Chapter 1

Theory and Implementation

1.1 Electronic structure methods

Electronic structure methods apply the principles of quantum mechanics to the evaluation of electron movement and position, thereby allowing chemists to derive the properties and interactions of molecules. At the most fundamental level, the wavefunction (Ψ) holds the description of a quantum system. The probability density is given by $|\Psi^2|$, such that the integral of $|\Psi^2|$ over a region of space returns the probability that a system will be found in that region. Therefore, for $|\Psi^2|$ that has been normalised, integrating over all space gives the probability of 1.

Operators acting on Ψ yield observable properties of the system. The operator returning the energy of the system is called the the Hamiltonian operator (**H**). Erwin Schrödinger first proposed his equation for the description of a quantum system using its wavefunction in 1929 [**?**]. The time-independent equation is given in short-form:

$$\mathbf{H}\Psi = E\Psi \tag{1.1}$$

where the Hamiltonian operator **H** is an eigenvalue, the wavefucntion Ψ is the eigenfunction and E is a scalar denoting the energy of the system. The general form of the Hamiltonian is given by:

$$\mathbf{H} = -\sum \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum \frac{\hbar^2}{2m_k}\nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \tag{1.2}$$

where all electrons are encompassed by i and j, and all nuclei by k and l. $\hbar$ is the reduced Planck's constant ($\hbar = \frac{h}{2\pi}$), m_e is the mass of an electron, m_k is the mass of the nucleus k, e is the charge of an electron, Z_k is the atomic number of k and r_{xy} is the distance

between particles x and y. ∇^2 refers to the Laplacian operator, which describes the divergence of the gradient of a field.In Cartesian space, this is defined as the second derivative of the gradient in the three dimensions:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{1.3}$$

The first and second terms of equation (1.2) correspond to the kinetic energy of the electrons and the nuclei, respectively. Electron - nuclear attraction is described by the third term, followed by the interelectronic and internuclear repulsive terms. The attractive and repulsive terms are potential energy terms and can be described classically. The kinetic energy terms are expressed as the eigenvalue of the kinetic energy operator ($\mathbf{T}$):

$$\mathbf{T} = -\frac{\hbar^2}{2m}\nabla^2 \tag{1.4}$$

When using **au!** (**au!**), the electronic mass, charge and reduced Planck's constant are reduced to 1. Equation 1.2 can therefore be written as:

$$\mathbf{H} = \mathbf{T}_e + \mathbf{T}_{nuc} + \mathbf{V}_{e-nuc} + \mathbf{V}_{e-e} + \mathbf{V}_{nuc-nuc} \tag{1.5}$$

1.1.1 Born-Oppenheimer Approximation

The motion of elections and nuclei are coupled in real systems, as electronic charge flows dynamically in response to the change in nuclear position and

The coupled motion of the particles in the system is

in the expressions of Schŕodinger's equation thus far.

1.1.2 Hartree Fock

Hartree Fock **HF!** (**HF!**)

1.1.2.1 Variational Principle

1.1.2.2 Open shell systems

[PLEASE REVAMP ALL OF THIS AND FILL IN DETAIL] The forced pairing of electrons of opposing spin into a shared orbital is referred to as the Restricted (RB3LYP) scheme. For systems without unpaired electrons, or "closed shell", this treatment is sufficient. For radicals and other species with unpaired electron spin such as transition metal complexes, an alternative model allowing singly occupied orbitals must be adopted. The

Restricted-Open (ROB3LYP) scheme maintains electron pairing within orbitals except in the case of the highest occupied molecular orbital (HOMO). This is singly occupied. An alternative model is Unrestricted (UB3LYP), where all electrons are unpaired and reside in their own orbitals (Figure 9). A caveat of the unrestricted model is its susceptibility to spin contamination, which has consequences at large bond separations where the bond has not completely broken.

