



UNIVERSIDAD DE OVIEDO

FACULTAD DE QUÍMICA

**BASIS SETS IN COMPUTATIONAL
CHEMISTRY: ANALYSIS OF THE SECOND
PERIOD DIATOMIC FLUORIDES**

(Physical Chemistry)

BACHELOR THESIS IN CHEMISTRY

Covadonga García Pérez

Oviedo, July, 2020

Table of contents

OBJECTIVES	- 1 -
1 INTRODUCTION	- 2 -
2 THEORY AND METHODS	- 3 -
2.1- <i>Approximation methods.....</i>	<i>- 4 -</i>
2.2- <i>The Born-Oppenheimer approximation.....</i>	<i>- 5 -</i>
2.3- <i>Hartree-Fock theory</i>	<i>- 6 -</i>
2.4- <i>Roothaan-Hall equations.....</i>	<i>- 9 -</i>
2.5- <i>Post Hartree-Fock methods</i>	<i>- 11 -</i>
2.6- <i>Basis sets</i>	<i>- 11 -</i>
2.6.1- <i>Types of basis functions.....</i>	<i>- 12 -</i>
2.6.2- <i>Classification of Gaussian Type Basis Sets</i>	<i>- 13 -</i>
2.6.3- <i>Contracted Basis Functions</i>	<i>- 14 -</i>
2.6.4- <i>Optimization of basis set parameters.....</i>	<i>- 15 -</i>
2.6.5- <i>Complete Basis Sets.....</i>	<i>- 16 -</i>
2.6.6- <i>Property Basis Sets</i>	<i>- 16 -</i>
3 COMPUTATIONAL METHODOLOGY	- 16 -
3.1- <i>Level of theory.....</i>	<i>- 17 -</i>
3.2- <i>Choosing a suitable Basis Set</i>	<i>- 17 -</i>
3.3- <i>Systems studied.....</i>	<i>- 20 -</i>
4 RESULTS AND DISCUSSION.....	- 22 -
5 CONCLUSIONS.....	- 29 -
6 BIBLIOGRAFY	- 30 -
6.1- <i>Web Resources</i>	<i>- 31 -</i>
7 APPENDIX	- 31 -
7.1- <i>Orca input.....</i>	<i>- 31 -</i>
7.2- <i>Energy tables.....</i>	<i>- 31 -</i>
7.3- <i>Geometries</i>	<i>- 38 -</i>

OBJECTIVES

This bachelor thesis lies within the branch of physical chemistry, being focused in the subfield of computational chemistry. Computational chemistry uses computers to solve, usually in an approximate manner, theoretical models of chemical problems. In its most simple form, when it is applied to model the behavior and properties of small molecules, theoretical chemistry and computational modeling (TCCM in the following) tries to solve the non-relativistic Schrödinger's equation (SE) by using general purpose computer programs. After seventy years, these codes have become widely accessible, and mastering their use has turned into one of the skills that a practicing chemist must dominate. To do that, a basic knowledge of quantum chemistry as well as of the basic strategies used to approximate the solution of the SE is needed. In the most successful methods devised, the so-called variational strategies, the solution of the SE, a partial differential equation, is mapped onto a linear-algebra problem by transforming the infinite dimensional Hilbert space into a finite one. With this approximate mapping, all functions are built as linear combinations of a set of fixed ones, and the problem reduces to finding the (linear) coefficients of these combinations. Since linear algebra is particularly suited for being programmed in standard computers, variational methods are very efficient. We call these fixed functions the *basis set* in TCCM. Learning its meaning as well as its impact on the calculation of the electronic structure of simple molecules is the aim of this thesis

More specifically, the objective of this thesis is to understand the meaning of a basis set, and to study and compare different basis functions in a limited set of diatomic molecules. We have employed probably the minimum level of theory that allows for a sensible description of the electronic structure of molecules: the Hartree-Fock (HF) method. This allows us to keep our eye on the role of the basis set, and not on that of the many other knobs that can be tuned in today's TCCM codes. As the latter are regarded, we have used the ORCA¹ (Version 4.1.2) package, which is offered for free for academic purposes. In order to interpret results as well as to prepare the input file to ORCA, we have also used the Avogadro² (Version 1.2.0) suite

1 INTRODUCTION

Chemistry is an experimental science that deals with the constructions, transformations and properties of molecular systems. However, since the advent of quantum mechanics and its first applications to chemical problems³, it has become clear that the basic laws of physics are able to rationalize the chemical world. This was soon recognized to open a new window to the chemical community: the possibility to model chemical phenomena *ab initio*: from our understanding of the laws that govern them. As Dirac⁴ put it, Chemistry had become a branch of Physics. Theoretical chemistry is, since then, the subfield where mathematical methods are combined with the fundamental laws of physics to study processes of chemical relevance.⁵ (Jensen)

Computational chemistry can be said to be a branch of theoretical chemistry whose goal is to create efficient mathematical approximations and computer programs to model chemical problems and, in its simplest form, to calculate the properties of molecules. It is to this very limited scope that we will refer in the following. The main goal in this very specific topic of TCCM is to solve the Schrödinger equation for a given system. This will provide its wavefunction, an object that contains all the information that we can extract from the system. To solve the SE different methods can be employed. Those methods that do not include empirical or semi-empirical parameters, the ones derived directly from theoretical principles, are the so-called *ab initio methods*. Since the SE cannot be solved exactly except in a few academic examples, even though *ab initio* methods do not use empirical parameters, they necessarily assume a number of mathematical approximations.

The simplest level of *ab initio* calculation of the electronic structure of a molecule employs the Hartree-Fock theory. Being a variational procedure, the approximate energies calculated in HF are equal or greater than the exact ones. In the Hartree-Fock theory the wavefunction is approximated as a Slater determinant composed of a set of occupied spatial orbitals, one per electron. The spatial orbitals are approximated as a linear combination of a set of fixed basis functions, used to expand the molecular orbitals. The basis functions are some of the most important factors that have to be chosen in the calculations as the selection of one or another set will lead to a better or worse accurate solution (and to a shorter or longer computer run). The parameters that characterize the basis functions are optimized following the variational principle once and for all, usually with the total energy, and no other properties, as a target. This means that basis sets that perform better energetically need not to be better when other observables are examined.

2 THEORY AND METHODS

One of the main interests in computational chemistry is to find solutions to the Schrödinger equation. The time-independent Schrödinger equation can be written as:

$$H\Psi = E\Psi. \quad (1)$$

In this equation H is the Hamiltonian operator built as the sum of the kinetic and potential energy of a system, Ψ is the wavefunction - it depends on the coordinates and spins of the electrons and the nuclei and it contains all the information about a system - and E is the total energy of the system.

The numerical solution of Schrödinger equation remains a difficult task. Exact solutions of the equation are limited to one-electron systems. For more than one electron system, approximations must be introduced. The approximations are used to reduce the equations to a form that can be solved, resulting in losing some degree of accuracy. The approaches that do not use any empirical data are called *ab initio methods*. When a limited amount of empirical data is used the methods are called *semi-empirical methods*. One of the first developed and easiest *ab initio method* is called Hartree-Fock (HF) theory, although nowadays most of the production calculations are performed using the Density Functional Theory (DFT).

In the Hartree-Fock theory the electron-electron repulsion is taken into account in an averaged way, so that each electron feels the field exerted by the average of the rest of the electrons. This is a variational calculation, therefore the obtained approximate energies - expressed in terms of the systems wavefunction - are always equal to or greater than the exact energy and tend to a limiting value called the Hartree-Fock limit.

To set up a Hartree-Fock calculation it is necessary to understand a hierarchy of decisions that take us from Eq. 1 for a molecule to a working computer code and numerical solution. Some of them are:

- The category of approximation method used
- The Born-Oppenheimer approximation
- The neglect of relativistic effects
- The wavefunction written as a Slater determinant
- The assumption that the variational solution is a linear combination of a finite number of basis functions

They are briefly summarized in the following sections.

2.1- Approximation methods

Approximate methods need to be used when exact solutions to the SE cannot be found. Two categories are widely used: the variational method and perturbation theory. Only the variational method is here below explained as it is the one used for the Hartree-Fock theory.

The variational methods are based on the variational principle that introduces a lower bound for the expectation value of the Hamiltonian computed from a well-behaved, otherwise general state vector. It states that any well-defined normalized function will provide an energy value equal or higher than the exact one.

$$\langle \Psi | A | \Psi \rangle \geq E_0, \quad (2)$$

where E_0 is the exact ground state energy.

This principle is applied by taking a “trial” wavefunction that contains some adjustable parameters called “variational parameters”. The variational parameters are adjusted until the energy of the trial wavefunctions is minimized. The resulting trial wavefunction and its corresponding energy are variational approximations to the exact wavefunction and energy.

The variational principle is easily demonstrated by noticing that any trial function can be written as a linear combination of the exact eigenfunctions of the problem (Ψ_i). The trial wavefunction can be written as:

$$\Phi = \sum_i c_i \Psi_i,$$

And the approximate energy that corresponds to this wavefunction is:

$$E(\Phi) = \frac{\langle \Phi | H | \Phi \rangle}{\langle \Phi | \Phi \rangle},$$

Substituting the expansion over the exact wavefunctions:

$$E(\Phi) = \frac{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \langle \Psi_i | H | \Psi_j \rangle}{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \langle \Psi_i | \Psi_j \rangle},$$

Taking into account the fact that $H\Psi_i = E_i\Psi_i$ and the orthonormality of the Ψ_i , then:

$$E(\Phi) = \frac{\sum_{i=0}^{\infty} c_i^2 E_i}{\sum_{i=0}^{\infty} c_i^2},$$

The exact ground state energy is subtracted from both sides:

$$E(\Phi) - E_0 = \frac{\sum_{i=0}^{\infty} c_i^2 E_i - E_0}{\sum_{i=0}^{\infty} c_i^2},$$

Since every term on the right-hand side of the equation is greater than or equal to zero, the left-hand side must also be greater than or equal to zero, so:

$$E(\Phi) \geq E_0.$$

The energy of any approximate wavefunction is always greater than or equal to the exact ground state energy E_0 .

With this in mind the variational method works by knowing that the energy of any approximate trial function is always above the true energy, then any variations in the trial functions - which decrease its energy - are, necessarily, making the approximate energy closer to the exact answer. If in the above expressions we built the trial function not as an infinite combination of eigensolutions of the Hamiltonian, but as a finite combination of a fixed set of functions, that we now call the basis set, we call this the linear variational principle. This is the one used in molecular quantum mechanics. When applied to the Hartree-Fock theory, it is usually called the Linear Combination of Atomic Orbitals (LCAO).

2.2- The Born-Oppenheimer approximation

An essential part of solving the Schrödinger equation is the Born-Oppenheimer approximation, where the coupling between the nuclei and electronic motion is neglected.⁶

The Born-Oppenheimer approximation rests on the fact that the nuclei are much more massive than the electrons, which allows saying that the nuclei are nearly fixed with respect to the electron motion. This approximation is applied in the Hamiltonian operator.

A system is described by two types of particles: the electrons (e) and the nuclei (n). Those particles are defined in a fixed frame of reference, where we denote (r) the set of electronic coordinates and (R) the set of nuclear coordinates.

The non-relativistic Hamiltonian can be written in a shorthand expression as:

$$H = T_e(r) + T_n(R) + V_{ne}(r, R) + V_{nn}(R) + V_{ee}(r). \quad (3)$$

The Hamiltonian takes into account five contributions to the total energy of a system: the kinetic energies of the electrons (T_e) and nuclei (T_n), the attraction of the electrons to the nuclei (V_{ne}) plus the interelectronic (V_{ee}) and internuclear (V_{nn}) repulsions.

