–401, 1981.

[36] Adamowicz, L., W. D. Laidig, and R. J. Bartlett. 'Analytical gradients for the coupled-cluster method.' *International Journal of Quantum Chemistry*, **26**(S18), 245–254, 1984.

[37] Stanton, J. F., J. Gauss, J. D. Watts, and R. J. Bartlett. 'A direct product decomposition approach for symmetry exploitation in many-body methods. I. energy calculations.' *The Journal of Chemical Physics*, **94**(6), 4334–4345, 1991.

[38] Koch, H., H. J. A. Jensen, P. Jørgensen, T. Helgaker, G. E. Scuseria, and H. F. Schaefer. 'Coupled cluster energy derivatives. analytic hessian for the closed-shell coupled cluster singles and doubles wave function: Theory and applications.' *The Journal of Chemical Physics*, **92**(8), 4924–4940, 1990.

[39] Paldus, J., J. Čížek, and I. Shavitt. 'Correlation problems in atomic and molecular systems. IV. extended coupled-pair many-electron theory and its application to the BH3 molecule.' *Physical Review A*, **5**(1), 50–67, 1972.

[40] Čížek, J. 'On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules.' In 'Advances in Chemical Physics,' volume 14, pages 35–89. John Wiley & Sons Inc., 1969.

[41] Jørgensen, P. and J. Simons. *Second Quantization-Based Methods in Quantum Chemistry*. Elsevier BV, 1981.

[42] Surján, P. R. *Second Quantized Approach to Quantum Chemistry*. Springer Berlin Heidelberg, 1989.

[43] Shavitt, I. and R. J. Bartlett. *Many–Body Methods in Chemistry and Physics*. Cambridge University Press, 2009.

[44] Purvis, G. D. and R. J. Bartlett. 'A full coupled-cluster singles and doubles model: The inclusion of disconnected triples.' *The Journal of Chemical Physics*, **76**(4), 1910–1918, 1982.

[45] Raghavachari, K., G. W. Trucks, J. A. Pople, and M. Head-Gordon. 'A fifth-order perturbation comparison of electron correlation theories.' *Chemical Physics Letters*, **157**(6), 479–483, 1989.

[46] Bartlett, R. J., J. Watts, S. Kucharski, and J. Noga. 'Non-iterative fifth-order triple and quadruple excitation energy corrections in correlated methods.' *Chemical Physics Letters*, **165**(6), 513–522, 1990.

[47] Pople, J. A., R. Krishnan, H. B. Schlegel, and J. S. Binkley. 'Derivative studies in Hartree–Fock and Møller–Plesset theories.' *International Journal of Quantum Chemistry*, **16**(S13), 225–241, 1978.

[48] Jørgensen, P. and T. Helgaker. 'Møller–Plesset energy derivatives.' *The Journal of Chemical Physics*, **89**(3), 1560–1570, 1988.

[49] Arponen, J. 'Variational principles and linked-cluster $\exp(S)$ expansions for static and dynamic many-body problems.' *Annals of Physics*, **151**(2), 311–382, 1983.

[50] Gauss, J., J. F. Stanton, and R. J. Bartlett. 'Coupled-cluster open-shell analytic gradients: Implementation of the direct product decomposition approach in energy gradient calculations.' *The Journal of Chemical Physics*, **95**(4), 2623–2638, 1991.

[51] Gauss, J., W. J. Lauderdale, J. F. Stanton, J. D. Watts, and R. J. Bartlett. 'Analytic energy gradients for open-shell coupled-cluster singles and doubles (CCSD) calculations using restricted open-shell Hartree–Fock (ROHF) reference functions.' *Chemical Physics Letters*, **182**(3-4), 207–215, 1991.

[52] Monkhorst, H. J. 'Calculation of properties with the coupled-cluster method.' *International Journal of Quantum Chemistry*, **12**(S11), 421–432, 1977.

[53] Kümmel, H. G. 'Expectation values and density matrices in the Coupled-Cluster theory.' *International Journal of Quantum Chemistry*, **24**(1), 79–84, 1983.

[54] Bartlett, R. J. 'Analytical evaluation of gradients in coupled-cluster and many-body perturbation theory.' In 'Geometrical Derivatives of Energy Surfaces and Molecular Properties,' pages 35–61. Springer Netherlands, 1986.

[55] Trucks, G. W., E. Salter, C. Sosa, and R. J. Bartlett. 'Theory and implementation of the MBPT density matrix. an application to one-electron properties.' *Chemical Physics Letters*, **147**(4), 359–366, 1988.

[56] Salter, E. A., G. W. Trucks, and R. J. Bartlett. 'Analytic energy derivatives in many-body methods. I. first derivatives.' *The Journal of Chemical Physics*, **90**(3), 1752–1766, 1989.

[57] Gauss, J. and J. F. Stanton. 'Coupled-cluster calculations of nuclear magnetic resonance chemical shifts.' *The Journal of Chemical Physics*, **103**(9), 3561–3577, 1995.

[58] Handy, N. C. and H. F. Schaefer. 'On the evaluation of analytic energy derivatives for correlated wave functions.' *The Journal of Chemical Physics*, **81**(11), 5031–5033, 1984.

[59] Dalgarno, A. and A. L. Stewart. 'On the perturbation theory of small disturbances.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **238**(1213), 269–275, 1956.

[60] Dalgarno, A. and A. L. Stewart. 'A perturbation calculation of properties of the 2pπ state of HeH^{2+}.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **240**(1221), 274–283, 1957.

[61] Dalgarno, A. and A. L. Stewart. 'A perturbation calculation of properties of the helium iso-electronic sequence.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **247**(1249), 245–259, 1958.

[62] Watts, J. D., J. Gauss, and R. J. Bartlett. 'Open-shell analytical energy gradients for triple excitation many-body, coupled-cluster methods: MBPT(4), CCSD+T(CCSD), CCSD(T) and QCISD(T).' *Chemical Physics Letters*, **200**(1-2), 1–7, 1992.

[63] QUEST. 'A rapid development platform for quantum electronic structure techniques.' `quest.codes`, 2017.

[64] Hohenberg, P. and W. Kohn. 'Inhomogeneous electron gas.' *Physical Review*, **136**(3B), B864–B871, 1964.

[65] Kohn, W. and L. J. Sham. 'Self-consistent equations including exchange and correlation effects.' *Physical Review*, **140**(4A), A1133–A1138, 1965.

[66] Levy, M. and J. P. Perdew. 'In defense of the Hohenberg-Kohn theorem and density functional theory.' *International Journal of Quantum Chemistry*, **21**(2), 511–513, 1982.

[67] Lieb, E. H. 'Density functionals for Coulomb systems.' *International Journal of Quantum Chemistry*, **24**(3), 243–277, 1983.

[68] Thomas, L. H. 'The calculation of atomic fields.' *Mathematical Proceedings of the Cambridge Philosophical Society*, **23**(05), 542, 1927.