1.1.3 Electron correlation

1.1.4 Post-Hartree Fock methods

1.1.4.1 MP2

1.1.5 Density functional theory

Solutions to the Schr'odinger's equation using reference

See here for a nice sum of the weaknesses: Some Fundamental Issues in Ground-State Density, Perdew 2009 (https://pubs.acs.org/doi/10.1021/ct800531s) Functional Theory: A Guide for the Perplexed [PLEASE REVAMP ALL OF THIS AND FILL IN DETAIL] Density functional theory derives from the Thomas-Femi-Dirac model, whereby the electron correlation is modelled via functionals of the electron density.[38,39] The total energy is defined by functionals split into the following terms:

$$E = E^T + E^V + E^J + E^X C$$

Where E^T is the kinetic energy term, arising from electron motion; E^V is the potential energy term arising from nuclear-electron attraction and nuclear-nuclear repulsion; E^J is the columbic repulsion term arising from electron-electron repulsion and E^{XC} is the exchange correlation term containing the remainder of the electron-electron interactions. The first three terms are purely classical, and correspond to the classical energy of the charge distribution (ρ). The exchange term is non-classical encompasses the exchange energy due to the antisymmetry of the wavefunction, and dynamic correlation of electron motion. It can be further divided into the exchange and correlation components: In the context of this report, the B3LYP functional will be used where the exchange is described as follows:

Here $c_0 = 0.2$, $c_X = 0.72$ and $c_C = 0.81$. The coefficient c_0 allows mixing of E_{HF^x} (Hartree Fock) and E_{LDA^x} (Local Density Approximation). E_{VWN3^C} (VWN3 local correlation) is mixed with E_{LYP^C} (Lee, Yang, Parr correlation function) via c_C. E_{B88^x} is Becke's

gradient-corrected exchange functional.

ωB97x-d is a range separated hybrid with D (self consistent exchange). It is more computationally demanding than just post HF style perturbation theory exchange (as in B3LYP, using MP2 exchange for example - fact check that). And according to the paper by Najibi and Goerigk even the D3 correction isn't better than the perturbation theory style exchange, for a massive amount of datasets that they tested. SO why even use this in the first place? Look to the origianl paper where it was published and see how they tout this method. Use those reasons to justify having chosen it to use on some of my systems.

1.1.5.1 Hohenburg-Kohn formalism

1.1.5.2 Kohn-Sham Equations

1.1.5.3 Exchange-correlation functionals

B3LYP! (**B3LYP!**) **wb97xd!** (**wb97xd!**)

1.1.5.4 Dispersion correction

Grimme

1.1.6 Basis sets

[FIX ME] A basis set is the collection of mathematical basis functions used in linear combination to construct the molecular orbitals. Split valence basis sets describe the core electrons with fewer basis functions than the interacting valence electrons, as they are not as significant in bonding or intermolecular interactions. In this study, the widely used Pople basis sets will be applied (Table 1. Examples of split valence basis sets.).40,41

Basis set superposition error (BSSE) is a false lowering of the energy that can occur when two species in a system approach one another to form a complex. Particle A borrows the extra basis functions belonging to particle B and an artificial stabilisation is observed. The error arises from the inconsistency in treatment between the individual particles at long separations and the complex at short distances. The effect is particularly pronounced for smaller basis sets. Counterpoise correction I used to circumvent BSSE, at the expense of higher computational resources required for the calculation.

1.1.7 Thermochemistry

(And vibrational frequencies)

1.1.8 (Transition State Theory & Free Energy Calculations)

1.1.9 Solvent models

1.2 Analysis

1.2.1 Topology analysis using QTAIM! (QTAIM!)

*paraphrase this: Topology analysis Theory

Some real space functions in Multiwfn are available for topology analysis, such as electron density, its Laplacian, orbital wavefunction, ELF, LOL ...