Following the Born-Oppenheimer approximation in which the nuclei are considered to be fixed in a position, the term $T_n(R)$ can be neglected and the term $V_{nn}(R)$ can be considered as a constant, so, the resulting Hamiltonian is called the electronic Hamiltonian operator - as it is solved for the electronic part -.

$$H_{el} = T_e(r) + V_{ne}(r, R) + V_{nn}(R) + V_{ee}(r). \quad (4)$$

The Schrödinger equation (1) can be written in electronic terms as:

$$H_{el}\Psi(r; R) = E_{el}\Psi(r; R). \quad (5)$$

This provides the electronic energy at each nuclear configuration, leading to the concept of potential energy surface which is basic in chemical reactivity.

2.3- Hartree-Fock theory

The dynamics of many-electron systems is very complex, and consequently requires elaborated computational methods. A significant simplification can be obtained by introducing independent-particle models, where the motion of one electron is considered to be independent of the dynamics of all the other electrons.

In HF, each electron is described by a one-electron function called orbital (spin-orbital when the spin is added) and the total wavefunction is given as a product of spin-orbitals; this being known as a Hartree Product. This form fails to satisfy the antisymmetry principle, which states that a wavefunction describing fermions should be antisymmetric with respect to the interchange of any set of space-spin coordinates.

The antisymmetry of the wavefunction can be achieved by building it as a Slater determinant (SD). In the SDs the columns label the spin-orbitals and the rows the electrons.

For a system with N electrons and spin-orbitals (ϕ) the Slater determinant is:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{bmatrix} \phi_1(1) & \phi_2(1) \cdots & \phi_N(1) \\ \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) \cdots & \phi_N(N) \end{bmatrix}. \quad (6)$$

The trial wavefunction (Φ) is taken as a single Slater determinant, implying that the electron-electron correlation is neglected.

Once the electronic Hamiltonian and the trial wavefunction are selected, the energy needs to be minimized by the variational principle and, from that, the Hartree-Fock equations will be obtained.

Coming back to the electronic Hamiltonian (4):

$$H_{el} = T_e + V_{ne} + V_{nn} + V_{ee}.$$

The terms V_{ne} and the T_e depend only on one electron coordinate. Term V_{ee} depends on two electron coordinates and so the operators can be grouped depending on the number of electron indices; then:

h_i corresponds to the one-electron operator and describes the motion of electron in the field of all nuclei.

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_a^N \text{nuclei} \frac{Z_a}{|R_a - r_i|} \quad (7)$$

g_{ij} corresponds to the two-electron operator giving the electron-electron repulsion.

$$g_{ij} = \frac{1}{|r_i - r_j|} \quad (8)$$

The electronic Hamiltonian (4) can be, then, written in a simpler way as:

$$H_{el} = \sum_i^N \text{elec} h_i + \sum_{j>i}^N \text{elec} g_{ij} + V_{nn}. \quad (9)$$

Knowing that the wavefunctions can be written as a Slater Determinant and the Hamiltonian in a simplified way, the energy of the system is obtained as follows.

For the calculation of the energy expression, the starting point is that $E_e = \langle \Phi | H | \Phi \rangle$ when the wavefunction is normalized. For the energy expression the variational theorem is applied, so it can be said that the parameters can be varied until the minimum possible energy is achieved. In simplified terms, in HF we try to find the best possible orbitals within a given Slater determinant.

Following some algebraic relationships, the electronic energy can be written in terms of integrals of the one- and two-electron operators:

$$E = \sum_{i=1}^{N_{elec}} h_i + \frac{1}{2} \sum_{i=1}^{N_{elec}} \sum_{j>i}^{N_{elec}} (J_{ij} - K_{ij}) + V_{nn}. \quad (10)$$

In this expression,

$$\begin{aligned} h_i &= \langle \phi_i(1) | h_i | \phi_i(1) \rangle, \\ J_{ij} &= \langle \phi_i(1)\phi_j(2) | g_{12} | \phi_i(1)\phi_j(2) \rangle, \\ K_{ij} &= \langle \phi_i(1)\phi_j(2) | g_{12} | \phi_j(1)\phi_i(2) \rangle. \end{aligned}$$

In the followings derivations, the energy is express in terms of Coulomb (J) and exchange (K) operators.

$$E = \sum_{i=1}^{Nelec} \langle \phi_i(1) | h_i | \phi_i(1) \rangle + \frac{1}{2} \sum_{ij}^{Nelec} (\langle \phi_j | J_i | \phi_j \rangle - \langle \phi_j | K_i | \phi_j \rangle) + V_{nn}.$$

$$J_i | \phi_j(2) \rangle = \langle \phi_i(1) | g_{12} | \phi_i(1) \rangle | \phi_j(2) \rangle,$$

$$K_i | \phi_j(2) \rangle = \langle \phi_i(1) | g_{12} | \phi_j(1) \rangle | \phi_i(2) \rangle.$$

The energy expression is symmetric and so the variational theorem holds. Now the set of spin-orbitals (or molecular orbitals, MOs) that make the energy a minimum need to be determined.

The Slater Determinant with the lowest energy will give the wavefunction for the assumed functional form of the Slater Determinant. The variation must be carried out in such a way that the MOs remain orthogonal and normalized. This is done by minimizing a Lagrange functional that uses a Lagrange multiplier for each of the orthonormality conditions:

$$L = E - \sum_i^{Nelec} \lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}),$$

$$\delta L = \delta E - \sum_i^{Nelec} \lambda_{ij} (\langle \delta \phi_i | \phi_j \rangle - \langle \phi_i | \delta \phi_j \rangle) = 0.$$

According to the variational principle, the best orbitals shall produce $\delta L = 0$.

Introducing the energy expression, the variation of the energy is given by:

$$\delta E = \sum_i^{Nelec} (\langle \delta \phi_i | h_i | \phi_i \rangle + \langle \phi_i | h_i | \delta \phi_i \rangle) +$$

$$\frac{1}{2} \sum_{ij}^{Nelec} \left(\langle \delta \phi_i | J_j - K_j | \phi_i \rangle + \langle \phi_i | J_j - K_j | \delta \phi_i \rangle + \right.$$

$$\left. \langle \delta \phi_j | J_i - K_i | \phi_j \rangle + \langle \phi_j | J_i - K_i | \delta \phi_j \rangle \right).$$

The third and fifth terms are identical as well as the fourth and sixth terms. They cancel out the $\frac{1}{2}$ factor and so it can be rewritten using the so-called Fock operator (F_i) as the variation:

$$\begin{aligned}
\delta E &= \sum_i^{Nelec} (\langle \delta\phi_i | h_i | \phi_i \rangle + \langle \phi_i | h_i | \delta\phi_i \rangle) + \\
&\sum_{ij}^{Nelec} (\langle \delta\phi_i | J_j - K_j | \phi_i \rangle + \langle \phi_i | J_j - K_j | \delta\phi_i \rangle), \\
\delta E &= \sum_i^{Nelec} (\langle \delta\phi_i | F_i | \phi_i \rangle + \langle \phi_i | F_i | \delta\phi_i \rangle), \\
F_i &= h_i + \sum_j^{Nelec} (J_j - K_j). \tag{11}
\end{aligned}$$

The Fock operator is an effective one-electron energy operator, describing the kinetic energy of an electron and the attraction to all the nuclei (h_i), as well as the repulsion to all the other electrons (via the J and K operators). The Fock operator is associated with the variation of the total energy, so it is substituted in the variation of the Lagrange function:

$$\delta L = \sum_i^{Nelec} (\langle \delta\phi_i | F_i | \phi_i \rangle + \langle \phi_i | F_i | \delta\phi_i \rangle) - \sum_i^{Nelec} \lambda_{ij} (\langle \delta\phi_i | \phi_j \rangle - \langle \phi_i | \delta\phi_j \rangle).$$

As said above, orbitals of interest are those that make $\delta L = 0$. The use of complex conjugate properties in the Lagrange function gives the final set of Hartree-Fock equations:

$$F_i \phi_i = \sum_j^{Nelec} \lambda_{ij} \phi_j, \tag{12}$$

It can be shown that one can subject the above orbitals to a rotation that makes the Lagrange multipliers diagonal. This leads to the canonical Fock equations, where $\lambda_{ii} = \varepsilon_i$ is called the orbital energy:

$$F_i \phi_i = \varepsilon_i \phi_i. \tag{13}$$

Notice that the Fock operator depends on the orbitals that we need to obtain, so that to solve the HF equations we need their solution. As usual, this problem is solved by using an initial guess for the orbitals, then solving the HF equations to get a new set of orbitals and iterating this procedure until there is no change neither in the orbitals nor in the total computed energy. This is the so-called Self-Consistent Field (SCF) procedure. To transform this technique into an efficient computational algorithm, another step needs to be taken.

2.4- Roothaan-Hall equations

The Hartree-Fock equations may be solved in a numerical way but most of the calculations use a basis set expansion to express the unknown MOs in terms of a set of known functions. Each MO

(ϕ) is expanded in terms of the basis functions (χ), also called atomic orbitals. As stated before, this linear expansion of the molecular orbitals is called the Linear Combination of Atomic Orbitals (LCAO).

$$\phi_i = \sum_{\alpha}^{M \text{ basis}} C_{\alpha i} \chi_{\alpha}. \quad (14)$$

Finding the HF orbitals is in this algebraic approximation equivalent to finding the $C_{\alpha i}$ parameters (expansion coefficients). Substituting the molecular orbitals in the Hartree-Fock equations for the expansion - in terms of the atomic orbitals - leads to:

$$F_i \sum_{\alpha}^{M \text{ basis}} C_{\alpha i} \chi_{\alpha} = \varepsilon_i \sum_{\alpha}^{M \text{ basis}} C_{\alpha i} \chi_{\alpha}. \quad (15)$$

Multiplying and integrating from the left by a basis functions yields the Roothaan-Hall equations:

$$FC = SC\varepsilon. \quad (16)$$

Where: F contains the Fock matrix elements, C is the matrix of the unknown expansion coefficients, S is the overlap matrix between basis functions and ε is the (diagonal) matrix of orbital energies.

$$F_{\alpha\beta} = \langle \chi_{\alpha} | F | \chi_{\beta} \rangle,$$

$$S_{\alpha\beta} = \langle \chi_{\alpha} | \chi_{\beta} \rangle.$$

For the determination of the MO coefficients, ($C_{\alpha i}$) the Fock matrix must be diagonalized, but the Fock matrix is only known if all the MO coefficients are known. So, the procedure starts with a guess of the coefficients, then form the F matrix and diagonalize it. The new set of coefficients is then used for calculating a new Fock matrix. There is a continuous process until the set of coefficients used for constructing the matrix equals those resulting from the diagonalization of the matrix. This set of coefficients determines a self-consistent field solution.⁷

Self-Consistent field steps:

1. Specify the system, basis function and electronic state of interest
2. Form overlap S matrix.
3. Guess initial MO $C_{\alpha i}$ coefficients.
4. Form Fock F matrix.
5. Solve $FC = SC\varepsilon$
6. Use new MO $C_{\alpha i}$ coefficients to build new Fock F matrix.
7. Repeat steps 5 and 6 until C no longer changes from an iteration to the next one.

2.5- Post Hartree-Fock methods

One of the limitations of HF calculations is that they do not include electron correlation. This means that HF takes into account the average effect of electron repulsion, but not the explicit electron-electron interaction. Within HF theory the probability of finding an electron at some location around an atom is determined by the distance from the nucleus but not the distance to the other electrons. This is not physically true, but it is the consequence of the central field approximation, which defines the HF method.

