DANIELA VELOSO COELHO

Building a Potential Energy Surface for the Ground Singlet State of the Hydrogen Peroxide

Faculdade de Ciências e Tecnologia

2016

DANIELA VELOSO COELHO

Building a Potential Energy Surface for the Ground Singlet State of the Hydrogen Peroxide

Doutoramento em Química

(Especialidade em Química Teórica)

Trabalho efetuado sob a orientação de:

Prof. Doutor João Carlos Pereira Peres Brandão

e

Professor Doutor Francisco Javier Aoiz Moleres

Faculdade de Ciências e Tecnologia

2016

Building a Potential Energy Surface for the Ground Singlet State of the Hydrogen Peroxide

Declaração de autoria de autor

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

(Daniela Veloso Coelho)

Agradecimentos

Em primeiro lugar quero manifestar o meu mais sincero agradecimento ao Prof. Doutor João Carlos Pereira Peres Brandão, meu orientador, pelos conhecimentos que adquiri sob a sua supervisão e pelo apoio ao longo de todos estes anos em que desenvolvi os trabalhos que levaram à conclusão da presente dissertação.

Ao Professor Doutor Francisco Javier Aoiz Moleres, meu coorientador, expresso o meu agradecimento pela valiosa análise crítica ao presente trabalho que em muito contribuiu para o seu aprimoramento e pela simpatia e disponibilidade demonstrada.

Agradeço particularmente ao meus colegas da Universidade do Algarve, em especial aos do grupo de Química Teórica, por todo o apoio e sugestões que me deram.

Ao Departamento de Química e Farmácia da Faculdade de Ciências e Tecnologia da Universidade do Algarve, em que estou integrada, agradeço a concessão das condições necessárias à realização do meu trabalho de investigação.

À Fundação para a Ciência e Tecnologia agradeço o apoio financeiro concedido através da bolsa SFRH/BD/64675/2009.

Ao Centro de Investigação em Química do Algarve agradeço o suporte financeiro de algumas das deslocações a conferências.

Agradeço à Dra. Augusta pela revisão da língua inglesa.

À minha família agradeço todo o sacrifício que fez, toda a compreensão e todo o carinho que sempre manisfestou, ajudando-me a superar as várias dificuldades com as quais me deparei ao longo dos últimos anos. Sem o seu apoio não teria sido possível terminar o doutoramento sendo, por isso, a ela que dedico o trabalho aqui apresentado.

Units

In this thesis, unless otherwise stated, it was employed the atomic unit system to describe the various quantities. The conversion to the International System of Units (SI) is given below:

1 a.u. of length (Bohr, a_0) = $5.29177249 \times 10^{-11}$ m;
1 a.u. of mass (m_e) = $9.1093897 \times 10^{-31}$ kg;
1 a.u. of charge (e) = $1.60217733 \times 10^{-19}$ C;
1 a.u. of energy (Hartree, E_h) = $4.3597482 \times 10^{-18}$ J;
1 a.u. of angular momentum ($\hbar$) = $1.05457266 \times 10^{-34}$ Js
1 a.u. of electrostatic constant ($4\pi\varepsilon_0$) = $1.11265006 \times 10^{-10}$ $\mathrm{J^{-1}C^2m^{-1}}$.

In addition to the atomic unit system, others were also used:

$1\,\mathrm{kcal\,mol^{-1}} = 6.9477014 \times 10^{-21}$ J;
$1\,\mathrm{cm^{-1}} = 1.9864475 \times 10^{-23}$ J;
1 Ångstrom ($\mathring{A}$) = 1×10^{-10} m;
$1^\circ = (\pi/180)$ rad.

Resumo

De entre os vários sistemas químicos de interesse, a molécula de peróxido de hidrogénio merece especial atenção, desempenhando um papel fundamental em muitos processos químicos nomeadamente, na combustão e na química da atmosfera e do meio interestelar. Por esta razão, esta molécula tem sido o tópico de enúmeros estudos quer teóricos, quer experimentais ao longo dos últimos anos.