[69] Fermi, E. 'Un metodo statistico per la determinazione di alcune priorietà dell'atome.' *Rendiconti dell'Accademia dei Lincei*, **6**(32), 602–607, 1927.

[70] Levy, M. 'Electron densities in search of Hamiltonians.' *Physical Review A*, **26**(3), 1200–1208, 1982.

[71] Eschrig, H. *The Fundamentals of Density Functional Theory*, volume 32 of *Teubner Texte zur Physik*. Vieweg+Teubner Verlag, 1996.

[72] Helgaker, T., P. Jørgensen, and J. Olsen. *Density Functional Theory: A Convex Treatment*. John Wiley & Sons Inc., New York, United States, In Preparation.

[73] Lieb, E. and M. Loss. *Analysis*, volume 14 of *Graduate Studies In Mathematics*. American Mathematical Society, Providence, United States, 2001.

[74] Weisstein, E. 'Field axioms.' *MathWorld – A Wolfram Web Resource*, 2015.

[75] v. Weizsäcker, C. F. 'Zur theorie der kernmassen.' *Zeitschrift für Physik*, **96**(7-8), 431–458, 1935.

[76] Rockafellar, R. T. *Convex Analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, United States, 1970.

[77] Kvaal, S. and T. Helgaker. 'Ground-state densities from the Rayleigh–Ritz variation principle and from density-functional theory.' *The Journal of Chemical Physics*, **143**, 1184106, 2015.

[78] Lammert, P. E. 'Differentiability of Lieb functional in electronic density functional theory.' *International Journal of Quantum Chemistry*, **107**(10), 1943–1953, 2007.

[79] Kvaal, S., U. Ekström, A. M. Teale, and T. Helgaker. 'Differentiable but exact formulation of density-functional theory.' *The Journal of Chemical Physics*, **140**(18), 18A518, 2014.

[80] Savin, A., C. Umrigar, and X. Gonze. 'Relationship of Kohn–Sham eigenvalues to excitation energies.' *Chemical Physics Letters*, **288**(2-4), 391–395, 1998.

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

[82] Buijse, M. A. and E. J. Baerends. 'Fermi holes and Coulomb holes.' In 'Density Functional Theory of Molecules, Clusters, and Solids,' pages 1–46. Springer Science + Business Media, 1996.

[83] Buijse, M. A. and E. J. Baerends. 'An approximate exchange-correlation hole density as a functional of the natural orbitals.' *Molecular Physics*, **100**(4), 401–421, 2002.

[84] Lieb, E. H. and S. Oxford. 'Improved lower bound on the indirect Coulomb energy.' *International Journal of Quantum Chemistry*, **19**(3), 427–439, 1981.

[85] Levy, M. 'Coordinate scaling requirements for approximating exchange and correlation.' In 'NATO ASI Series,' pages 11–31. Springer Science + Business Media, 1995.

[86] Slater, J. C. 'A simplification of the Hartree–Fock method.' *Physical Review*, **81**(3), 385–390, 1951.

[87] Sawada, K. 'Correlation energy of an electron gas at high density.' *Physical Review*, **106**(2), 372–383, 1957.

[88] Levy, M. and J. P. Perdew. 'Tight bound and convexity constraint on the exchange-correlation-energy functional in the low-density limit, and other formal tests of generalized-gradient approximations.' *Physical Review B*, **48**(16), 11638–11645, 1993.

[89] Ceperley, D. M. and B. J. Alder. 'Ground state of the electron gas by a stochastic method.' *Physical Review Letters*, **45**(7), 566–569, 1980.

[90] Perdew, J. P. and Y. Wang. 'Accurate and simple analytic representation of the electron-gas correlation energy.' *Physical Review B*, **45**(23), 13244–13249, 1992.

[91] Becke, A. D. 'Density-functional exchange-energy approximation with correct asymptotic behavior.' *Physical Review A*, **38**(6), 3098–3100, 1988.

[92] Perdew, J. P., K. Burke, and M. Ernzerhof. 'Generalized gradient approximation made simple.' *Physical Review Letters*, **77**(18), 3865–3868, 1996.

[93] Becke, A. D. and M. R. Roussel. 'Exchange holes in inhomogeneous systems: A coordinate-space model.' *Physical Review A*, **39**(8), 3761–3767, 1989.

[94] Lee, C., W. Yang, and R. G. Parr. 'Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.' *Physical Review B*, **37**(2), 785–789, 1988.

[95] Becke, A. D. 'Density-functional thermochemistry. III. the role of exact exchange.' *The Journal of Chemical Physics*, **98**(7), 5648, 1993.

[96] Vosko, S. H., L. Wilk, and M. Nusair. 'Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis.' *Canadian Journal of Physics*, **58**(8), 1200–1211, 1980.

[97] Langreth, D. and J. Perdew. 'The exchange-correlation energy of a metallic surface.' *Solid State Communications*, **17**(11), 1425–1429, 1975.

[98] Gunnarsson, O. and B. I. Lundqvist. 'Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism.' *Physical Review B*, **13**(10), 4274–4298, 1976.

[99] Gunnarsson, O. and B. I. Lundqvist. 'Erratum: Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism.' *Physical Review B*, **15**(12), 6006–6006, 1977.

[100] Langreth, D. C. and J. P. Perdew. 'Exchange-correlation energy of a metallic surface: Wave-vector analysis.' *Physical Review B*, **15**(6), 2884–2901, 1977.

[101] Yang, W. 'Generalized adiabatic connection in density functional theory.' *The Journal of Chemical Physics*, **109**(23), 10107, 1998.

[102] Toulouse, J., F. Colonna, and A. Savin. 'Exchange–correlation potentials and local energies per particle along nonlinear adiabatic connections.' *Molecular Physics*, **103**(20), 2725–2734, 2005.

[103] Teale, A. M., S. Coriani, and T. Helgaker. 'Range-dependent adiabatic connections.' *The Journal of Chemical Physics*, **133**(16), 164112, 2010.

[104] Hellmann, H. *Einführung in die Quantenchemie*. Franz Deuticke, Leipzig & Wien, 1937.

[105] Feynman, R. P. 'Forces in molecules.' *Physical Review*, **56**(4), 340–343, 1939.

[106] Savin, A., F. Colonna, and R. Pollet. 'Adiabatic connection approach to density functional theory of electronic systems.' *International Journal of Quantum Chemistry*, **93**(3), 166–190, 2003.

[107] Levy, M. 'Universal variational functionals of electron densities, first–order density matrices, and natural spin–orbitals and solution of the v–representability problem.' *Proceedings of the National Academy of Sciences*, **76**(12), 6062–6065, 1979.

[108] Gidopoulos, N. I. 'Progress at the interface of wave-function and density-functional theories.' *Physical Review A*, **83**(4), 040502, 2011.

[109] Davidson, E. R. *Reduced Density Matrices in Quantum Chemistry*, volume 6 of *Theoretical Chemistry*. Elsevier BV, 1976.