Some calculation types begin with a HF calculation and then they correct for correlation. Some of these methods are: Møller-Plesset perturbation theory (MP), the generalized valence bond (GVB) method, multiconfigurational self-consistent field (MCSCF), configuration interaction (CI) and couple cluster theory (CC) as a group. All these methods are referred to as correlated calculations. Correlation is important for many different reasons; including correlation, generally, improves the accuracy of computed energies and molecular geometries.

2.6- Basis sets

We are now in a position to understand what basis sets are used in the calculation of the electronic structure of molecules. In the derivation of the Roothaan-Hall equations the concept of basis functions was introduced, and a basis function can be defined as a linear combination of mathematical functions. The set of basis functions used for a particular calculation with a specific set of parameters are called the basis set. The type of basis function used influences the accuracy of the calculations: typically, the smaller the basis the poorer the representation. It also impact the computational effort of *ab initio methods*. If the number of basis functions is M , then the number of two-electron integrals to be computed scales - formally at least - as M^4 so it is, of course, of high importance to make the basis set as small as possible without compromising the accuracy.⁸

An ideal basis set should have a number of properties such as:

- The basis functions should reflect the nature of the problem.
- Basis sets should allow calculations for not only the energy and energy-related quantities but also for a range of other properties.
- For atom-dependent basis functions, the basis sets should be available for all atoms in the periodic table.

The requirement of computational efficiency is often incompatible with some of the properties, such as the universality of the basis functions. A basis set that is capable of a given accuracy for a variety of properties will always be larger than a tuned one for a specific purpose.

2.6.1- Types of basis functions

There are two great categories of basis functions commonly used in electronic structure calculations for isolated molecules, known as the Slater Type Orbitals (STOs) and the Gaussian Type Orbitals (GTOs). In the case of periodic solids, it is also customary to employ planewaves, but we will not discuss them.

Slater Type Orbitals appear in the solution of the SE equation for a hydrogen-like atomic system. This means that STOs should be well suited for achieving a good representation of AOs with only a few functions. An STO is formally the following three-dimensional function,

$$\chi_{\xi,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{n-1}e^{-\xi r}. \quad (17)$$

where N is a normalization constant, $Y_{l,m}$ are spherical harmonic functions, ξ is the orbital exponent – a parameter that needs to be optimized-and n is the principal quantum number. Calculation of 3- and 4- center two-electron integrals with STO functions is difficult and computationally inefficient.

The common way to overcome this limitation in electronic structure calculations is to replace the STOs by GTOs. They can be written in terms of polar or cartesian coordinates:

$$\begin{aligned} \chi_{\xi,n,l,m}(r, \theta, \varphi) &= NY_{l,m}(\theta, \varphi)r^{2n-2-l}e^{-\xi r^2}, \\ \chi_{\xi,l_x,l_y,l_z}(x, y, z) &= Nx^{l_x}y^{l_y}z^{l_z}e^{-\xi r^2}. \end{aligned} \quad (18)$$

The sum of l_x , l_y and l_z determines the type of orbital (0 for s functions, 1 for p , 2 for d , and so on) and the orbital exponent ξ defines the radial extent of the function.

Most of the *ab initio* calculations use GTOs as basis functions. In particular the cartesian GTOs, since the molecular integral formulas over them are straightforward to derive and to code.

The r^2 dependence in the exponent makes the GTOs inferior to the STOs, since it is analytically known that the behavior of the exact wavefunction both in the vicinities of the nuclei (Kato's cusp theorem⁹) and in the large distance asymptotic regime¹⁰ is exponential. The usual way of improving the accuracy is to use linear combinations of a few GTOs to describe a single STO. Although the number of basis functions (also the integrals) are increased, the ease by which these integrals can be calculated makes GTOs the preferred basis functions in practical electronic structure calculations.

2.6.2- Classification of Gaussian Type Basis Sets

A given basis set is uniquely defined by the number of functions of each type (*s*-,*p*-,*d*-, etc), the exponent value for each function and how the functions are combined into fixed linear combinations. We call each of these functions a primitive. The combination of these three sets of parameters provides a lot of degrees of freedom, which is reflected by the large number of referenced proposed basis sets.¹¹

Combining the full set of basis functions, known as the primitive GTOs (PGTOs), into a smaller set of functions by forming fixed linear combinations is known as basis set contraction, and the resulting functions are called contracted GTOs (CGTOs).

A first classification according to the number of contracted functions used for describing the occupied AOs can be made. If we use the smallest number of possible functions, we call the basis set minimal. Only enough AOs functions are employed to contain all the electrons of the neutral atom, i.e. one per electron. They are denoted as single zeta (SZ) basis sets. For the Hydrogen and Helium atoms this means a single *s*-functions (1*s*), which can be of up or down spin in the second case; instead for the second row it means two *s*-functions (1*s* and 2*s*) and one set of *p*-functions (2*p_x*, 2*p_y*, 2*p_z*), and so on.

If the number of basis functions is double the number of electrons, then it is called of the double zeta type (DZ). A DZ basis employs two *s*-functions (1*s* and 1*s'*) for Hydrogen and Helium and four *s*-functions (1*s*, 1*s'*, 2*s* and 2*s'*) and two sets of *p*-functions (2*p* and 2*p'*) for second row elements. Doubling the number of basis functions allows for a much better description of the electron distribution.

The next one in the hierarchy basis set size is a triple zeta basis (TZ). It contains three times as many primitives as the minimum basis. The increase in the number of basis functions can continue into quadruple zeta (QZ) and quintuple zeta and beyond (5Z, 6Z, etc). Increasing the number of functions allows the basis set to describe the bonding in all the directions and environments. It is important to notice, however, that basis sets are usually optimized in free atoms, and that the molecular environment tends to be rather different than the atomic one. For instance, spherical symmetry is no longer present, so deformations of the electron density are expected. Similarly, intra- or intermolecular charge transfer effects may increase or decrease the number of electrons locally, making it necessary to increase the number of basis functions to accommodate for them.

Improvement of the computational efficiency of basis sets can thus be guided by chemical knowledge of the expected properties of the solutions. There are three simple possibilities that are used in this regard: core electrons, for instance, tend to stay frozen when a molecule is formed, so the number of basis functions used to represent them may be smaller than in the case of the valence. Similarly, distortion from sphericity may advice the use of extra functions with other

angular dependence. Finally, if we deal with anions, or with electron rich regions due to charge transfer, extra functions to take this larger number of electrons will be needed. This leads to the concepts of split valence basis, polarization, and diffuse basis functions.

Since the core electrons of a molecule are less affected by the chemical environment than the valence electrons, they are sometimes treated with a minimal basis set while the valence electrons are treated with a larger basis set such as a DZ or TZ. This is what is called the Split Valence Basis Set (SV), and it is useful to reduce the time required for large molecules calculation.

Other way of increasing the size of the basis set is by adding polarization functions. A polarization function is any higher angular momentum orbital used in a basis set that is not normally occupied in the separated atom. The use of polarization basis functions allows for the atomic electron densities to be polarized in order to better represent the molecule's electron density. Polarization functions are needed to obtain a better representation of the electron density in the bonding regions; an example is that the Hydrogen atom only contains a *s*-orbital (spherical) but when the Hydrogen atom is bonded to a more electronegative atom such as Lithium, the electron density is no longer spherical. The addition of a *p*-orbital function will allow the electron density to be shifted (or polarized) in the direction of the bond. Polarization functions are usually *p*-orbitals, *d*-orbitals or *f*-orbitals since the *p*-orbitals introduces a polarization of the *s*-orbitals and the *d*-orbitals can be used for polarizing *p*-orbitals and so on. Adding a single set of polarization functions to the DZ basis forms a Double Zeta plus Polarizations type basis (DZP). There is a variation where polarization functions are only added to non-hydrogen atoms, considering that hydrogens often have a "passive" role. As Hydrogen often account for a large number of atoms in the system, a saving of three basis functions per each Hydrogen is significant. Diffuse basis functions are extra basis functions that are added to the basis set to represent very broad electron distribution. They are very useful for representing the electron density in ions or in intermolecular complexes.

2.6.3- Contracted Basis Functions

Contraction denotes the practice of forming predefined linear combinations of primitive basis functions which are held fixed, thus decreasing the number of linear parameters to be computed. It improves computational efficiency, but, at the same time, it carries a degradation of the accuracy that can be achieved. The main effort in contracting basis sets is to achieve the largest computational saving without affecting the accuracy of the primitive basis set.

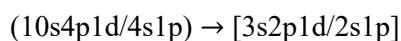
In a contraction we take a set of *k* primitive Gaussian basis functions (PGTOs) to form a new fixed contracted (the CGTO).

$$\chi(\text{CGTO}) = \sum_i^k \alpha_i \chi_i(\text{PGTO}). \quad (19)$$

Here α_i are the contraction coefficients, which are optimized once and for all as the orbital exponents of the functions.

Contracting a basis set will always increase the energy, since it is a restriction of the number of variational parameters, and makes the basis set less flexible, but will significantly reduce the computational cost.

Specification of a basis set in terms of primitive and contracted functions is given by the notation:



The basis in parentheses is the number of primitives for the heavy atoms (second row elements) before the slash and for hydrogen atoms after the slash, all before contraction. After contraction takes place, we indicate the final number of primitives in square brackets. This notation provides no information about how the contraction is done.

There are two different schemes for basis set contractions: segmented and general contraction. In a segmented contraction, each primitive function only contributes to one contracted function and both the exponent and the contraction coefficient are simultaneously optimized by a variational optimization. In a general contraction all the primitive functions are allowed to contribute to all contracted functions but with different contraction coefficients. Most basis sets in use contain elements of both; segmented and general contraction schemes at the same time. The primitive functions with large exponents describe, primarily, the inner-shell orbitals while the smaller exponent primitive functions mainly contribute to the description of the outer-valence orbitals.

2.6.4- Optimization of basis set parameters

The exponents of the primitive functions and the contraction coefficients are basis set parameters that must be assigned values by a suitable criterion. The most common method for that purpose is to employ the variational principle and to minimize the energy as a function of these variables.¹²

The basis sets parameters describing the occupied AOs can be optimized at the atomic level, while polarization functions must be optimized for molecular systems. The diffuse functions for the occupied orbitals can be variationally optimized by minimizing the HF energy of selected anions, while keeping fixed the exponents of the underlying basis set.

The common procedure of optimizing basis set parameters by minimizing the total energy has the consequence that only the energetically important region in function space for the target systems is adequately covered.¹³

The optimization of basis sets is something not to really worry about, as most of the optimized basis sets of different qualities and sizes are available in tables, websites or built in within the available computer programs.

2.6.5- Complete Basis Sets

Solution of the HF equations with an infinite complete basis set would provide the exact Hartree-Fock energy. This is defined as the HF limit. Carrying out such calculations is of course impossible, but one may try to check how the energy varies on increasing the basis set size and extrapolate from this the HF limit with a fair degree of confidence. This extrapolation is called a Complete Basis Set (CBS) limit extrapolation. This procedure tries to remove any error due to the linear combination of atomic orbitals approximations. The extrapolation is based on a minimum of three separate calculations with increasingly larger basis sets. There are some basis sets than have been expressly designed for that purpose (cc-pVnZ for example), as they increase in size in a consistent fashion with each increment on n . One can imagine plotting some particular computed property as a function on n^{-1} and extrapolating for a curve through those points, then back to intercept the one corresponding to $n = \infty$.

2.6.6- Property Basis Sets

Design of basis sets for molecular properties is not as straightforward as for the energy; there is no equivalent to the variational principle for properties. A common procedure is to manually identify deficiencies in a standard (energy-optimized) basis set and then adding necessary functions for performance improving.