Do ponto de vista teórico, o conhecimento das propriedades moleculares do sistema de peróxido de hidrogénio, assim como de qualquer outro sistema molecular, requer, *à priori*, o conhecimento da Superfície de Energia Potencial (SEP), função que governa o movimento dos núcleos num potencial criado pelas energias eletrónica e de repulsão internuclear definidas para cada configuração nuclear do sistema e cujo conceito tem por génese a separação de Born-Oppenheimer.

A construção de uma SEP com precisão adequada tendo em conta as propriedades do sistema que se pretendem avaliar, acarreta vários condicionamentos quer de carácter formal, quer de carácter computacional, não obstante, continua a ser uma fonte primordial de informação para a compreensão de muitos processos químicos e muitos têm sido os avanços no sentido da criação e aplicação de métodos e técnicas mais eficientes e expeditos para a sua obtenção.

Posto isto, o objetivo da presente tese de doutoramento incide na construção de uma SEP global com precisão química para a molécula de peróxido de hidrogénio no seu estado fundamental, para posteriormente ser utilizada em estudos de dinâmica molecular.

Para a representação analítica da SEP aplicou-se o método da dupla expansão multi-corpos (DBME). Com este formalismo, o potencial do sistema em consideração é dado pelo sumatório sobre todos os fragmentos representativos dos termos a 1, 2, 3 e 4-corpos o que garante a dissociação correta do sistema. Cada um destes termos é por sua vez separado nas suas componentes de curtas e longas distâncias as quais apresentam dependências funcionais distintas. Para o sistema em estudo, foi necessário recorrer a uma matriz 3×3 a fim de reproduzir com precisão todos os canais de dissociação nomeadamente, $O(^1D)$ + $H_2O(X\ ^1A_1)$, $OH(X\ ^2\Pi)$ + $OH(X\ ^2\Pi)$, $O_2(a\ ^1\Delta_g)$ + $H_2(X\ ^1\Sigma_g^+)$ e $H(^2S)$ + $HO_2(X\ ^2A'')$,

de acordo com as regras de Wigner-Witmer.

Na primeira parte da dissertação são descritos os princípios teóricos que serviram de base à investigação desenvolvida. Assim, abordar-se-á de forma sucinta o conceito de superfície de energia potencial à luz da aproximação de Born-Oppenheimer, assim como os sistemas de coordenadas mais frequentemente utilizados na sua representação e as características mais relevantes da sua topologia. Os princípios básicos associados aos métodos de cálculo de estrutura eletrónica, chamados de cálculos *ab initio* por recorrerem apenas aos primeiros princípios, serão igualmente abordados, dando particular ênfase aos aplicados na arquitetura da presente SEP.

Foram efetuados cálculos de estrutura eletrónica com base em dois métodos teóricos, o método Coupled-Cluster (CC) e o método de cálculo baseado na Teoria de Perturbação Multiconfiguracional, sendo a utilização de um em detrimento de outro, condicionada pelas caraterísticas eletrónicas das espécies que prevalecem na zona específica que se pretende estudar. O elevado custo, em termos computacionais, dos métodos de cálculo *ab initio*, que rapidamente cresce com o tamanho da base de funções utilizada, foi a força impulsionadora do desenvolvimento de um esquema de extrapolação/escalonamento para prever com elevada precisão energias calculadas usando bases de funções menos dispendiosas.

A forma funcional utilizada para representar as interações a 4-corpos a curtas distâncias baseia-se numa soma de funções polinomiais do quarto grau, cada uma multiplicada por um termo de amortecimento, ambos centrados em geometrias de referência específicas. Relativamente às interações a 4-corpos a longas distâncias, apenas foram consideradas as interações eletrostáticas dipolo-dipolo entre dois radicais $OH(^2\Pi)$.

Devido à invariância do Hamiltoneano eletrónico em relação à permutação de quaisquer átomos idênticos, a SEP para um sistema molecular constituído por dois ou mais átomos da mesma espécie química, deve satisfazer esta simetria permutacional. A implementação da simetria permutacional na SEP foi efetuada com o recurso a funções polinomiais invariantes. Para tal, aplicou-se o teorema de Molien para determinar o número mínimo de termos, a chamada base de integridade, para construir a função invariante e o número de termos de cada grau que devem ser utilizados.