[110] Colonna, F. and A. Savin. 'Correlation energies for some two- and four-electron systems along the adiabatic connection in density functional theory.' *The Journal of Chemical Physics*, **110**(6), 2828, 1999.

[111] Teale, A. M., S. Coriani, and T. Helgaker. 'The calculation of adiabatic-connection curves from full configuration-interaction densities: Two-electron systems.' *The Journal of Chemical Physics*, **130**(10), 104111, 2009.

[112] Teale, A. M., S. Coriani, and T. Helgaker. 'Accurate calculation and modeling of the adiabatic connection in density functional theory.' *The Journal of Chemical Physics*, **132**(16), 164115, 2010.

[113] Yang, W. and Q. Wu. 'Direct method for optimized effective potentials in density-functional theory.' *Physical Review Letters*, **89**(14), 143002, 2002.

[114] Wu, Q. and W. Yang. 'A direct optimization method for calculating density functionals and exchange–correlation potentials from electron densities.' *The Journal of Chemical Physics*, **118**(6), 2498, 2003.

[115] Fermi, E. and E. Amaldi. 'Le orbite ∞s degli elementi.' *Memorie dell'Accademia d'Italia*, **6**(1), 119–149, 1934.

[116] Dunning, T. H. 'Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen.' *The Journal of Chemical Physics*, **90**(2), 1007, 1989.

[117] Kendall, R. A., T. H. Dunning, and R. J. Harrison. 'Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions.' *The Journal of Chemical Physics*, **96**(9), 6796–6806, 1992.

[118] Wu, Q. and W. Yang. 'Algebraic equation and iterative optimization for the optimized effective potential in density functional theory.' *Journal of Theoretical and Computational Chemistry*, **2**(04), 627–638, 2003.

[119] 'Numba.' `numba.pydata.org`, 2017. Version 0.32.

[120] Lam, S. K., A. Pitrou, and S. Seibert. 'Numba.' In 'Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC – LLVM'15,' ACM Press, 2015.

[121] Ernzerhof, M., J. P. Perdew, and K. Burke. 'Coupling-constant dependence of atomization energies.' *International Journal of Quantum Chemistry*, **64**(3), 285–295, 1997.

[122] Cohen, A. J., P. Mori-Sánchez, and W. Yang. 'Challenges for density functional theory.' *Chemical Reviews*, **112**(1), 289–320, 2012.

[123] Burke, K. 'Perspective on density functional theory.' *The Journal of Chemical Physics*, **136**(15), 150901, 2012.

[124] Kryachko, E. S. and E. V. Ludeña. 'Density functional theory: Foundations reviewed.' *Physical Reports*, **544**(2), 123–239, 2014.

[125] Tao, J., J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. 'Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids.' *Physical Review Letters*, **91**(14), 2003.

[126] Perdew, J. P., A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. 'Workhorse semilocal density functional for condensed matter physics and quantum chemistry.' *Physical Review Letters*, **103**(2), 2009.

[127] Zhao, Y. and D. G. Truhlar. 'Density functionals with broad applicability in chemistry.' *Accounts of Chemical Research*, **41**(2), 157–167, 2008.

[128] Bartlett, R. J. '*Ab initio* DFT and its role in electronic structure theory.' *Molecular Physics*, **108**(21-23), 3299–3311, 2010.

[129] Sharp, R. T. and G. K. Horton. 'A variational approach to the unipotential many-electron problem.' *Physical Review*, **90**(2), 317–317, 1953.

[130] Talman, J. D. and W. F. Shadwick. 'Optimized effective atomic central potential.' *Physical Review A*, **14**(1), 36–40, 1976.

[131] Kadantsev, E. S. and M. J. Stott. 'Variational method for inverting the Kohn–Sham procedure.' *Physical Review A*, **69**(1), 2004.

[132] van Leeuwen, R. and E. J. Baerends. 'Exchange-correlation potential with correct asymptotic behavior.' *Physical Review A*, **49**(4), 2421–2431, 1994.

[133] Burke, K., F. G. Cruz, and K.-C. Lam. 'Unambiguous exchange-correlation energy density.' *The Journal of Chemical Physics*, **109**(19), 8161, 1998.

[134] Mirtschink, A., M. Seidl, and P. Gori-Giorgi. 'Energy densities in the strong-interaction limit of density functional theory.' *Journal of Chemical Theory and Computation*, **8**(9), 3097–3107, 2012.

[135] Gori-Giorgi, P., G. Vignale, and M. Seidl. 'Electronic zero-point oscillations in the strong-interaction limit of density functional theory.' *Journal of Chemical Theory and Computation*, **5**(4), 743–753, 2009.

[136] Becke, A. D. 'Real-space post-Hartree–Fock correlation models.' *The Journal of Chemical Physics*, **122**(6), 064101, 2005.

[137] Becke, A. D. and E. R. Johnson. 'A unified density-functional treatment of dynamical, nondynamical, and dispersion correlations.' *The Journal of Chemical Physics*, **127**(12), 124108, 2007.

[138] Perdew, J. P., V. N. Staroverov, J. Tao, and G. E. Scuseria. 'Density functional with full exact exchange, balanced nonlocality of correlation, and constraint satisfaction.' *Physical Review A*, **78**(5), 2008.

[139] Pistol, M. E. and C. O. Almbladh. 'Adiabatic connections and properties of coupling-integrated exchange–correlation holes and pair densities in density functional theory.' *Chemical Physics Letters*, **480**(1-3), 136–139, 2009.

[140] Buijse, M. A., E. J. Baerends, and J. G. Snijders. 'Analysis of correlation in terms of exact local potentials: Applications to two-electron systems.' *Physical Review A*, **40**(8), 4190–4202, 1989.

[141] Slater, J. C. *The Self Consistent Field for Molecules and Solids*, volume 4 of *Quantum Theory of Molecules and Solids*. McGraw-Hill, New York, 1974.

[142] Becke, A. D. 'Hartree-Fock exchange energy of an inhomogeneous electron gas.' *International Journal of Quantum Chemistry*, **23**(6), 1915–1922, 1983.

[143] Gori-Giorgi, P. and A. Savin. 'System-adapted correlation energy density functionals from effective pair interactions.' *Philosophical Magazine*, **86**(17-18), 2643–2659, 2006.

[144] Seidl, M., J. P. Perdew, and S. Kurth. 'Density functionals for the strong-interaction limit.' *Physical Review A*, **62**(1), 2000.

[145] Cohen, A. J., P. Mori-Sánchez, and W. Yang. 'Assessment and formal properties of exchange-correlation functionals constructed from the adiabatic connection.' *The Journal of Chemical Physics*, **127**(3), 034101, 2007.

[146] Gori-Giorgi, P., J. G. Ángyán, and A. Savin. 'Charge density reconstitution from approximate exchange-correlation holes.' *Canadian Journal of Chemistry*, **87**(10), 1444–1450, 2009.