3 COMPUTATIONAL METHODOLOGY

After presenting the meaning and classification of the basis sets commonly used in electronic structure calculations, a particular choice of the computational environment with which we will examine some examples needs to be made. There is a significant number of electronic structure computer programs available in the literature. Some of them are completely open software, while others are free only for academic purposes. Finally, some of them lie behind a company paywall. Almost every computational chemist is familiar with a small number of general purpose big packages: Gaussian, Gamess, Dalton, Orca, NWChem, etc⁷. These software-packages are self-sufficient and include many built-in quantum chemical methods. Being coded for research purposes, they usually stand within the input-code-output paradigm. The codes read an input file (appendix 7.1) in which the method, system and basis functions, that the user wishes to use, are specified and the program generates an output with the desired calculations and properties.

Usually both the input and the output files are generated and analyzed with additional codes that ease this task. It is at this step that graphical interfaces may appear.

Most of the programs share the same global strategy, and a calculation is divided into a set of independent tasks:

1. Preparation of the calculation:
 - a. Steps that the computation should follow
 - b. Specification of the system's geometry
 - c. Specification of the system's characteristics (charge, spin state)
 - d. Choice of the basis set to be employed
 - e. Choice of the level of theory that will be used: HF, DFT, etc.
2. Calculation of the basic integrals.
3. Preparation of an initial guess
4. SCF calculation.
5. Rest of calculations
 - a. Post-SCF calculations
 - b. Search for stationary points in the geometric configuration
6. Properties calculations

As stated, we have used the ORCA package for this Thesis. It is a free software available for all the common operating systems. ORCA describes itself as an *ab initio*¹, *DFT and semi-empirical electronic structure package*. All the basic theory levels are implemented in ORCA, ranging from Hartree-Fock to DFT and multi-reference models.

3.1- Level of theory

The level of theory chosen for this Thesis is the Hartree-Fock theory. As already explained, the Hartree-Fock method is an *ab initio method* within the Born-Oppenheimer approximation that neglects relativistic effects. It is the cheapest of all the possibilities available in ORCA, but it serves well the purpose of studying the role of basis sets. The method is selected in the input by specifying the keyword HF.¹⁴

3.2- Choosing a suitable Basis Set

The choice of a basis set is nothing but a compromise between accuracy and computational efficiency, and both factors apply to the wavefunctions selection as well. As a general rule, the more sophisticated a method is the larger the basis set it requires. Since the Hartree-Fock method is not a high-performance method, it is usually used with low quality basis such as DZP, unless the HF results are used as a component in more elaborated extrapolation schemes.

Basis sets are available from many different sources and often provided as an integral part of electronic structure programs so they can be employed by just specifying an input suitable keyword.

The different selected basis sets for this work come from three of the main research groups that have worked on basis set selection over the years: the Pople, the Dunning and the Alrichs groups, and these names label them. These groups of basis sets are some of the most used for calculations and they are highly studied and optimized. The main difference between them is the optimization method of the exponents and contraction coefficients. From each group at least three basis sets were selected for this work. The nomenclature of each of them is described below:

1. Basis sets from Pople's group:

The basis sets published by the group of John Pople (who was a Nobel prize winner in 1998) are amongst the most popular basis sets in *ab initio* molecular calculations. They offer low to high quality basis sets. The most used basis functions are the split-valence basis sets in which minimal basis functions are used for core orbitals, but valence orbitals are split into two or three parts. The parameters were optimized by variational procedures with the restriction the same exponents are used for both the *s*- and *p*-functions in the valence, decreasing the flexibility of the basis set but increasing the computational efficiency. The nomenclature is *K-nlmG* in which *K* corresponds to the number of PGTOs used for describing the core orbitals and *nlm* indicate how many functions the valence orbitals are split into as well as the number of PGTOs used to describe them. If there are only two values (*nl*) it corresponds to a double-zeta and if there are three values (*nlm*), it corresponds to a triple-zeta. Some of the most common Pople group basis sets used are the 3-21G, 6-31G and 6-311G.

In order to improve the basis sets with polarization functions or diffuse functions the nomenclature used is * and + respectively. The polarization functions are indicated with * that means that one set of d-type polarization functions is added to each non-hydrogen atom in the molecule and ** means that one set of *d*-type polarization functions is added to non-hydrogen and one set of *p*-type polarization functions is added to Hydrogen. The diffuse functions are indicated with + that means that one set of *sp*-type diffuse basis functions is added to non-hydrogen atoms. The use of a ++ label means that one set of *sp*-type diffuse functions is added to each non-hydrogen atom and one *s*-type diffuse function is added to Hydrogen atoms.

2. Correlation consistent basis sets (Dunning group):

The *ab initio* total energy of an HF wavefunction is in error with respect to the true nonrelativistic energy by an amount called "the correlation energy". The correlation consistent (cc) basis sets come from an upgrade from the Dunning basis sets (DZ,TZ...). The cc basis sets were made to recover the correlation energy of the valence electrons.

The cc means that the basis sets are designed such, those functions that contribute similar amounts of correlation energy, are included at the same level - independent of function type-. They provide systematic improvements that converge toward the complete basis set limit. The nomenclature of the basis sets is cc-pVNZ in where the cc corresponds to correlation consistent, the p to polarized, the V indicates that they are valence-only basis sets and N indicates the level of multiplicity (D=double, T=triple...)

The smallest member of this series and thus often the starting point for correlated calculations is the cc polarized double zeta (cc-pVDZ) set. There are also the cc-pVTZ, cc-pVQZ, etc, sets.

3. Alhrichs group Basis Sets:

These basis sets were developed as second generation basis sets and are labelled as the def2 basis sets, in particular def2-SVP, def2-TZVP, def2-TZVPP, def2-QZVP, where SV denotes split valence, TZV denotes valence triple zeta and so on. Parameters optimization was done by means of gradient techniques. The def2 basis sets are designed to provide consistent accuracy across the whole periodic table whereas this is not true for the more commonly used basis sets from Pople and Dunning groups.

The basis sets from the Pople and Ahlrichs groups were designed following a segmented contraction scheme; instead the Dunning group basis sets were designed following a general contraction scheme.

The basis sets chosen from the three groups for this work are those who do not have a high computer-processing time while, at the same time, they provide a good representation of each category. In total thirteen (13) different basis sets were chosen, and the table below contains a brief description of each of them. Characteristics such as whether the basis sets are double, triple or quadrupole-zeta, if polarized or diffuse functions are added, and the contraction followed depending on whether the elements are from the s- or the p- block, can be found in Table 1 below.

Table 1. Description of the basic features of the basis sets used in this work. The quality, origin, standard name, the use of polarization or diffuse functions, and the contraction scheme are specified for second row *s*- and *p*-block elements (Li-Ne)

Quality	Group	Name	Polarization or diffuse functions	Composition	
				<i>s</i> -Block elements	<i>p</i> -Block elements
SV	Alhrichs	Def2-SVP	Polarization	$(7s4p) \rightarrow [3s2p]$	$(7s4p1d) \rightarrow [3s2p1d]$
DZ	Pople	3-21G	None	$(6s3p) \rightarrow [3s2p]$	$(6s3p) \rightarrow [2s2p]$
DZ	Pople	6-31G	None	$(10s4p) \rightarrow [3s2p]$	$(10s4p) \rightarrow [3s2p]$
DZ	Pople	6-31G*	Polarization	$(10s4p1d) \rightarrow [3s2p1d]$	$(10s4p1d) \rightarrow [3s2p1d]$
DZ	Pople	6-31G+	Diffuse	$(10s4p1d) \rightarrow [3s2p1d]$	$(10s4p1d) \rightarrow [3s2p1d]$
DZ	Dunning	cc-pVDZ	Polarization	$(9s4p1d) \rightarrow [3s2p1d]$	$(9s4p1d) \rightarrow [3s2p1d]$
TZ	Pople	6-311G	None	$(11s5p2d) \rightarrow [4s3p2d]$	$(11s5p2d) \rightarrow [4s3p2d]$
TZ	Pople	6-311G*	Polarization	$(11s5p2d1f) \rightarrow [4s3p2d1f]$	$(11s5p2d1f) \rightarrow [4s3p2d1f]$
TZ	Pople	6-311G+	Diffuse	$(11s5p2d1f) \rightarrow [4s3p2d1f]$	$(11s5p2d1f) \rightarrow [4s3p2d1f]$
TZ	Dunning	cc-pVTZ	Polarization	$(11s5p2d1f) \rightarrow [4s3p2d1f]$	$(10s5p2d1f) \rightarrow [4s3p2d1f]$
TZ	Alhrichs	Def2-TZVP	Polarization	$(11s4p1d) \rightarrow [5s3p1d]$	$(11s6p2d1f) \rightarrow [5s3p2d1f]$
QZ	Dunning	cc-pVQZ	Polarization	$(12s6p3d2f1g) \rightarrow$ $[5s4p3d2f1g]$	$(12s6p3d2f1g) \rightarrow$ $[5s4p3d2f1g]$
QZ	Alhrichs	Def2-QZVP	Polarization	$(15s7p2d1f) \rightarrow [7s4p2d1f]$	$(15s8p3d2f1g) \rightarrow$ $[7s4p3d2f1g]$

3.3- Systems studied

The systems studied in this work are diatomic molecules of the second period of the periodic table. In computational chemistry the second period of the periodic table is also known as the first row, so some confusion may arise. In total 10 systems have been used. they are homonuclear diatomic molecules (Li₂, Be₂, B₂, C₂, N₂, O₂ and F₂) and heteronuclear diatomic molecules (LiF, BF and NF). Not all the fluorides corresponding to the second period were selected, only the ones with an even number of electrons. With this, we avoid the problem of constructing appropriate single determinant functions for non-singlet states. This is a severe problem that is usually solved by allowing up and down spin electrons to be described by different functions. This approach is known as *unrestricted* but gives rise to wave functions which are not spin eigenfunctions. Several of their properties are collected in Table 2: the experimental electronic ground state, the bond order, the number of electrons the system contains, the harmonic vibrational frequency, the equilibrium internuclear distance and the point group. Some of these values are used for comparison in the results and discussion Section.

Table 2. Some properties of the molecules studied in this work. Data taken from the NIST Webbook¹⁵.

Systems	Electronic Ground state	Bond order	Number of electrons	Vibrational frequency (cm^{-1})	Internuclear distance (r_e) (Å)	Point group
Li ₂	$1\sum_g^+$	1	6	351.43	2.673	$D_{\infty h}$
Be ₂	$1\sum_g^+$	1	8	-	2.460	$D_{\infty h}$
B ₂	$3\sum_g^-$	2	10	1051.3	1.590	$D_{\infty h}$
C ₂	$1\sum_g^+$	3	12	1854.71	1.243	$D_{\infty h}$
N ₂	$1\sum_g^+$	3	14	2358.57	1.098	$D_{\infty h}$
O ₂	$3\sum_g^-$	2	16	1580.19	1.208	$D_{\infty h}$
F ₂	$1\sum_g^+$	1	18	916.64	1.412	$D_{\infty h}$
LiF	$1\sum^+$	1	12	910.34	1.564	$C_{\infty v}$
BF	$1\sum^+$	1	14	1402.1	1.304	$C_{\infty v}$
NF	$3\sum^-$	1	16	1141.37	1.317	$C_{\infty v}$

Several systems are isoelectronic such as N₂ and BF, O₂ and NF and C₂ and LiF. The couple O₂ and NF are systems that may give very similar results due to the closeness of the elements N, O and F in the periodic table, so their properties are expected to be similar. Instead, the couples conformed by N₂ and BF and C₂ and LiF are well separated so their properties will differ more. The Be₂ system is quite special; Although Be has a $1s^2 2s^2$ electronic configuration which indicates that there should not exist a covalent bond in the diatomic, it is well known that it suffers from an important multiconfigurational character. This means that the Be₂ system won't follow the trends of the other systems.