The topology analysis technique proposed by Bader was firstly used for analyzing electron density in "atoms in molecules" (AIM) theory, which is also known as "the quantum theory of atoms in molecules" (QTAIM), this technique has also been extended to other real space functions, e.g. the first topology analysis research of ELF for small molecules is given by Silvi and Savin, see Nature, 371, 683. In topology analysis language, the points at where gradient norm of function value is zero (except at infinity) are called as critical points (CPs), CPs can be classified into four types according to how many eigenvalues of Hessian matrix of real space function are negative.

(3,-3): All three eigenvalues of Hessian matrix of function are negative, namely the local maximum. For electron density analysis and for heavy atoms, the position of (3,-3) are nearly identical to nuclear positions, hence (3,-3) is also called nuclear critical point (NCP). Generally the number of (3,-3) is equal to the number of atoms, only in rarely cases the former can be more than (e.g. Li2) or less than (e.g. KrH+) the latter.

(3,-1): Two eigenvalues of Hessian matrix of function are negative, namely the second-order saddle point. For electron density analysis, (3,-1) generally appears between attractive atom pairs and hence commonly called as bond critical point (BCP). The value of real space functions at BCP have great significance, for example the value of XXX and the sign of at BCP are closely related to bonding strength and bonding type respectively in analogous bonding type (The Quantum Theory of Atoms in Molecules-From Solid State to DNA and Drug Design, p11); the potential energy density at BCP has been shown to be highly correlated with hydrogen bond energies (Chem. Phys. Lett., 285, 170); local information entropy at BCP is a good indicator of aromaticity (Phys. Chem. Chem. Phys., 12, 4742).

(3,+1): Only one eigenvalue of Hessian matrix of function is negative, namely first-order saddle point (like transition state in potential energy surface). For electron density

analysis, (3,+1) generally appears in the center of ring system and displays steric effect, hence (3,+1) is often named as ring critical point (RCP).

(3,+3): None of eigenvalues of Hessian matrix of function are negative, namely the local minimum. For electron density analysis, (3,+3) generally appears in the center of cage system (e.g. pyramid P4 molecule), hence is often referred to as cage critical point (CCP).

(table 1.1)

Table 1.1: Features of different types of critical point from **QTAIM!** topological analysis.

Critical Point	Label	Derivation	Attribute	Representation
Nuclear (**NCP!**)	(3,-3)	All 3 eivalues of the Hessian matrix are negative	Local maximum	Atomic nuclei
Bonding (**BCP!**)	(3,-1)	2 negative eigenvalues of Hessian matrix	2nd order saddle point	Bonding site
Ring (**RCP!**)	(3,+1)	1 negative eigenvalue of Hessian matrix	1st order saddle point	Steric point or centre of ring system
Cage (**CCP!**)	(3,+3)	No negative eigenvalue of Hessian matrix	Local minimum	Centre of cage system

The positions of CPs are searched by Newton method, one need to assign an initial guess point, then the Newton iteration always converge to the CP that is closest to the guess point. By assigning different guesses and doing iteration for each of them, all CPs could be found. Once searches of CPs are finished, one should use Poincaré-Hopf relationship to verify if all CPs may have been found, the relationship states that (for isolated system)

$n(3,-3) - n(3,-1) + n(3,+1) - n(3,+3) = 1$

If the relationship is unsatisfied, then some of CPs must be missing, you may need to try to search those CPs by different guesses. However even if the relationship is satisfied, it does not necessarily mean that all CPs have been found. Notice that the function spaces of ELF/LOL and Laplacian of XX are much more complex than XXX, it is very difficult to locate all CPs for these functions, especially for middle and large system, so, you can stop trying for searching CPs once all CPs that you are interested in have been found.

The maximal gradient path linking BCP and associated two local maxima of density is termed as "bond path", which reveals atomic interaction path for all kinds of bonding. The collection of bond paths is known as molecular graph, which provides an unambiguous definition of molecular structure. Bond path can be straight line or curve, obviously for the latter case the length of bond path is longer than the sum of the distances between BCP and

associated two (3,-3) CPs.