[147] Olsen, J., P. Jørgensen, and J. Simons. 'Passing the one-billion limit in full configuration-interaction (FCI) calculations.' *Chemical Physics Letters*, **169**(6), 463–472, 1990.

[148] Helgaker, T. and P. Jørgensen. 'Configuration-interaction energy derivatives in a fully variational formulation.' *Theoretica Chimica Acta*, **75**(2), 111–127, 1989.

[149] Dalton. 'A molecular electronic structure program.' `daltonprogram.org`, 2015.

[150] Aidas, K., C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenaes, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjaergaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnaes, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren. 'The Dalton quantum chemistry program system.' *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **4**(3), 269–284, 2014.

[151] Lindh, R., P.-Å. Malmqvist, and L. Gagliardi. 'Molecular integrals by numerical quadrature. I. radial integration.' *Theoretical Chemistry Accounts*, **106**(3), 178–187, 2001.

[152] Filippi, C., C. J. Umrigar, and M. Taut. 'Comparison of exact and approximate density functionals for an exactly soluble model.' *The Journal of Chemical Physics*, **100**(2), 1290, 1994.

[153] Umrigar, C. J. and X. Gonze. 'Accurate exchange-correlation potentials and total-energy components for the helium isoelectronic series.' *Physical Review A*, **50**(5), 3827–3837, 1994.

[154] Peach, M. J. G., A. M. Teale, and D. J. Tozer. 'Modeling the adiabatic connection in H_2.' *The Journal of Chemical Physics*, **126**(24), 244104, 2007.

[155] Seidl, M., J. P. Perdew, and M. Levy. 'Strictly correlated electrons in density-functional theory.' *Physical Review A*, **59**(1), 51–54, 1999.

[156] Seidl, M., P. Gori-Giorgi, and A. Savin. 'Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities.' *Physical Review A*, **75**(4), 2007.

[157] Perdew, J., S. Kurth, and M. Seidl. 'Exploring the adiabatic connection between weak and strong–interaction limits in density functional theory.' *International Journal of Modern Physics B*, **15**(10n11), 1672–1683, 2001.

[158] Liu, Z.-F. and K. Burke. 'Adiabatic connection for strictly correlated electrons.' *The Journal of Chemical Physics*, **131**(12), 124124, 2009.

[159] Vuckovic, S., L. O. Wagner, A. Mirtschink, and P. Gori-Giorgi. 'Hydrogen molecule dissociation curve with functionals based on the strictly correlated regime.' *Journal of Chemical Theory and Computation*, **11**(7), 3153–3162, 2015.

[160] Görling, A. and M. Levy. 'Exact Kohn-Sham scheme based on perturbation theory.' *Physical Review A*, **50**(1), 196–204, 1994.

[161] Cancio, A. C., C. E. Wagner, and S. A. Wood. 'Laplacian-based models for the exchange energy.' *International Journal of Quantum Chemistry*, **112**(24), 3796–3806, 2012.

[162] Cancio, A., C. Fong, and J. Nelson. 'Exchange-correlation hole of the Si atom: A quantum Monte Carlo study.' *Physical Review A*, **62**(6), 2000.

[163] Gori-Giorgi, P. and A. Savin. 'Degeneracy and size consistency in electronic density functional theory.' *Journal of Physics: Conference Series*, **117**, 012017, 2008.

[164] Savin, A. 'Is size-consistency possible with density functional approximations?' *Chemical Physics*, **356**(1-3), 91–97, 2009.

[165] Savin, A. 'On degeneracy, near-degeneracy and density functional theory.' In 'Recent Developments and Applications of Modern Density Functional Theory,' volume 4 of *Theoretical and Computational Chemistry*, pages 327–357. Elsevier BV, Amsterdam, 1996.

[166] Modrzejewski, M., M. Lesiuk, Ł. Rajchel, M. Szczęśniak, and G. Chałasiński. 'A first-principles-based correlation functional for harmonious connection of short-range correlation and long-range dispersion.' *The Journal of Chemical Physics*, **137**(20), 204121, 2012.

[167] Grimme, S., J. Antony, S. Ehrlich, and H. Krieg. 'A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu.' *The Journal of Chemical Physics*, **132**(15), 154104, 2010.

[168] Gori-Giorgi, P. and J. P. Perdew. 'Short-range correlation in the uniform electron gas: Extended Overhauser model.' *Physical Review B*, **64**(15), 2001.

[169] Becke, A. D. 'Correlation energy of an inhomogeneous electron gas: A coordinate-space model.' *The Journal of Chemical Physics*, **88**(2), 1053, 1988.

[170] Rajagopal, A. K., J. C. Kimball, and M. Banerjee. 'Short-ranged correlations and the ferromagnetic electron gas.' *Physical Review B*, **18**(5), 2339–2345, 1978.

[171] Ekström, U., L. Visscher, R. Bast, A. J. Thorvaldsen, and K. Ruud. 'Arbitrary-order density functional response theory from automatic differentiation.' *Journal of Chemical Theory and Computation*, **6**(7), 1971–1980, 2010.

[172] Tellgren, E. I., A. Soncini, and T. Helgaker. 'Nonperturbative *ab initio* calculations in strong magnetic fields using London orbitals.' *The Journal of Chemical Physics*, **129**(15), 154114, 2008.

[173] Perdew, J. P., J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. 'Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation.' *Physical Review B*, **46**(11), 6671–6687, 1992.

[174] Perdew, J. P., M. Ernzerhof, and K. Burke. 'Rationale for mixing exact exchange with density functional approximations.' *The Journal of Chemical Physics*, **105**(22), 9982, 1996.

[175] Goerigk, L. and S. Grimme. '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*, **7**(2), 291–309, 2011.

[176] Sharkas, K., J. Toulouse, and A. Savin. 'Double-hybrid density-functional theory made rigorous.' *The Journal of Chemical Physics*, **134**(6), 064113, 2011.

[177] Furche, F. 'Molecular tests of the random phase approximation to the exchange-correlation energy functional.' *Physical Review B*, **64**(19), 2001.

[178] Heßelmann, A. and A. Görling. 'Random phase approximation correlation energies with exact Kohn–Sham exchange.' *Molecular Physics*, **108**(3-4), 359–372, 2010.

[179] Ángyán, J. G., R.-F. Liu, J. Toulouse, and G. Jansen. 'Correlation energy expressions from the adiabatic-connection fluctuation–dissipation theorem approach.' *Journal of Chemical Theory and Computation*, **7**(10), 3116–3130, 2011.

[180] Zhou, Y., H. Bahmann, and M. Ernzerhof. 'Construction of exchange-correlation functionals through interpolation between the non-interacting and the strong-correlation limit.' *The Journal of Chemical Physics*, **143**(12), 124103, 2015.

[181] Ernzerhof, M. 'Construction of the adiabatic connection.' *Chemical Physics Letters*, **263**(3-4), 499–506, 1996.