A definição da base de integridade passa pelo recurso a operadores de projeção. Com base na aplicação destes operadores, determinaram-se as sete invariantes a serem utilizadas na contrução de cada polinómio. Sendo polinómios do quarto grau teremos, então, um total de 75 termos por polinómio.

O processo de ajuste baseia-se num ajuste multifuncional linear envolvendo um total de 21 geometrias de referência, o que perfaz $21 \times 75 = 1575$ parâmetros a ajustar. De entre as referências, 11 foram escolhidas de modo a cobrir as zonas do espaço configuracional de

maior relevância na dinâmica das reações que ocorrem na SEP. A atribuição dos pontos *ab initio* às correspondentes geometrias centrais processou-se através do recurso a uma técnica de partição denominada por k-means em que os pontos são distribuídos com base na minimização da distância euclidiana. As restantes geometrias, definidas de modo a cobrir o espaço configuracional não abrangido pelas onze estruturas de referência fixas, foram otimizadas durante o processo de atribuição dos pontos de modo a que cada uma tivesse pelo menos 300 pontos na sua vizinhança.

Tendo em conta a opção por um ajuste multifuncional, a abordagem passou por utilizar funções gaussianas modificadas para definir os termos de amortecimento de modo a minimizar a interferência entre os vários termos polinomiais. Para manter o ajuste linear, os parâmetros destas funções gaussianas foram definidos a partir da dispersão dos pontos na vizinhança de cada centro.

Características relevantes da função de energia potencial modelada foram investigadas através da determinação de pontos de estacionariedade, do cálculo de barreiras energéticas e da análise de gráficos ilustrativos de cortes a uma dimensão ao longo de coordenadas selecionadas bem como, de gráficos de contornos referentes a vários canais reacionais. Na ausência de dados experimentais precisos, a aferição da qualidade da superfície é estabelecida por comparação com resultados teóricos da literatura.

Por fim, as principais conclusões serão sumariadas. Futuros melhoramentos e possíveis aplicações da presente superfície de energia potencial são igualmente sugeridos.

Palavras Chave: Química Teórica, Superfície de Energia Potencial, Peróxido de Hidrogénio, Cálculos *ab initio* de estrutura eletrónica

Abstract

The hydrogen peroxide system has attracted widespread attention due to its overwhelming importance in many chemical processes such as combustion, atmospheric and interstellar chemistry.

The main goal of the present PhD thesis is the construction of a six-dimensional Potential Energy Surface (PES) for the ground state of the H_2O_2 system with chemical accuracy in order to achieve trustworthy kinetic data. The analytical representation of the PES is based on a double many-body expansion (DMBE) formalism, where a 3×3 matrix is used to accurately reproduce all the dissociation channels.

The first part of this thesis concerns the concept of PES, the coordinates used to represent it as well as its major topological features. A survey of the theoretical framework of the *ab initio* calculations and the general strategy to obtain the analytical representation of the PES are also reported. The second part focuses on the description of the *ab initio* electronic calculations performed for mapping the most important regions of the configuration space. An extrapolation/scaling scheme is proposed to accomplish high-quality *ab initio* energies. The parameters used in this procedure are obtained by interpolation among several reference geometries.

This new DMBE hydrogen peroxide PES also accounts for the electrostatic dipole-dipole interaction between two OH($^2\Pi$) fragments. The functional form used to represent the short-range interactions is based on a sum of polynomial functions of the fourth degree multiplied by a range factor, both built with intrinsic permutation symmetry and centred at specific reference geometries, to which the *ab initio* points computed are assigned based on a k-means algorithm.

Finally, important features of the model function are characterized and the major conclusions are summarized. Further improvements and possible applications for the present potential are also outlined.