4 RESULTS AND DISCUSSION

The electronic computer programs offer a wide range of properties that can be calculated, going from the total energy of the system to the vibration or rotation frequencies. As most of the basis sets have variational parameters optimized for the total energy, the main property, in which the work is focused, is the systems total energy. Other properties such as the internuclear distance or the harmonic vibrational force constant have also been computed and are commented later in the discussion to compare the accuracy of the basis set with respect to these properties.

All values obtained from the calculations can be found in Appendix 7.2. Tables are numbered and so a cross-relation can be followed. We use atomic units, so energies are given in Hartrees (E_h), that has an equivalence of $1 E_h = 27.2113824$ eV.

We have used two different approaches to examine the performance of basis sets:

- I. Changing the basis set at a fixed geometry of the system.
- II. Changing the basis set with an optimization of the geometry of the system.

The initial selected geometries for both approaches were obtained from the CCCBDB (Computational chemistry comparison and Benchmark database)¹⁶ and they are experimental data; these geometries are found in Appendix 7.3 table 9. From Table 10 to Table 19 the optimized geometries generated by ORCA are collected for all the systems.

Calculations are done following these two approaches in order to see how the basis sets affects the geometry. In the case of the optimization of the geometry it will give a better result as the objective of geometry optimization is to find an atomic arrangement which makes the molecule most stable and molecules are most stable when their energy level is optimized.

There is not a criterion to follow to indicate which basis sets are “the best or the worst”, but there are some theoretical concepts that provide an indication of which basis sets must give a higher accurate result than others. For example, the number of primitives functions used to describe an orbital can vary from minimal, double, triple and so on, and it is known that as size increases, results should improve.

Tables 3 and 4 in Appendix 7.2 provide values obtained for the total energy of the systems at fixed and optimized geometry. After some detailed analyses, several features arise:

- The worst basis set amongst the ones examined is the 3-21G for all the systems
- The best basis set for most of the systems is def2-QZVP but for Li_2 the best basis set is cc-pVQZ and for Be_2 the best basis set is 6-311G+.
- Inside each different group of basis sets the poorest values are obtained from the smallest basis set.
- In the basis sets of the Pople group the basis functions of the same size give better results with the addition of diffuse and polarize functions.
- The optimization of the geometry looks for the most stable arrangement and in the case of diatomic molecules, the difference with respect to the fixed geometry is of about 0.01

E_h . It is expected that if the molecules were bigger -in number of atoms- the optimization of the geometry will improve considerable the results of the total energy

In order to obtain a visual representation of the results, the energy values for the Li_2 system with optimized geometry are represented with respect to the corresponding basis set in Figure 1:

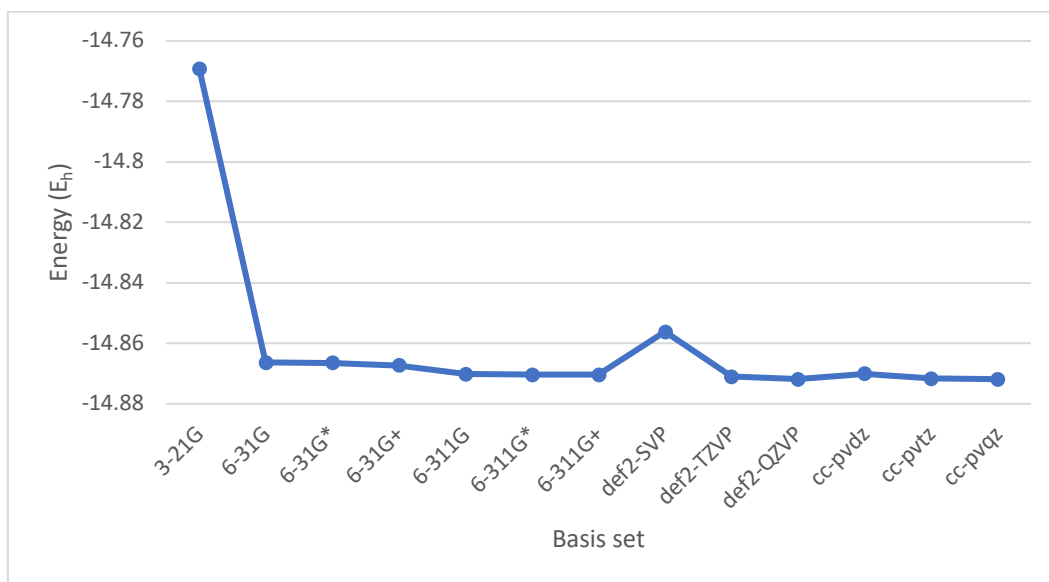


Figure 1. Evolution of the energy in Li_2 with the different basis sets explored. The geometry has been optimized in all the basis sets.

Figure 1 represents the Li_2 system. It is of small size (it only has 6 electrons) and its total energy is about $-15 E_h$. This value is relatively easy to rationalize if we take into account the hydrogenoid expression $E_n = -\frac{Z^2}{2n^2}$, which provides a value of $-4.5 E_h$ for a Li-like $1s$ function and of about -1.1 for a Li-like $2s$ electron. Two basis sets stand out from the rest: the 3-21G set and the def2-SVP set. The first one offers a much higher energy value than the rest concluding that “it is not a good-enough” basis set. For the def2-SVP set, it presents a bump a little higher in energy than the tendency around it. The def2-SVP set is a split-valence polarized one, in which the core and the valence are described differently. The basis sets after the def2-SVP are from the same group and it can be seen that with an increase of their size the energy value improves a lot.

In Figure 1 the difference between functions of the same size (DZ, TZ...) is not well visible, for that reason in Figure 2 the basis sets 3-21G and def2-SVP are removed:

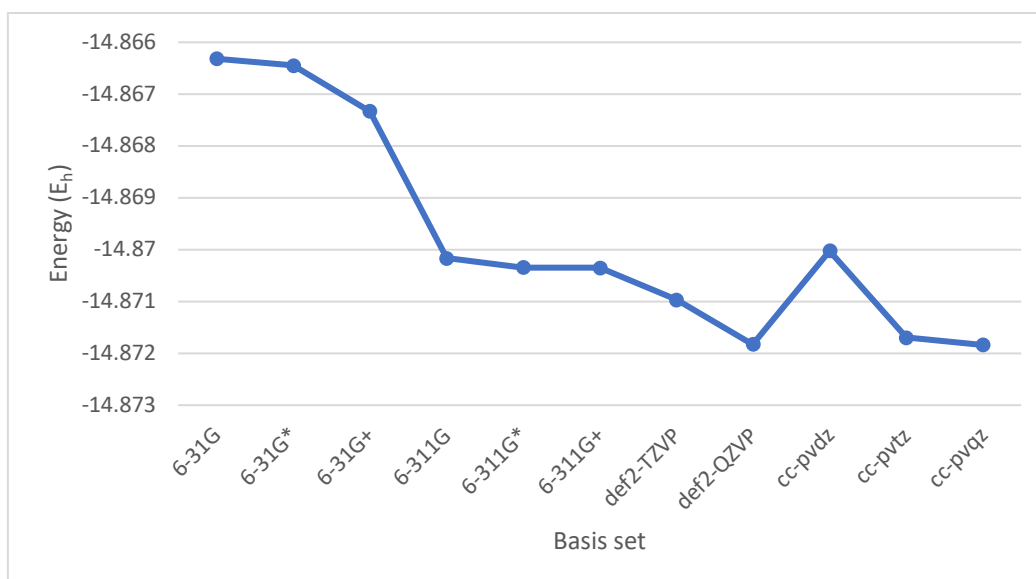


Figure 2. Evolution of the energy in Li_2 with the different basis sets explored. The geometry has been optimized in all the basis sets. The 3-21G and the def2-SVP sets have been deleted.

Figure 2 shows the difference in energy between the other sets, which is better visualized.

The range in energy values between the best and the worse basis set is $6 mE_h$; that is 3.8 kcal/mol. This is considered a high difference as the so-called chemical accuracy is prescribed at the 1 kcal/mol level.

It can also be noticed that basis sets of the same size - but different groups - offer divergent values. In the three triple-Zeta basis sets that are represented (6-311G, def2-TZVP and cc-pVTZ) there is an energy difference of $2 mE_h$ (2.26 kcal/mol). This difference in between functions with similar size is due to the method followed from each group for the optimization of the variational parameters. The importance in the optimization of the parameters can also be seen by comparing the cc-pVDZ function and the 6-311G function, the first one is a DZ and the second one a TZ, but their energy values are pretty similar.

Most of the basis sets used in the work do not consider explicitly the ionic character of the systems and so they do offer bad results for them. In Pople's basis sets we can specifically tune for that by adding diffuse or polarization functions *ad hoc* in order to obtain a better representation of the orbitals. For ionic systems the addition of diffuse functions helps to improve the results. One of the systems of study is Be_2 , it offers better results with the addition of diffuse functions. This is due to the fact that Be_2 can be said to be a Van der Waals system. These systems are described by their intermolecular interactions, so the addition of diffuse functions helps to construct better the chemical environment. The behaviour of the Pople group sets in this system is presented in Figure 3, below, with the energy values following the two approaches:

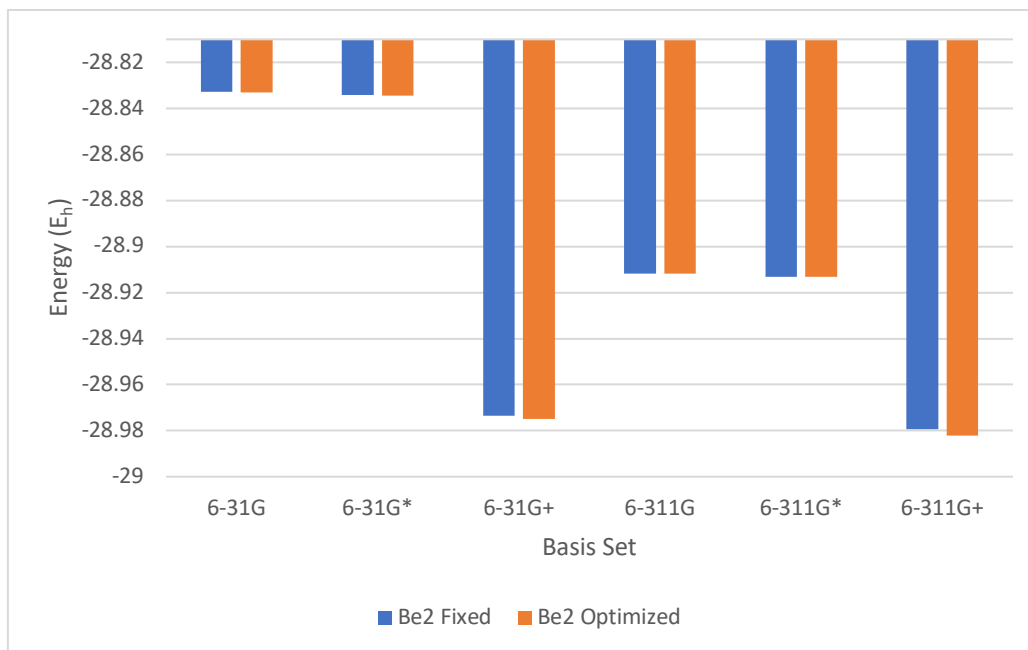


Figure 3. Energies for the Be_2 molecule at fixed geometry (blue) and at basis set optimized geometry (orange) for the basis sets of the Pople Group. The optimized geometries for each basis sets are collected in appendix 7.3.