[182] Burke, K., M. Ernzerhof, and J. P. Perdew. 'The adiabatic connection method: a non-empirical hybrid.' *Chemical Physics Letters*, **265**(1-2), 115–120, 1997.

[183] Mori-Sánchez, P., A. J. Cohen, and W. Yang. 'Self-interaction-free exchange-correlation functional for thermochemistry and kinetics.' *The Journal of Chemical Physics*, **124**(9), 091102, 2006.

[184] Irons, T. J. P. and A. M. Teale. 'The coupling constant averaged exchange–correlation energy density.' *Molecular Physics*, **114**(3-4), 484–497, 2016.

[185] Malet, F. and P. Gori-Giorgi. 'Strong correlation in Kohn–Sham density functional theory.' *Physical Review Letters*, **109**(24), 2012.

[186] Mirtschink, A., M. Seidl, and P. Gori-Giorgi. 'Derivative discontinuity in the strong-interaction limit of density-functional theory.' *Physical Review Letters*, **111**(12), 2013.

[187] Malet, F., A. Mirtschink, J. C. Cremon, S. M. Reimann, and P. Gori-Giorgi. 'Kohn–Sham density functional theory for quantum wires in arbitrary correlation regimes.' *Physical Review B*, **87**(11), 2013.

[188] Mendl, C. B., F. Malet, and P. Gori-Giorgi. 'Wigner localization in quantum dots from Kohn–Sham density functional theory without symmetry breaking.' *Physical Review B*, **89**(12), 2014.

[189] Malet, F., A. Mirtschink, K. J. H. Giesbertz, L. O. Wagner, and P. Gori-Giorgi. 'Exchange–correlation functionals from the strong interaction limit of DFT: applications to model chemical systems.' *Physical Chemistry Chemical Physics*, **16**(28), 14551, 2014.

[190] Chen, H., G. Friesecke, and C. B. Mendl. 'Numerical methods for a Kohn–Sham density functional model based on optimal transport.' *Journal of Chemical Theory and Computation*, **10**(10), 4360–4368, 2014.

[191] Malet, F., A. Mirtschink, C. B. Mendl, J. Bjerlin, E. O. Karabulut, S. M. Reimann, and P. Gori-Giorgi. 'Density-functional theory for strongly correlated bosonic and fermionic ultracold dipolar and ionic gases.' *Physical Review Letters*, **115**(3), 2015.

[192] Buttazzo, G., L. D. Pascale, and P. Gori-Giorgi. 'Optimal-transport formulation of electronic density-functional theory.' *Physical Review A*, **85**(6), 2012.

[193] Cotar, C., G. Friesecke, and C. Klüppelberg. 'Density functional theory and optimal transportation with Coulomb cost.' *Communications on Pure and Applied Mathematics*, **66**(4), 548–599, 2012.

[194] Mendl, C. B. and L. Lin. 'Kantorovich dual solution for strictly correlated electrons in atoms and molecules.' *Physical Review B*, **87**(12), 2013.

[195] Benamou, J.-D., G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. 'Iterative Bregman projections for regularized transportation problems.' *SIAM Journal on Scientific Computing*, **37**(2), A1111–A1138, 2015.

[196] Colombo, M., L. D. Pascale, and S. D. Marino. 'Multimarginal optimal transport maps for one-dimensional repulsive costs.' *Journal Canadien de Mathématiques*, **67**(2), 350–368, 2014.

[197] Colombo, M. and F. Stra. 'Counterexamples in multimarginal optimal transport with Coulomb cost and spherically symmetric data.' *Mathematical Models and Methods in Applied Sciences*, **26**(06), 1025–1049, 2016.

[198] Görling, A. and M. Levy. 'Correlation-energy functional and its high-density limit obtained from a coupling-constant perturbation expansion.' *Physical Review B*, **47**(20), 13105–13113, 1993.

[199] Levy, M. 'Density-functional exchange correlation through coordinate scaling in adiabatic connection and correlation hole.' *Physical Review A*, **43**(9), 4637–4646, 1991.

[200] Seidl, M., J. P. Perdew, and S. Kurth. 'Density functionals for the strong-interaction limit.' *Physical Review A*, **62**(1), 2000.

[201] Peach, M. J. G., A. M. Miller, A. M. Teale, and D. J. Tozer. 'Adiabatic connection forms in density functional theory: H_2 and the He isoelectronic series.' *The Journal of Chemical Physics*, **129**(6), 064105, 2008.

[202] Bahmann, H. and M. Kaupp. 'Efficient self-consistent implementation of local hybrid functionals.' *Journal of Chemical Theory and Computation*, **11**(4), 1540–1548, 2015.

[203] Janesko, B. G., A. V. Krukau, and G. E. Scuseria. 'Self-consistent generalized Kohn-Sham local hybrid functionals of screened exchange: Combining local and range-separated hybridization.' *The Journal of Chemical Physics*, **129**(12), 124110, 2008.

[204] Toulouse, J., K. Sharkas, E. Brémond, and C. Adamo. 'Communication: Rationale for a new class of double-hybrid approximations in density-functional theory.' *The Journal of Chemical Physics*, **135**(10), 101102, 2011.

[205] Grimme, S. 'Semiempirical hybrid density functional with perturbative second-order correlation.' *The Journal of Chemical Physics*, **124**(3), 034108, 2006.

[206] Werner, H.-J., F. R. Manby, and P. J. Knowles. 'Fast linear scaling second-order Møller–Plesset perturbation theory (MP2) using local and density fitting approximations.' *The Journal of Chemical Physics*, **118**(18), 8149, 2003.

[207] Reine, S., E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansik, S. Hoøst, and P. Salek. 'Variational and robust density fitting of four-center two-electron integrals in local metrics.' *The Journal of Chemical Physics*, **129**(10), 104101, 2008.

[208] Domínguez-Soria, V. D., G. Geudtner, J. L. Morales, P. Calaminici, and A. M. Köster. 'Robust and efficient density fitting.' *The Journal of Chemical Physics*, **131**(12), 124102, 2009.

[209] Seidl, M. 'Strong-interaction limit of density-functional theory.' *Physical Review A*, **60**(6), 4387–4395, 1999.

[210] Wagner, L. O. and P. Gori-Giorgi. 'Electron avoidance: A nonlocal radius for strong correlation.' *Physical Review A*, **90**(5), 2014.

[211] Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. 'Gaussian 09 Revision E.01.' `gaussian.com`. Gaussian Inc. Wallingford CT 2009.

[212] 'TURBOMOLE v6.2.' `turbomole.com`, 2010. A development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007.

[213] Antaya, H., Y. Zhou, and M. Ernzerhof. 'Approximating the exchange energy through the nonempirical exchange-factor approach.' *Physical Review A*, **90**(3), 2014.