Keywords: Theoretical Chemistry, Potential Energy Surface, Hydrogen Peroxide, *Ab initio* Electronic Structure Calculations

Contents

List of Figures

List of Tables

Acronyms

aug-cc-pVXZ	augmented correlation-consistent polarized valence basis sets
BSSE	Basis Set Superposition Error
CASPT2	Complete Active Space with Second-order Perturbation Theory
CASSCF	Complete Active Space Self-Consistent Field
CC	Coupled-Cluster Theory
CCD	Coupled-Cluster with Double excitations
cc-pCVXZ	correlation-consistent polarized core-valence basis sets
cc-pVXZ	correlation-consistent polarized valence basis sets
CCSD	Coupled-Cluster with Single and Double and Perturbative Triple excitations
CGTOs	Contracted Gaussian-type Orbitals
CI	Configuration Interaction
CISD	Configuration Interaction with Single and Double excitations
CNP	Complete Nuclear Permutation
CP	Counterpoise correction
CSF	Configuration State Functions
DBME	Double Many-Body Expansion
DFT	Density Functional Theory
EHF	*extended* Hartree-Fock
FORS	Full Optimized Reaction Space
FV CASSCF	Full Valence Complete Active Space Self-Consistent Field
GTOs	Gaussian-type orbitals
HF	Hartree-Fock

Acronyms

HOMO	Highest Occupied Molecular Orbital
LCAO	Linear Combination of Atomic Orbitals
LUMO	Lowest Unoccupied Molecular Orbital
MBE	Many-Body Expansion
MCSCF	Multiconfiguration Self-Consistent Field
MPPT	Møller-Plesset Perturbation Theory
MRCI	Multireference Configuration Interaction
MRCISD	Multireference Configuration Interaction with Single and Double Excitations
PES	Potential Energy Surface
PT	Perturbation Theory
QCISD	Quadratic Configuration Interaction with Single and Double excitations
QCT	Quasiclassical Trajectory
SCF	Self-Consistent Field
STOs	Slater-type orbitals
SV	Split-valence Basis Set
SVD	Single Value Decomposition
ZPE	Zero Point Energy

Part I

Theoretical Background

Chapter 1

The concept of the Potential Energy Surface

The Potential Energy Surface (PES) plays an important role in the understanding of molecular phenomena such as molecular structure, reaction dynamics and spectroscopy. Therefore, it is not difficult to perceive the effort applied during the past years, mainly in the branch of theoretical chemistry, in the conception of suitable methods to its construction. Given this, the major challenge is to obtain the solutions to the nonrelativistic time-independent Schrödinger equation, since the exact description of a physical system in its stationary state is given by the resolution of this equation regarding to all the particles that constitute the system. Unfortunately, this is far from being a trivial task since the analytical solution is only possible for a few simple systems like the hydrogen atom [1] and, for this reason, approximations must be considered.

One of the most important approximations concerning the understanding of molecular properties is the Born-Oppenheimer approximation [2] and it is on this assumption that lies the concept of the Potential Energy Surface.

1.1 The Born-Oppenheimer approximation

Considering a general molecular system comprising N nuclei and n electrons, the nonrelativistic time-independent Hamiltonian can be written as a sum of five terms,

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_{i=1}\nabla_i^2 + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} - \sum_{A=1}^{N}\sum_{i=1}^{n}\frac{Z_A e^2}{4\pi\varepsilon_0 r_{Ai}} + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B e^2}{4\pi\varepsilon_0 r_{AB}} - \sum_{A=1}^{N}\frac{\hbar^2}{2M_A}\nabla_A^2, \tag{1.1}$$

where i, j refer to the electrons and A, B refer to the nuclei.

In Equation 1.1 the distance between the electron i and nucleus A is written as $r_{iA} = |r_i - r_A|$, and similar definitions are adopted for the distances r_{ij} and r_{AB}. The constants Z_A and M_A stand for the atomic number and mass of the nucleus A. The Laplacian operators ∇_i^2 and ∇_A^2 give the differentiation with respect to the coordinates of the electron i and to the coordinates of the nucleus A, respectively. The terms represent, in order of appearance, the electronic kinetic energy, the electron-electron repulsions, the electron-nuclear attraction, the nuclear-nuclear repulsion and the nuclear kinetic energy operators.