In Figure 3 the basis sets represented are the 6-31G (DZ) and the 6-311G (TZ) together with those coming from adding polarization (*) and diffuse (+) functions. It can be observed how the addition of diffuse functions improves the energy considerably. It is the case that the diffuse DZ value is better than the non-diffuse TZ one. The values obtained with 6-31G+ and 6-311G+ are some of the best values obtained for all the basis sets studied as it can be seen in Tables 3 and 4. Improvements thanks to the addition of diffuse functions prove how important is to use basis sets that can describe the appropriate chemical environment of a system, and warns about a blind use of computational prescriptions.

Another way of representing the outcome of our calculations is by showing the difference in energy when varying the basis set with respect to a given reference. This allows to compare systems with widely varying total energies on the same footing. We have chosen as a reference basis set the 3-21G one, as it is the poorest one for all systems. The values of this energy difference are collected in Tables 5 and 6 of the Appendix 7.2. The data is better visualized in Fig. 4, that contains the values for the optimized geometry. Notice that is plot $E(\text{basis}) - E(6-31G)$, so that better sets display larger positive differences.

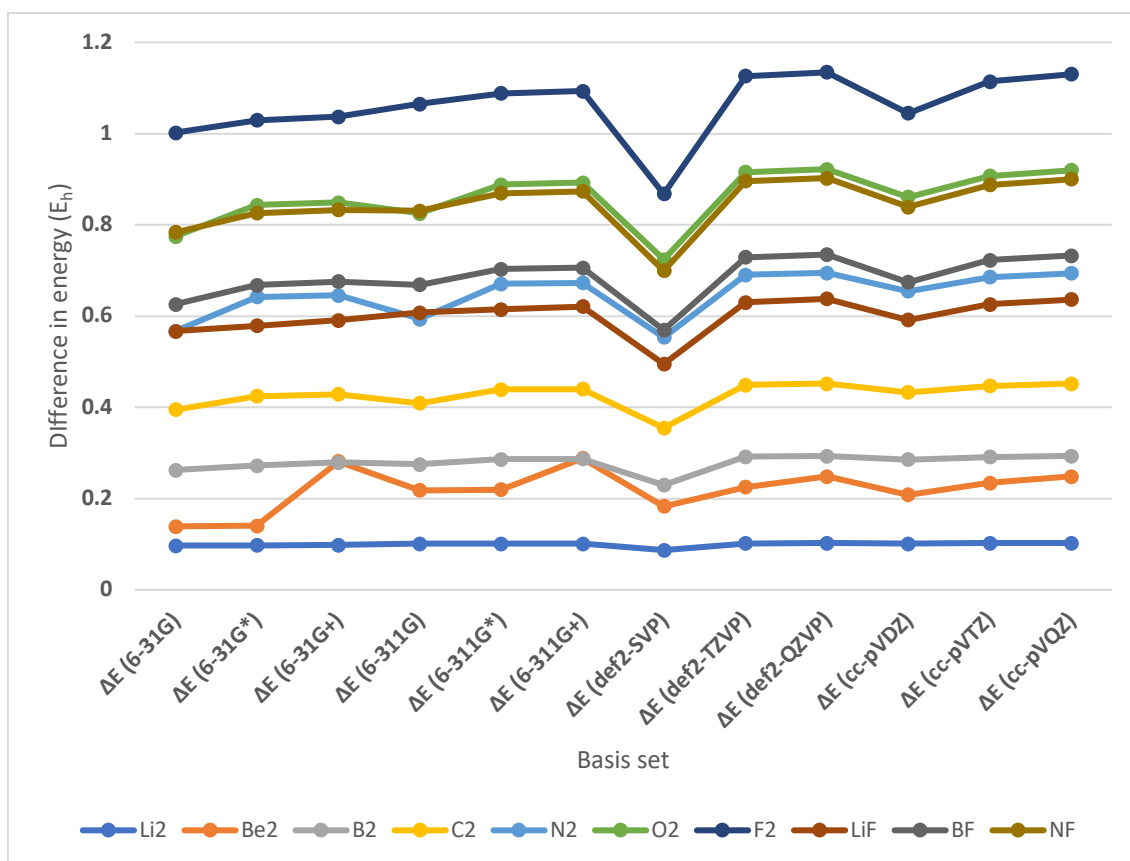


Figure 4. Energy difference with respect to the 3-21G basis set for all the systems at their optimized geometry that can be found in appendix 7.3.

The main considerations that can be extracted from Figure 4 are:

- Except for the Be_2 and the N_2 systems, the rest of systems follow a tendency marked by the depression at the def2-SVP set.
- The isoelectronic pairs of systems (same number of electrons) do offer very similar results. In the case of O_2 and NF (green and light brown lines, respectively) they are almost overlapping; this indicating that both systems have very similar properties. The pair formed by N_2 and BF (light blue and dark grey lines, respectively) does not show a high overlap - as the others – although a clear similarity is evident. Even though they are isoelectronic, the difference of the elements in the position of the periodic table is higher. Finally, the couple C_2 and LiF (yellow and dark brown lines, respectively) does not come close at all. Charge transfer is so large in this case for LiF that it destroys any possible similarity.
- Considering the size of the basis sets; as they increase the energy difference increases as expected. Trend is $\text{DZ} < \text{TZ} < \text{QZ}$.
- It can be again noticed how necessary diffuse functions are for diberyllium.

As said before, most basis sets, including the ones used for these calculations, are only optimized for the energy and so their performance as other properties are examined may not follow the same trend. To examine this possibility, we have computed the harmonic vibrational frequency at the potential energy curve minimum as well as the internuclear distance for our systems. Notice that both quantities have only a definite meaning at minima Table 2 collects together with the corresponding values obtained from the NIST (National Institute of Standards and Technology) database¹⁵ As the method followed is Hartree-Fock, it is expected for the property values to diverge from the reference ones, as HF works differently with each system. Tables 7 and 8 in Appendix 7.2 show the value for each property with optimized geometry: the internuclear distance (r_e) in Angstrom and the vibrational frequency ν_e in cm^{-1} .

The observed tendency in these two properties and probability if others were studied, is that, as the size of the basis set increases the results tend to stabilize to a certain value. Given the limited size of the basis sets, it is unclear how close these results are to the HF limit, which is the limiting asymptote.

The behavior of the internuclear distance does not seem to follow a well-defined trend for any system. This is well-known in the case of HF, which over- or underestimates bond distances depending on the type of chemical bond existing between the atoms examined. In general, the higher basis sets in size (TZ and QZ) do stabilize in a value for the internuclear distance. As an example, Figure 5 shows the values for the C_2 molecule:

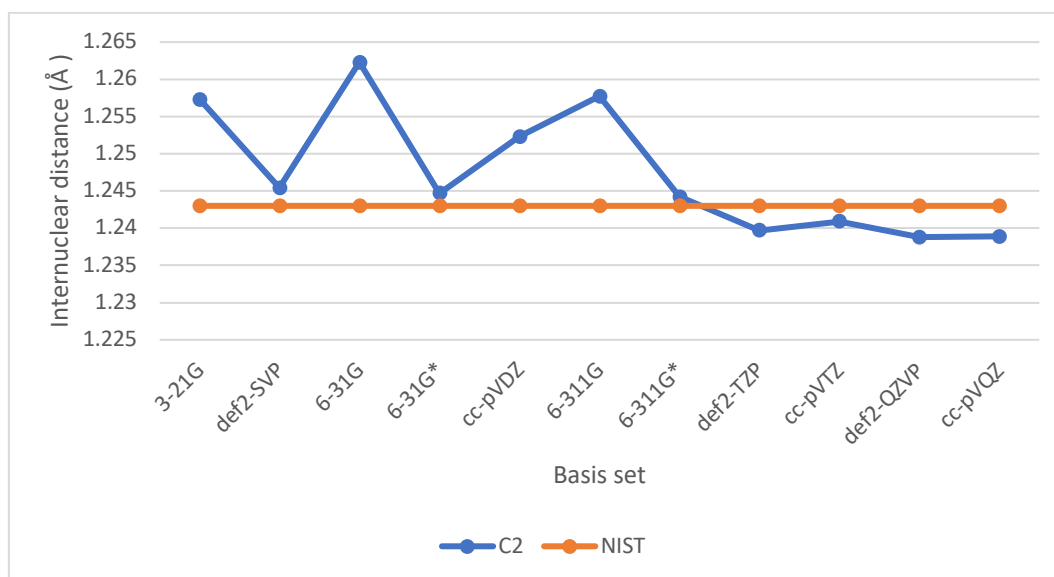


Figure 5. Evolution of the optimized internuclear distance with the basis set for the C_2 system. The NIST reference (orange) is also given. The basis sets are ordered in increasing size.

The basis sets in Figure 5 are ordered in a different way to the one used until now. Now we organize them according to their size, this allowing us to observe how the values stabilize as

we approach completeness. In the case of C_2 , values are pretty close to the reference one with the highest difference being of 0.02 Å. This is a typical accuracy for HF data.

As well as the internuclear distance values, the ones obtained for the vibrational constant are diverse and vary with respect to the reference value. In Figure 6, below, a representation of the C_2 systems vibrational constant appears.

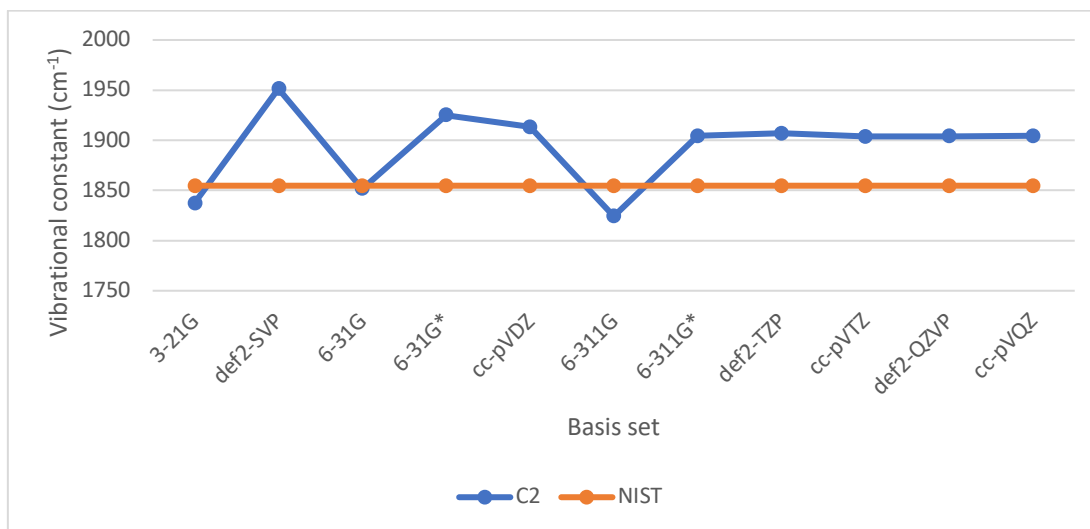


Figure 6. Evolution of the computed harmonic vibrational frequency ν_e with the basis set in C_2 . The reference NIST distance is also plotted in orange. The size of the basis set increases to the right.

The evolution of the vibrational frequency is found in Fig. 6, and it shows that apart from some low-quality Pople's sets, the frequency in this system stabilizes relatively soon. The sets 3-21G, 6-31G and 6-311G are closer to the reference and the other basis sets give values above the reference one. This behavior follows that of the internuclear distance and shows that sometimes low-quality sets provide better answers than high-quality ones. This is completely spurious. An interesting point is also obvious. Since the asymptotic HF distance in dicarbon is shorter than the experimental one, the asymptotic frequency is larger. This is also quite general. Harmonic force constants increase with decreasing distances: shorter bonds tend to be tighter.