[214] Bahmann, H., Y. Zhou, and M. Ernzerhof. 'The shell model for the exchange-correlation hole in the strong-correlation limit.' *The Journal of Chemical Physics*, **145**(12), 124104, 2016.

[215] Vuckovic, S., T. J. P. Irons, A. Savin, A. M. Teale, and P. Gori-Giorgi. 'Exchange–correlation functionals via local interpolation along the adiabatic connection.' *Journal of Chemical Theory and Computation*, **12**(6), 2598–2610, 2016.

[216] Mirtschink, A., C. J. Umrigar, J. D. Morgan, and P. Gori-Giorgi. 'Energy density functionals from the strong-coupling limit applied to the anions of the He isoelectronic series.' *The Journal of Chemical Physics*, **140**(18), 18A532, 2014.

[217] Fuchs, M., Y. M. Niquet, X. Gonze, and K. Burke. 'Describing static correlation in bond dissociation by Kohn–Sham density functional theory.' *The Journal of Chemical Physics*, **122**(9), 094116, 2005.

[218] Woon, D. E. and T. H. Dunning. 'Gaussian basis sets for use in correlated molecular calculations. V. core-valence basis sets for boron through neon.' *The Journal of Chemical Physics*, **103**(11), 4572–4585, 1995.

[219] Seidl, M., S. D. Marino, A. Gerolin, L. Nenna, K. J. H. Giesbertz, and P. Gori-Giorgi. 'The strictly-correlated electron functional for spherically symmetric systems revisited.' *In Preparation*.

[220] Pascale, L. D. 'Optimal transport with Coulomb cost. approximation and duality.' *ESAIM: Mathematical Modelling and Numerical Analysis*, **49**(6), 1643–1657, 2015.

[221] Fabiano, E., P. Gori-Giorgi, M. Seidl, and F. D. Sala. 'Interaction-strength interpolation method for main-group chemistry: Benchmarking, limitations, and perspectives.' *Journal of Chemical Theory and Computation*, **12**(10), 4885–4896, 2016.

[222] Perdew, J. P. 'What do the Kohn-Sham orbital energies mean? how do atoms dissociate?' In 'Density Functional Methods In Physics,' pages 265–308. Springer US, 1985.

[223] Tempel, D. G., T. J. Martínez, and N. T. Maitra. 'Revisiting molecular dissociation in density functional theory: A simple model.' *Journal of Chemical Theory and Computation*, **5**(4), 770–780, 2009.

[224] Gritsenko, O. V. and E. J. Baerends. 'Effect of molecular dissociation on the exchange-correlation Kohn-Sham potential.' *Physical Review A*, **54**(3), 1957–1972, 1996.

[225] Kong, J. and E. Proynov. 'Density functional model for nondynamic and strong correlation.' *Journal of Chemical Theory and Computation*, **12**(1), 133–143, 2016.

[226] Zhang, I. Y., P. Rinke, J. P. Perdew, and M. Scheffler. 'Towards efficient orbital-dependent density functionals for weak and strong correlation.' *Physical Review Letters*, **117**(13), 2016.

[227] Gutlé, C. and A. Savin. 'Orbital spaces and density-functional theory.' *Physical Review A*, **75**(3), 2007.

[228] Grimme, S. and A. Hansen. 'A practicable real-space measure and visualization of static electron-correlation effects.' *Angewandte Chemie International Edition*, **54**(42), 12308–12313, 2015.

[229] Baerends, E. J. and O. V. Gritsenko. 'A quantum chemical view of density functional theory.' *The Journal of Physical Chemistry A*, **101**(30), 5383–5403, 1997.

[230] Levy, M. and J. P. Perdew. 'Hellmann-Feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms.' *Physical Review A*, **32**(4), 2010–2021, 1985.

[231] Ferrón, A. and P. Serra. 'Stability of two-electron diatomic molecules.' *Journal of Physics B: Atomic and Molecular Physics*, **40**(5), 995–1002, 2007.

[232] Amovilli, C. and N. H. March. 'Hookean atom with four electrons: On the formation of a tetrahedral Wigner molecule in the weak trapping limit.' *Physical Review A*, **83**(4), 2011.

[233] Amovilli, C. and N. H. March. 'Electrostatic potentials at the nucleus for isoelectronic series of light atomic ions using the QMC method in relation to DFT.' *Journal of Mathematical Chemistry*, **53**(8), 1725–1732, 2015.

[234] Grabowski, P. E. and K. Burke. 'Quantum critical benchmark for electronic structure theory.' *Physical Review A*, **91**(3), 2015.

[235] Amovilli, C. and N. H. March. 'Variational quantum Monte Carlo results for N_2, N_2^+ and C_2^- utilising the four-dimensional density of Bright Wilson.' *Physics and Chemistry of Liquids*, **55**(3), 281–290, 2016.

[236] Seidl, M., S. Vuckovic, and P. Gori-Giorgi. 'Challenging the Lieb–Oxford bound in a systematic way.' *Molecular Physics*, **114**(7-8), 1076–1085, 2016.

[237] Chan, G. K.-L. and N. C. Handy. 'Optimized Lieb–Oxford bound for the exchange-correlation energy.' *Physical Review A*, **59**(4), 3075–3077, 1999.

[238] Lewin, M. and E. H. Lieb. 'Improved Lieb-Oxford exchange-correlation inequality with a gradient correction.' *Physical Review A*, **91**(2), 2015.

[239] Ditchfield, R. 'Self-consistent perturbation theory of diamagnetism.' *Molecular Physics*, **27**(4), 789–807, 1974.

[240] Wolinski, K., J. F. Hinton, and P. Pulay. 'Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations.' *Journal of the American Chemical Society*, **112**(23), 8251–8260, 1990.

[241] Cheeseman, J. R., G. W. Trucks, T. A. Keith, and M. J. Frisch. 'A comparison of models for calculating nuclear magnetic resonance shielding tensors.' *The Journal of Chemical Physics*, **104**(14), 5497–5509, 1996.

[242] Helgaker, T. and P. Jørgensen. 'An electronic Hamiltonian for origin independent calculations of magnetic properties.' *The Journal of Chemical Physics*, **95**(4), 2595–2601, 1991.

[243] Ruud, K., T. Helgaker, K. L. Bak, P. Jørgensen, and H. J. A. Jensen. 'Hartree–Fock limit magnetizabilities from London orbitals.' *The Journal of Chemical Physics*, **99**(5), 3847–3859, 1993.

[244] Helgaker, T., M. Jaszuński, and K. Ruud. '*Ab Initio* methods for the calculation of NMR shielding and indirect spin-spin coupling constants.' *Chemical Reviews*, **99**(1), 293–352, 1999.

[245] Soncini, A. and P. Fowler. 'Non-linear ring currents: effect of strong magnetic fields on π-electron circulation.' *Chemical Physics Letters*, **400**(1-3), 213–220, 2004.

[246] London, F. 'Théorie quantique des courants interatomiques dans les combinaisons aromatiques.' *Journal de Physique et le Radium*, **8**(10), 397–409, 1937.