At this point we shall introduce the system of atomic units widely used to report atomic and molecular calculations. In this system of units the electron's mass (m_e), the proton's charge (e), the unit of angular momentum ($\hbar$), and the vacuum permittivity ($4\pi\varepsilon_0$) each have the numerical value of 1 [1, 3]. Hence, through the introduction of this set of units, Equation 1.1 becomes,

$$\hat{H} = -\frac{1}{2}\sum_{i=1}\nabla_i^2 + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} - \sum_{A=1}^{N}\sum_{i=1}^{n}\frac{Z_A}{r_{Ai}} + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B}{r_{AB}} - \sum_{A=1}^{N}\frac{1}{2M_A}\nabla_A^2. \qquad (1.2)$$

Unless otherwise noted, from now on, atomic units will be used in this thesis.

Once established the Hamiltonian, the wave functions and energies of the system are the solutions to the time-independent Schrödinger equation,

$$\hat{H}\Psi(\mathbf{r}, \mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R}), \qquad (1.3)$$

where $\mathbf{r}$ and $\mathbf{R}$ designate the electronic and nuclear coordinates, respectively. The first approach to solve Equation 1.3 is to take advantage of the fact that the nuclei are much heavier than the electrons. The mass disparity between both particles allows us to say that the electrons move much faster than the nuclei and so we can consider the nuclei nearly fixed with respect to the electron motion. This approximation, known as the Born-Oppenheimer approximation [2], is one of the cornerstones of theoretical studies reducing significantly the mathematical treatment behind the application of the Schrödinger equation to atomic and molecular systems. The basic idea is that, due to the higher velocities of the electrons they are able to adapt themselves instantaneously to any change in the nuclear configuration. This means that the motion of the nuclei can be studied by considering that the nuclei move on a PES, which are solutions to the electronic Schrödinger equation, including internuclear repulsion, for different arrangements of the nuclei.

So, within this aproximation, Equation 1.2 can be split into an electronic Hamiltonian

for fixed internuclear distances, known as the clamped-nuclei (CN) Hamiltonian [4],

$$\hat{H}_{CN} = -\frac{1}{2}\sum_{i=1}\nabla_i^2 + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} - \sum_{A=1}^{N}\sum_{i=1}^{n}\frac{Z_A}{r_{Ai}} + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B}{r_{AB}}, \tag{1.4}$$

and a term, $\hat{T}_N$, referring to the nuclear kinetic energy operator,

$$\hat{T}_N = -\sum_{A=1}^{N}\frac{1}{2M_A}\nabla_A^2. \tag{1.5}$$

Since the nuclear-nuclear repulsion term is independent from the electronic coordinates and is a constant for a fixed nuclear configuration, it can be omitted from Equation 1.4 yielding a purely electronic Hamiltonian $\hat{H}_{elec}$,

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}\nabla_i^2 + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} - \sum_{A=1}^{N}\sum_{i=1}^{n}\frac{Z_A}{r_{Ai}}. \tag{1.6}$$

Considering the nuclei to be static, the motion of the electrons is dictated by the Schrödinger equation involving only $\hat{H}_{elec}$,

$$\hat{H}_{elec}(\mathbf{r},\mathbf{R})\phi(\mathbf{r};\mathbf{R}) = E_{elec}(\mathbf{R})\phi(\mathbf{r};\mathbf{R}); \tag{1.7}$$

where $\phi(\mathbf{r};\mathbf{R})$ is the electronic wave function, which depends explicitly on the electronic coordinates and parametric on the nuclear coordinates. After finding the value of $E_{elec}(\mathbf{R})$, for a fixed arrangement of the nuclei, the clamped-nuclei energy which is given by the electronic energy including internuclear repulsion, $U(\mathbf{R})$, can easily be obtained by solving the internuclear repulsion term and then add it to E_{elec} [1],

$$U(\mathbf{R}) = E_{elec}(\mathbf{R}) + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B}{r_{AB}}. \tag{1.8}$$

It is this clamped-nuclei potential, $U(\mathbf{R})$, that we shall consider as the PES for the nuclear motion in the Born-Oppenheimer approximation [5].