5 CONCLUSIONS

Performance comparisons of different basis sets are an important ingredient of contemporary computational chemistry. Typically, the choice is driven by a wish for high data accuracy and confined by how costly the calculation process will become. In this Thesis a few basis set performance studies have been undertaken using the Pople, Dunning and Alhrichs basis sets for a limited set of diatomic molecules. We have used for that the ORCA software package.

The basis sets cannot be compared based on just one property as they vary in many factors such as: the procedure used for the optimization of the variational parameters, the method used for the calculations and the systems to be analyzed. In this Thesis the level of theory employed was the Hartree-Fock method, that is based on the variational principle. Relativistic effects have been neglected. With the data collected in this thesis the conclusions obtained are collected below:

- Hartree-Fock is the simplest *ab initio method* but allows to perform a preliminary study of different basis sets and helps for future calculations.
- Two approaches were followed in the calculations: fixed geometry and optimized geometry. As expected, the better results were obtained from the optimized geometry. Knowing that for the systems studied, diatomic molecules, the optimization of the geometry allows an improvement in the total energy of the system; it can be said that the optimization of the geometry is recommended to use in all calculations and specially for bigger molecules.
- The addition of diffuse functions is convenient when calculating ionic systems or systems with non-covalent bonds.
- The Pople basis sets do offer low quality results in general even when using large TZ sets.
- The Dunning group sets gives really good results, with a clear improvement with the size of the functions and for all type of systems.
- The Alhrichs group contains a poor basis sets, the def2-SVP, that gives values far from the expected ones for a basis set of that size, but, as the size of the basis sets is increased, results improve.
- Considering that most of the basis sets are optimized for the total energy, we have observed that other properties tend to stabilize to their asymptotic limit only when large sets, for instance, of quadruple-zeta quality, are used.

6 BIBLIOGRAFY

- (1) Neese, F.; Ripplinger, C.; Auer, A. A.; Anh, T. Van; Datta, D.; Souza, B. De; Izsak, R.; Lang, L.; Lu, Q.; Gastel, V.; Sivalingam, K.; Stoychev, G.; Wennmohs, F. Orca Manual.
- (2) Avogadro - Free cross-platform molecular editor - Avogadro <https://avogadro.cc/> (accessed Jun 15, 2020).
- (3) Weinbaum, S. The Normal State of the Hydrogen Molecule. *J. Chem. Phys.* **1933**, *1* (8), 593–596. <https://doi.org/10.1063/1.1749333>.
- (4) Dirac, P. A. . *The Principles of Quantum Mechanics*; 1958. <https://doi.org/10.1017/cbo9781316276105.005>.
- (5) Jensen, F. *Introduction to Computational Chemistry*; 2007. <https://doi.org/10.1017/CBO9781107415324.004>.
- (6) Lewars, E. G. *Computational Chemistry*.
- (7) Cramer, C. J. *Essentials of Computational Chemistry*.
- (8) Kojic-Prodic, B.; Cody, V. *Reviews in Computational Chemistry*; 1993; Vol. 30. [https://doi.org/10.1016/0263-7855\(93\)85039-s](https://doi.org/10.1016/0263-7855(93)85039-s).
- (9) Kato, T. On the Eigenfunctions of Many-particle Systems in Quantum Mechanics. *Commun. Pure Appl. Math.* **1957**, *10* (2), 151–177. <https://doi.org/10.1002/cpa.3160100201>.
- (10) Silverstone, HJ, Gailitis, M. On the Use of Asymptotic Expansions. 1987.
- (11) Huzinaga, S. Basis Sets for Molecular Calculations. *Comput. Phys. Reports* **1985**, *2* (6), 281–339. [https://doi.org/10.1016/0167-7977\(85\)90003-6](https://doi.org/10.1016/0167-7977(85)90003-6).
- (12) Huzinaga, S.; Klobukowski, M.; Tatewaki, H. The Well-Tempered GTF Basis Sets and Their Applications in the SCF Calculations on N 2 , CO, Na 2 , and P 2 . *Can. J. Chem.* **1985**, *63* (7), 1812–1828. <https://doi.org/10.1139/v85-302>.
- (13) Petersson, G. A.; Zhong, S.; Montgomery, J. A.; Frisch, M. J. On the Optimization of Gaussian Basis Sets. *J. Chem. Phys.* **2003**, *118* (3), 1101–1109. <https://doi.org/10.1063/1.1516801>.
- (14) Neese, F.; Ripplinger, C.; Auer, A. A.; Anh, T. Van; Datta, D.; Souza, B. De; Izsak, R.; Lang, L.; Lu, Q.; Gastel, V.; Sivalingam, K.; Stoychev, G.; Wennmohs, F. Jump-Start Guide The ORCA Quantum Chemistry Suite. **2018**, No. March.
- (15) Bienvenido al Libro del Web de Química del NIST <https://webbook.nist.gov/>
- (16) Johnson III, R. D. NIST Computational Chemistry Comparison and Benchmark Database NIST Standard Reference Database Number 101. <https://doi.org/10.18434/T47C7Z>.

6.1- Web Resources

- a) <https://sites.google.com/site/orcainputlibrary/> (Date of access 10-02-2020)
- b) <https://orcaforum.kofo.mpg.de/app.php/portal> (Date of access 10-02-2020)
- c) <https://www.basissetexchange.org> (Date of access 15-02-2020)
- d) <https://webbook.nist.gov> (Date of access 15-03-2020)
- e) <https://cccbdb.nist.gov> (Date of access 15-03-2020)

7 APPENDIX

7.1- Orca input

The input file has the general structure:

- (1) # Comments – Anything can be written here, and it won't affect the calculations
- (2) ! Keywords – Method, basis sets...
- (3) *xyz Charge. Mult
- (4) Atom A 0.000 0.000 0.000
- (5) Atom B 0.000 0.000 0.000
- (6) *
- (7) ...

Keywords can be given in any order and can be written in more than one line. The coordinates can be written in cartesian or in z-matrix.

7.2- Energy tables

Table 3. Energy values in E_h at fixed geometry for all the basis sets

Systems/ Functions	Li ₂	Be ₂	B ₂	C ₂	N ₂	O ₂	F ₂	LiF	BF	NF
3-21G	-14.7686	-28.6938	-48.7492	-74.9540	-108.3003	-148.6854	-197.6441	-106.3535	-123.4322	-152.8441
6-31G	-14.657	-28.8328	-49.0116	-75.3485	-108.8678	-149.4614	-198.6461	-106.9209	-124.0573	-153.6290
6-31G*	-14.8659	-28.8341	-49.0231	-75.3786	-108.9419	-149.5271	-198.6699	-106.9332	-124.1012	-153.6701
6-31G+	-14.8667	-28.9736	-49.0299	-75.3827	-108.9453	-149.5326	-198.6780	-106.9452	-124.1095	-153.6774
6-311G	-14.8698	-28.9118	-49.0244	-75.3628	-108.8942	-149.5118	-198.7089	-106.9620	-124.1017	-153.6760
6-311G*	-14.8700	-28.9131	-49.0369	-75.3937	-108.9694	-149.5697	-198.7271	-106.9686	-124.1366	-153.7131
6-311G+	-14.8700	-28.9793	-49.0376	-75.3944	-108.9717	-149.5738	-198.7318	-106.9751	-124.1390	-153.7176
def2-SVP	-14.8553	-28.8768	-48.9800	-75.3092	-108.8522	-149.4050	-198.5068	-106.8489	-124.0028	-153.5422
def2-TZVP	-14.8705	-28.9179	-49.0422	-75.4036	-108.9881	-149.5971	-198.7644	-106.9841	-124.1621	-153.7396
def2-QZVP	-14.8715	-28.9420	-49.0439	-75.4064	-108.9923	-149.6036	-198.7722	-106.9918	-124.1679	-153.7462
cc-pVDZ	-14.8695	-28.9022	-49.0355	-75.3869	-108.9541	-149.5429	-198.6857	-106.9455	-124.1075	-153.6833
cc-pVTZ	-14.8713	-28.9282	-49.0414	-75.4014	-108.9835	-149.5893	-198.7520	-106.9801	-124.1557	-153.7315
cc-pVQZ	-14.8715	-28.9424	-49.0436	-75.4058	-108.9911	-149.6012	-198.7683	-106.9907	-124.1658	-153.7434

Table 4. Energy values in E_h at optimized geometries for all the basis sets

Systems/ Functions	Li_2	Be_2	B_2	C_2	N_2	O_2	F_2	LiF	BF	NF
3-21G	-14.7693	-28.6939	-48.7529	-74.9543	-108.3010	-148.6872	-197.6442	-106.3542	-123.4336	-152.8458
6-31G	-14.8663	-28.8330	-49.0153	-75.3491	-108.8680	-149.4617	-198.6461	-106.9209	-124.0596	-153.6295
6-31G*	-14.8664	-28.8345	-49.0252	-75.3786	-108.9432	-149.5307	-198.6735	-106.9332	-124.1012	-153.6713
6-31G+	-14.8673	-28.9750	-49.0324	-75.3827	-108.9466	-149.5361	-198.6815	-106.9452	-124.1095	-153.6784
6-311G	-14.8702	-28.9118	-49.0281	-75.3631	-108.8946	-149.5121	-198.7091	-106.9620	-124.1021	-153.6763
6-311G*	-14.8703	-28.9131	-49.0394	-75.3937	-108.9719	-149.5752	-198.7322	-106.9686	-124.1368	-153.7151
6-311G+	-14.8704	-28.9820	-49.0401	-75.3944	-108.9741	-149.5792	-198.7373	-106.9751	-124.1392	-153.7195
def2-SVP	-14.8562	-28.8768	-48.9825	-75.3093	-108.8542	-149.4114	-198.5118	-106.8491	-124.0030	-153.5450
def2-TZP	-14.8710	-28.9194	-49.0445	-75.4036	-108.9914	-149.6024	-198.7705	-106.9841	-124.1624	-153.7417
def2-QZVP	-14.8718	-28.9420	-49.0461	-75.4064	-108.9957	-149.6091	-198.7784	-106.9918	-124.1684	-153.7483
cc-pVDZ	-14.8700	-28.9022	-49.0387	-75.3870	-108.9556	-149.5478	-198.6889	-106.9456	-124.1075	-153.6850
cc-pVTZ	-14.8717	-28.9282	-49.0438	-75.4015	-108.9866	-149.5945	-198.7580	-106.9801	-124.1561	-153.7335
cc-pVQZ	-14.8718	-28.9425	-49.0459	-75.4058	-108.9945	-149.6067	-198.7745	-106.9907	-124.1663	-153.7456

Table 5. Energy difference in E_h at fixed geometry with respect to the 3-21G basis set