[247] Ditchfield, R., D. P. Miller, and J. A. Pople. 'Self-consistent molecular-orbital methods. VII. convergence of Gaussian expansions of Slater-type atomic orbitals in calculations of first- and second-order properties.' *The Journal of Chemical Physics*, **53**(2), 613–619, 1970.

[248] Hall, G. G. 'Gauge invariant Gaussian orbitals and the *ab initio* calculation of diamagnetic susceptibility for molecules.' *International Journal of Quantum Chemistry*, **7**(1), 15–25, 1973.

[249] Epstein, S. T. 'Gauge invariance, current conservation and GIAOs.' *The Journal of Chemical Physics*, **58**(4), 1592–1595, 1973.

[250] Dirac, P. A. M. 'The quantum theory of the electron.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **117**(778), 610–624, 1928.

[251] Pauli, W. 'Zur quantenmechanik des magnetischen elektrons.' *Zeitschrift für Physik*, **43**(9-10), 601–623, 1927.

[252] Ditchfield, R., D. P. Miller, and J. A. Pople. 'Self-consistent molecular orbital methods. XI. molecular orbital theory of NMR chemical shifts.' *The Journal of Chemical Physics*, **54**(10), 4186–4193, 1971.

[253] Ditchfield, R. 'Molecular orbital theory of magnetic shielding and magnetic susceptibility.' *The Journal of Chemical Physics*, **56**(11), 5688–5691, 1972.

[254] Ditchfield, R. 'On molecular orbital theories of NMR chemical shifts.' *Chemical Physics Letters*, **15**(2), 203–206, 1972.

[255] Tellgren, E. I., T. Helgaker, and A. Soncini. 'Non-perturbative magnetic phenomena in closed-shell paramagnetic molecules.' *Physical Chemistry Chemical Physics*, **11**(26), 5489, 2009.

[256] Tellgren, E. I., S. S. Reine, and T. Helgaker. 'Analytical GIAO and hybrid-basis integral derivatives: application to geometry optimization of molecules in strong magnetic fields.' *Physical Chemistry Chemical Physics*, **14**(26), 9492, 2012.

[257] Lange, K. K., E. I. Tellgren, M. R. Hoffmann, and T. Helgaker. 'A paramagnetic bonding mechanism for diatomics in strong magnetic fields.' *Science*, **337**(6092), 327–331, 2012.

[258] Tellgren, E. I. and H. Fliegl. 'Non-perturbative treatment of molecules in linear magnetic fields: Calculation of anapole susceptibilities.' *The Journal of Chemical Physics*, **139**(16), 164118, 2013.

[259] Tellgren, E. I., A. M. Teale, J. W. Furness, K. K. Lange, U. Ekström, and T. Helgaker. 'Non-perturbative calculation of molecular magnetic properties within current-density functional theory.' *The Journal of Chemical Physics*, **140**(3), 034101, 2014.

[260] Reimann, S., U. Ekström, S. Stopkowicz, A. M. Teale, A. Borgoo, and T. Helgaker. 'The importance of current contributions to shielding constants in density-functional theory.' *Physical Chemistry Chemical Physics*, **17**(28), 18834–18842, 2015.

[261] Reynolds, R. D. and T. Shiozaki. 'Fully relativistic self-consistent field under a magnetic field.' *Physical Chemistry Chemical Physics*, **17**(22), 14280–14283, 2015.

[262] Stopkowicz, S., J. Gauss, K. K. Lange, E. I. Tellgren, and T. Helgaker. 'Coupled-cluster theory for atoms and molecules in strong magnetic fields.' *The Journal of Chemical Physics*, **143**(7), 074110, 2015.

[263] Furness, J. W., J. Verbeke, E. I. Tellgren, S. Stopkowicz, U. Ekström, T. Helgaker, and A. M. Teale. 'Current density functional theory using meta-generalized gradient exchange-correlation functionals.' *Journal of Chemical Theory and Computation*, **11**(9), 4169–4181, 2015.

[264] Furness, J. W., U. Ekström, T. Helgaker, and A. M. Teale. 'Electron localisation function in current-density-functional theory.' *Molecular Physics*, **114**(7-8), 1415–1422, 2016.

[265] Hampe, F. and S. Stopkowicz. 'Equation-of-motion coupled-cluster methods for atoms and molecules in strong magnetic fields.' *The Journal of Chemical Physics*, **146**(15), 154105, 2017.

[266] Obara, S. and A. Saika. 'Efficient recursive computation of molecular integrals over Cartesian Gaussian functions.' *The Journal of Chemical Physics*, **84**(7), 3963, 1986.

[267] Honda, M., K. Sato, and S. Obara. 'Formulation of molecular integrals over Gaussian functions treatable by both the Laplace and Fourier transforms of spatial operators by using derivative of Fourier-kernel multiplied Gaussians.' *The Journal of Chemical Physics*, **94**(5), 3790, 1991.

[268] Kiribayashi, S., T. Kobayashi, M. Nakano, and K. Yamaguchi. 'Self-consistent-field calculations of molecular magnetic properties using gauge-invariant atomic orbitals.' *International Journal of Quantum Chemistry*, **75**(4-5), 637–643, 1999.

[269] Ishida, K. 'ACE algorithm for the rapid evaluation of the electron-repulsion integral over Gaussian-type orbitals.' *International Journal of Quantum Chemistry*, **59**(3), 209–218, 1996.

[270] Ishida, K. 'Molecular integrals over the gauge-including atomic orbitals.' *The Journal of Chemical Physics*, **118**(11), 4819, 2003.

[271] Dupuis, M. 'Evaluation of molecular integrals over Gaussian basis functions.' *The Journal of Chemical Physics*, **65**(1), 111, 1976.

[272] King, H. F. and M. Dupuis. 'Numerical integration using Rys polynomials.' *Journal of Computational Chemistry*, **21**(2), 144–165, 1976.

[273] McMurchie, L. E. and E. R. Davidson. 'One- and two-electron integrals over Cartesian Gaussian functions.' *Journal of Computational Chemistry*, **26**(2), 218–231, 1978.

[274] Tachikawa, M. and M. Shiga. 'Evaluation of atomic integrals for hybrid Gaussian type and plane-wave basis functions via the McMurchie–Davidson recursion formula.' *Physical Review E*, **64**(5), 056706, 2001.

[275] Kuchitsu, T., J. Okuda, and M. Tachikawa. 'Evaluation of molecular integral of Cartesian Gaussian type basis function with complex-valued center coordinates and exponent via the McMurchie–Davidson recursion formula and its application to electron dynamics.' *International Journal of Quantum Chemistry*, **109**(3), 540–548, 2009.

[276] LONDON. 'A quantum chemistry program for plane–wave/GTO hybrid basis sets and finite magnetic field calculations.' `londonprogram.org`.