Let us see an example. In the complex shown below, the imidazole plane is vertical to magnesium porphyrin plane, the nitrogen in imidazole coordinated to magnesium. Magenta, orange and yellow spheres correspond to (3,-3), (3,-1) and (3,+1) critical points, brown lines denote bond paths.

1.2.2   (Transition state searches)

[Maybe this goes into a smaller metholodgy section?] Transition state searches are called through the Opt=TS, QST or QST3 keywords. The Opt=TS method in GAUSSIAN attempts to optimise the given "guess" geometry to a transition state. The guess structure can be obtained from a geometry scan, manually constructed or generated using the QST2 function. In many cases, a TS alone will not be able to isolate the suitable transition state and is usually used in conjunction the QST2 or QST3 methods, and combined with other techniques such as frequency calculations. The QST2 option is able to generate a transition state geometry using the Synchronous Transit Quasi-Newton (STQN) method42. Here the transition geometry is midway between a given reactant and product. The corresponding atoms labels must match in both the starting and end products. QST3 performs similarly, but also considers a guess transition state (Figure 10). It is widely acknowledged that transition state searching is challenging; in addition to the techniques above, the task requires perhaps a certain measure of chemical intuition.

Frequency calculations

1.2.3   (Intrinsic Reaction Coordinate)

[Maybe this goes into a smaller metholodgy section?] IRC calculations begin at the saddle point and descend the PES in both the forwards and backwards direction of the reaction co-ordinate dictated by the normal mode of the imaginary frequency. In a similar manner to a geometry scan, geometry optimisations are performed at each step point. Its purpose is to connect the two minima leading to the found transition state, thereby confirming whether the found transition state corresponds to your reactants and products of interest. IRC calculations are called via the IRC keyword, with specifications of whether the forward or backward reaction is to be scanned, step size and maximum number of steps allowed.

1.2.4   (PES Scans)

[Maybe this goes into a smaller methodology section?] Relaxed potential energy surface (PES) scans, or geometry scans are used to probe the local energy landscape corresponding

to specific change in geometry. During the course of a scan, a selected bond length, angle or dihedral is adjusted in incremental steps, as specified by the given scan parameters. At each step, the adjusted parameter is frozen and a geometry optimisation is performed, allowing the rest of the system to relax around the modified bond. Each scan yields a PES of the explored pathway, presented in a reaction co-ordinate diagram (see Figure 17, section 2.4.1). An energy maximum followed by a trough indicates a transition state and intermediate reaction product, respectively. The structural co-ordinates at the points of interest are extracted and used for subsequent frequency calculations, transition state searches and validated using intrinsic reaction co-ordinate methods. To explore the predicted degradation mechanisms, the scanning parameter was assigned to the bond undergoing the most significant transformation during a particular step of the mechanism. In the case that more than one significant bond was altered, multiple scans with different bond specifications were compared. Geometry scans were performed on the optimised reactant geometry using the Opt=ModRedundant keyword.

1.3 Spectroscopy

1.3.1 Nuclear Magnetic Resonance Spectra

Refer them to an actual resource, so you don't have to explain the theory of **NMR!** (**NMR!**), but give an overview of the experimental, then give the theoretical detail about how Gaussian calculates it, and any discrepancies between experimental and calculated.

Comment on the different methods used to calculate NMR parameters - a tiny literature review, if you will, and say which you'll be using, why - even if it's becuase of the fact that it's built into Gaussian (obviously big it up, if that is the case) - and state any caveats.

Maybe have a look at: Accurate Calculation of NMR Chemical Shifts, Jurgen Gauss, 1995 as a basis, but you'll also need something more current. Any review papers out there?

1.3.1.1 Computational implementation

1.3.2 Infra-Red Spectra

Intro to the history spiel.

1.3.2.1 Computational implementation

Caveats and scaling factors.