Systems/ Functions	Li ₂	Be ₂	B ₂	C ₂	N ₂	O ₂	F ₂	LiF	BF	NF
3-21G	-14.7693	-28.6939	-48.7529	-74.9543	-108.3010	-148.6872	-197.6442	-106.3542	-123.4336	-152.8458
ΔE (6-31G)	0.0971	0.1391	0.2624	0.3948	0.5671	0.7745	1.0019	0.5667	0.6260	0.7837
ΔE (6-31G*)	0.0972	0.1406	0.2723	0.4244	0.6422	0.8435	1.0292	0.5790	0.6677	0.8255
ΔE (6-31G+)	0.0981	0.2811	0.2794	0.4284	0.6456	0.8489	1.0373	0.5910	0.6759	0.8326
ΔE (6-311G)	0.1009	0.2179	0.2752	0.4088	0.5936	0.8249	1.0649	0.6078	0.6685	0.8305
ΔE (6-311G*)	0.1011	0.2192	0.2864	0.4394	0.6710	0.8880	1.0880	0.6144	0.7033	0.8693
ΔE (6-311G+)	0.1011	0.2881	0.2872	0.4401	0.6731	0.8921	1.0930	0.6209	0.7056	0.8738
ΔE (def2-SVP)	0.0870	0.1829	0.2296	0.3550	0.5533	0.7242	0.8675	0.4949	0.5694	0.6992
ΔE (def2-TZP)	0.1017	0.2254	0.2916	0.4493	0.6905	0.9153	1.1262	0.6300	0.7289	0.8959
ΔE (def2-QZVP)	0.1026	0.2481	0.2932	0.4521	0.6948	0.9220	1.1342	0.6376	0.7348	0.9026
ΔE (cc-pVDZ)	0.1008	0.2083	0.2858	0.4328	0.6546	0.8607	1.0447	0.5914	0.6740	0.8392
ΔE (cc-pVTZ)	0.1024	0.2343	0.2909	0.4472	0.6856	0.9073	1.1138	0.6259	0.7225	0.8877
ΔE (cc-pVQZ)	0.1026	0.2485	0.2930	0.4515	0.6935	0.9195	1.1302	0.6365	0.7327	0.8998

Table 6. Energy difference in E_h at optimized geometries with respect to the 3-21G basis set

Systems/ Functions	Li ₂	Be ₂	B ₂	C ₂	N ₂	O ₂	F ₂	LiF	BF	NF
3-21G	-14.7686	-28.6938	-48.7492	-74.9540	-108.3003	-148.6854	-197.6441	-106.3535	-123.4322	-152.8441
ΔE (6-31G)	0.0971	0.1390	0.2624	0.3946	0.5675	0.7760	1.0019	0.5674	0.6252	0.7849
ΔE (6-31G*)	0.0972	0.1403	0.2739	0.4247	0.6416	0.8417	1.0257	0.5797	0.6690	0.8260
ΔE (6-31G+)	0.0981	0.2798	0.2807	0.4287	0.6450	0.8471	1.0338	0.5917	0.6773	0.8333
ΔE (6-311G)	0.1012	0.2180	0.2752	0.4088	0.5939	0.8263	1.0648	0.6085	0.6695	0.8319
ΔE (6-311G*)	0.1013	0.2192	0.2877	0.4397	0.6692	0.8842	1.0829	0.6151	0.7045	0.8690
ΔE (6-311G+)	0.1014	0.2855	0.2884	0.4404	0.6714	0.8883	1.0876	0.6216	0.7068	0.8735
ΔE (def2-SVP)	0.0867	0.1830	0.2308	0.3553	0.5519	0.7196	0.8626	0.4954	0.5706	0.6982
ΔE (def2-TZP)	0.1019	0.2241	0.2930	0.4496	0.6879	0.9116	1.1202	0.6306	0.7300	0.8955
ΔE (def2-QZVP)	0.1028	0.2481	0.2947	0.4524	0.6921	0.9182	1.1280	0.6383	0.7358	0.9021
ΔE (cc-pVDZ)	0.1008	0.2083	0.2863	0.4329	0.6539	0.8575	1.0415	0.5921	0.6754	0.8393
ΔE (cc-pVTZ)	0.1027	0.2343	0.2922	0.4475	0.6832	0.9038	1.1079	0.6266	0.7236	0.8874
ΔE (cc-pVQZ)	0.1028	0.2486	0.2944	0.4518	0.6908	0.9157	1.1241	0.6372	0.7337	0.8994

Table 7. Internuclear distance in Angstrom at optimized geometry for all the basis sets

Systems/ Functions	Li ₂	Be ₂	B ₂	C ₂	N ₂	O ₂	F ₂	LiF	BF	NF
NIST	2.673	2.46	1.59	1.243	1.098	1.208	1.412	1.564	1.304	1.317
3-21G	2.8157	2.4117	1.7004	1.2573	1.0828	1.2419	1.4025	1.5203	1.3026	1.3635
6-31G	2.8155	2.3876	1.6971	1.2623	1.081	1.1933	1.4125	1.5712	1.3182	1.343
6-31G*	2.8111	2.3523	1.6686	1.2447	1.0781	1.1648	1.3472	1.5567	1.2613	1.2801
6-311G	2.7846	2.4794	1.7001	1.2577	1.086	1.1935	1.3968	1.5651	1.2882	1.3395
6-311G*	2.7844	2.4185	1.6766	1.2442	1.0703	1.154	1.3314	1.559	1.2524	1.2869
def2-SVP	2.8549	2.4385	1.6749	1.2454	1.0735	1.1512	1.3341	1.5406	1.2518	1.262
def2-TZP	2.798	2.2545	1.6744	1.2397	1.0661	1.1541	1.3282	1.5548	1.2492	1.2682
def2-QZVP	2.7839	2.4345	1.6739	1.2388	1.0654	1.1532	1.327	1.5581	1.2451	1.2671
cc-pVDZ	2.8086	2.4412	1.6875	1.2523	1.0773	1.1569	1.3476	1.5788	1.2717	1.2729
cc-pVTZ	2.7836	2.4322	1.6761	1.2409	1.0671	1.1549	1.3291	1.5671	1.2485	1.2689
cc-pVQZ	2.7839	2.4266	1.674	1.2389	1.0656	1.1535	1.3275	1.5537	1.2453	1.2675

Table 8. Vibrational constant in cm^{-1} at optimized geometry for all the basis sets

Systems/ Functions	Li ₂	Be ₂	B ₂	C ₂	N ₂	O ₂	F ₂	LiF	BF	NF
NIST	339.36	268.97	872.2	1904.22	2729.31	1981.35	1263.81	924.72	1395.48	1128.55
3-21G	342.11	320.72	820.36	1837.37	2612.13	1599.75	1295.83	954.94	1659.3	1309.54
6-31G	340.57	348.83	854.42	1851.73	2661.1	1762.49	1141.5	1019.43	1502.54	1239.26
6-31G*	338.09	-	885.24	1925.08	2754.25	1989.26	1226.42	956.62	1426.12	1166.42
6-311G	336.49	267.01	825.47	1824.25	2591.94	1815.07	1105.12	972.42	1395.15	1234.26
6-311G*	338.67	-	875.41	1904.48	2734.05	1997.11	1224.72	959.31	1385.45	1131.5
def2-SVP	326.8	300.18	914.97	1951.39	2791.82	2048.36	1222.3	957.63	1422.23	1126
def2-TZP	333.89	385.21	873.29	1906.93	2734.85	1979.02	1267.61	928.34	1415	1130.23
def2-QZVP	339.36	261.13	871.84	1904.05	2730.13	1981.21	1264.25	926.84	1394.63	1126.71
cc-pVDZ	338.78	295.61	879.11	1913.42	2757.86	1996.79	1180.11	1042.6	1423.24	1130.49
cc-pVTZ	339.53	276.35	872.82	1903.72	2731.27	1978.8	1266.78	944.11	1412.69	1137.31
cc-pVQZ	339.36	268.97	872.2	1904.22	2729.31	1981.35	1263.81	924.72	1395.48	1128.55

7.3- Geometries

The internuclear distances of the systems were obtained from the CCCBDB¹⁶. These geometries were used directly as well as optimized for each of the basis sets studied.

Table 9. Experimental internuclear distances used in the input for the calculations following both approaches.

System	R_e (Å)
Li₂	2.6730
Be₂	2.4600
B₂	1.5900
C₂	1.2425
N₂	1.0980
O₂	1.2075
F₂	1.4119
LiF	1.5514
BF	1.2734
NF	1.3114

In Tables 10 to 19 the optimized geometries generated by ORCA are collected for all the systems and corresponding basis sets.

Table 10. Optimized internuclear distances for the Li₂ molecule for all the basis sets.

Li₂	R_e (Å)
3-21G	2.8157
6-31G	2.8155
6-311G	2.7846
def2-SVP	2.8549
def2-TZVP	2.7980
def2-QZVP	2.7839
cc-pVDZ	2.8085
cc-pVTZ	2.7836
cc-pVQZ	2.7839

Table 11. Optimized geometries for the Be₂ systems at all the basis sets in cartesian coordinates.

Be₂	R_e (Å)
3-21G	2.4117
6-31G	2.3876
6-311G	2.4793
def2-SVP	2.4385
def2-TZVP	2.2545
def2-QZVP	2.4344
cc-pVDZ	2.4412
cc-pVTZ	2.4322
cc-pVQZ	2.4266

Table 12. Optimized geometries for the B₂ systems at all the basis sets in cartesian coordinates.

B₂	R_e (Å)
3-21G	1.7004
6-31G	1.6870
6-311G	1.7000
def2-SVP	1.6750
def2-TZVP	1.6744
def2-QZVP	1.6739
cc-pVDZ	1.6875
cc-pVTZ	1.6818
cc-pVQZ	1.6740

Table 13. Optimized geometries for the C₂ systems at all the basis sets in cartesian coordinates.

C₂	R_e (Å)
3-21G	1.2573
6-31G	1.2622
6-311G	1.2577
def2-SVP	1.2454
def2-TZVP	1.2397
def2-QZVP	1.2388
cc-pVDZ	1.2523
cc-pVTZ	1.2409
cc-pVQZ	1.2389

Table 14. Optimized geometries for the N₂ systems at all the basis sets in cartesian coordinates.

N₂	R_c (Å)
3-21G	1.0828
6-31G	1.0891
6-311G	1.0860
def2-SVP	1.0735
def2-TZVP	1.0661
def2-QZVP	1.0654
cc-pVDZ	1.0773
cc-pVTZ	1.0671
cc-pVQZ	1.0659

Table 15. Optimized geometries for the O₂ systems at all the basis sets in cartesian coordinates.

O₂	R_c (Å)
3-21G	1.2419
6-31G	1.1933
6-311G	1.1935
def2-SVP	1.1512
def2-TZVP	1.1542
def2-QZVP	1.1532
cc-pVDZ	1.1570
cc-pVTZ	1.1549
cc-pVQZ	1.1535

Table 16. Optimized geometries for the F₂ systems at all the basis sets in cartesian coordinates.

F₂	R_c (Å)
3-21G	1.4025
6-31G	1.4125
6-311G	1.3967
def2-SVP	1.3341
def2-TZVP	1.3282
def2-QZVP	1.3270
cc-pVDZ	1.3476
cc-pVTZ	1.3291
cc-pVQZ	1.3275

Table 17. Optimized geometries for the LiF systems at all the basis sets in cartesian coordinates.

LiF	R_e (Å)
3-21G	1.5203
6-31G	1.5712
6-311G	1.5651
def2-SVP	1.5406
def2-TZVP	1.5548
def2-QZVP	1.5581
cc-pVDZ	1.5788
cc-pVTZ	1.5671
cc-pVQZ	1.5537

Table 18. Optimized geometries for the BF systems at all the basis sets in cartesian coordinates.

BF	R_e (Å)
3-21G	1.3026
6-31G	1.3182
6-311G	1.2882
def2-SVP	1.2518
def2-TZVP	1.2492
def2-QZVP	1.2451
cc-pVDZ	1.2717
cc-pVTZ	1.2485
cc-pVQZ	1.2453

Table 19. Optimized geometries for the NF systems at all the basis sets in cartesian coordinates.

NF	R_e (Å)
3-21G	1.3635
6-31G	1.3430
6-311G	1.3395
def2-SVP	1.2620
def2-TZVP	1.2682
def2-QZVP	1.2671
cc-pVDZ	1.2729
cc-pVTZ	1.2689
cc-pVQZ	1.2674