[277] BAGEL. 'Brilliantly advanced general electronic-structure library.' `nubakery.org`. Published under the GNU General Public License.

[278] Flocke, N. 'On the use of shifted Jacobi polynomials in accurate evaluation of roots and weights of Rys polynomials.' *The Journal of Chemical Physics*, **131**(6), 064107, 2009.

[279] Frisch, M. J., J. A. Pople, and J. S. Binkley. 'Self-consistent molecular orbital methods 25. supplementary functions for Gaussian basis sets.' *The Journal of Chemical Physics*, **80**(7), 3265, 1984.

[280] Binkley, J. S., J. A. Pople, and W. J. Hehre. 'Self-consistent molecular orbital methods. 21. small split-valence basis sets for first-row elements.' *Journal of the American Chemical Society*, **102**(3), 939–947, 1980.

[281] Gill, P. M. W. 'Molecular integrals over Gaussian basis functions.' In 'Advances in Quantum Chemistry,' pages 141–205. Elsevier BV, 1994.

[282] Johnson, B. G., P. M. W. Gill, J. A. Pople, and D. J. Fox. 'Computing molecular electrostatic potentials with the PRISM algorithm.' *Chemical Physics Letters*, **206**(1-4), 239–246, 1993.

[283] Gill, P. M. W., B. G. Johnson, J. A. Pople, and S. W. Taylor. 'Modeling the potential of a charge distribution.' *The Journal of Chemical Physics*, **96**(9), 7178–7179, 1992.

[284] Head-Gordon, M. and J. A. Pople. 'A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations.' *The Journal of Chemical Physics*, **89**(9), 5777, 1988.

[285] Rys, J., M. Dupuis, and H. F. King. 'Computation of electron repulsion integrals using the Rys quadrature method.' *Journal of Computational Chemistry*, **4**(2), 154–157, 1983.

[286] Johnson, B. G., P. M. W. Gill, and J. A. Pople. 'The efficient transformation of (**m**0|**n**0) to (**ab**|**cd**) two–electron repulsion integrals.' *Chemical Physics Letters*, **206**(1-4), 229–238, 1993.

[287] Makowski, M. 'Simple yet powerful techniques for optimization of horizontal recursion steps in Gaussian-type two-electron integral evaluation algorithms.' *International Journal of Quantum Chemistry*, **107**(1), 30–36, 2006.

[288] Ryu, U., Y. S. Lee, and R. Lindh. 'An efficient method of implementing the horizontal recurrence relation in the evaluation of electron repulsion integrals using Cartesian Gaussian functions.' *Chemical Physics Letters*, **185**(5-6), 562–568, 1991.

[289] Schlegel, H. B. and M. J. Frisch. 'Transformation between Cartesian and pure spherical harmonic Gaussians.' *International Journal of Quantum Chemistry*, **54**(2), 83–87, 1995.

[290] Obara, S. and A. Saika. 'General recurrence formulas for molecular integrals over Cartesian Gaussian functions.' *The Journal of Chemical Physics*, **89**(3), 1540, 1988.

[291] Shavitt, I. 'The Gaussian function in calculations of statistical mechanics and quantum mechanics.' In 'Methods in Computational Physics,' volume 3, pages 1–45. Academic Press New York, 1963.

[292] Saunders, V. R. 'An introduction to molecular integral evaluation.' In 'Computational Techniques in Quantum Chemistry and Molecular Physics,' pages 347–424. Springer Netherlands, 1975.

[293] Boys, S. F. 'Electronic wave functions. I. a general method of calculation for the stationary states of any molecular system.' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **200**(1063), 542–554, 1950.

[294] Helgaker, T., P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons Inc., 2000.

[295] Čársky, P. and M. Polášek. 'Incomplete Gamma Fm(x) functions for real negative and complex arguments.' *Journal of Computational Physics*, **143**(1), 259–265, 1998.

[296] Mathar, R. J. 'Numerical representations of the incomplete gamma function of complex-valued argument.' *Numerical Algorithms*, **36**(3), 247–264, 2004.

[297] Ishida, K. 'Accurate and fast algorithm of the molecular incomplete gamma function with a complex argument.' *Journal of Computational Chemistry*, **25**(5), 739–748, 2004.

[298] Colle, R., A. Fortunelli, and S. Simonucci. 'A mixed basis set of plane waves and Hermite–Gaussian functions. analytic expressions of prototype integrals.' *Il Nuovo Cimento D*, **9**(8), 969–977, 1987.

[299] Colle, R., A. Fortunelli, and S. Simonucci. 'Hermite–Gaussian functions modulated by plane waves: a general basis set for bound and continuum states.' *Il Nuovo Cimento D*, **10**(7), 805–818, 1988.

[300] Golub, G. H. and J. H. Welsch. 'Calculation of Gauss quadrature rules.' *Mathematics of Computation*, **23**(106), 221–221, 1969.

[301] Saylor, P. E. and D. C. Smolarski. 'Why Gaussian quadrature in the complex plane?' *Numerical Algorithms*, **26**(3), 251–280, 2001.

[302] Wilf, H. S. *Mathematics for the physical sciences*. John Wiley & Sons Inc., 1962.

[303] Sack, R. A. and A. F. Donovan. 'An algorithm for Gaussian quadrature given modified moments.' *Numerische Mathematik*, **18**(5), 465–478, 1971.

[304] Horn, R. A. and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 1985.

[305] Dongarra, J. J., J. W. Demmel, and S. Ostrouchov. 'LAPACK: A linear algebra library for high-performance computers.' In 'Computational Statistics,' pages 23–28. Physica-Verlag HD, 1992.

[306] Lindh, R., U. Ryu, and B. Liu. 'The reduced multiplication scheme of the Rys quadrature and new recurrence relations for auxiliary function based two-electron integral evaluation.' *The Journal of Chemical Physics*, **95**(8), 5889, 1991.

[307] Schmelcher, P. and L. S. Cederbaum. 'Molecules in strong magnetic fields: Properties of atomic orbitals.' *Physical Review A*, **37**(3), 672–681, 1988.

[308] Kubo, A. 'The hydrogen molecule in strong magnetic fields: Optimizations of anisotropic Gaussian basis sets.' *The Journal of Physical Chemistry A*, **111**(25), 5572–5581, 2007.

[309] Adamowicz, L., E. I. Tellgren, and T. Helgaker. 'Non-Born–Oppenheimer calculations of the HD molecule in a strong magnetic field.' *Chemical Physics Letters*, **639**, 295–299, 2015.

[310] Gill, P. M. W. and J. A. Pople. 'The PRISM algorithm for two–electron integrals.' *International Journal of Quantum Chemistry*, **40**(6), 753–772, 1991.

[311] Almlöf, J. and P. R. Taylor. 'Atomic natural orbital (ANO) basis sets for quantum chemical calculations.' In 'Advances in Quantum Chemistry,' pages 301–373. Elsevier BV, 1991.