Since the Hamiltonian operator is an hermitian operator the adiabatic eigenfunctions of the electronic Hamiltonian $\phi_i(\mathbf{r};\mathbf{R})$ can be chosen to be an orthonormal complete basis set for each nuclear configuration. The complete wave function of the system can be obtained as an expansion in the complete set of the electronic functions being the

coefficient of the expansion, $\chi(\mathbf{R})$, functions of the nuclear coordinates [6],

$$\Psi(\mathbf{r},\mathbf{R}) = \sum_i \chi_i(\mathbf{R})\phi_i(\mathbf{r};\mathbf{R}), \tag{1.9}$$

Subtituing Equation 1.9 into the Schrödinger equation 1.3, it gives

$$\hat{H}\sum_i \chi_i(\mathbf{R})\phi_i(\mathbf{r};\mathbf{R}) = E\sum_i \chi_i(\mathbf{R})\phi_i(\mathbf{r};\mathbf{R}). \tag{1.10}$$

Considering the orthogonality of $\phi_i(\mathbf{r};\mathbf{R})$, if we multiply both terms of the previous equation by $\phi_j(\mathbf{r};\mathbf{R})$ and proceed to the integration over the electronic coordinates, we obtain the following set of differential equations with respect to the nuclear coordinates,

$$\sum_i \left\langle \phi_j(\mathbf{r};\mathbf{R}) \mid \hat{H} \mid \phi_i(\mathbf{r};\mathbf{R})\chi_i(\mathbf{R}) \right\rangle = E\chi_j(\mathbf{R}), \tag{1.11}$$

where the Dirac notation is applied to represent the integrals over the electronic coordinates. Replacing in this last equation, sequentially, Equation 1.2, Equation 1.6 and Equation 1.7, and after some algebraic manipulation we obtain

$$\left[E_j(\mathbf{R}) + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B}{r_{AB}} - E\right]\chi_j(\mathbf{R}) - \sum_i\sum_{A=1}^{N}\left\langle \phi_j(\mathbf{r};\mathbf{R}) \mid \frac{\nabla_A^2}{2M_A} \mid \phi_i(\mathbf{r};\mathbf{R})\chi_i(\mathbf{R}) \right\rangle = 0. \tag{1.12}$$

The operator ∇_A^2 involves the differentiation concerning the Ath nucleus and so, it acts explicitly on the nuclear wave function and parametrically on the electronic wave functions. Developing this operator leads to the following equation,

$$\left[-\sum_{A=1}^{N}\frac{\nabla_A^2}{2M_A} + E_j(\mathbf{R}) + \sum_{A=1}^{N}\sum_{B>1}^{N}\frac{Z_A Z_B}{r_{AB}} - E\right]\chi_j(\mathbf{R}) = \hat{C}_{ji}\chi_i(\mathbf{R}), \tag{1.13}$$

where $\hat{C}_{ji}$ is the coupling operator, which is defined as,

$$\hat{C}_{ji} = \sum_{A=1}^{N}\left[\left\langle \phi_j(\mathbf{r};\mathbf{R}) \mid \frac{\nabla_A^2}{2M_A} \mid \phi_i(\mathbf{r};\mathbf{R}) \right\rangle + \left\langle \phi_j(\mathbf{r};\mathbf{R}) \mid \frac{\nabla_A}{M_A} \mid \phi_i(\mathbf{r};\mathbf{R}) \right\rangle \nabla_A\right]. \tag{1.14}$$

The coupling terms in the previous equation give the variation of the electronic wave function regarding the nuclear coordinates. Since, as we have already mentioned, the nuclei move much more slowly than the electrons, these coupling terms can be considered as negligible. By doing so, Equation 1.13 becomes a set of uncoupled equations, the

independent Schrödinger equations for the nuclear motion,

$$\left[-\sum_{A=1}^{N} \frac{\nabla_A^2}{2M_A} + E_j(\mathbf{R}) + \sum_{A=1}^{N} \sum_{B>1}^{N} \frac{Z_A Z_B}{r_{AB}} \right] \chi_j(\mathbf{R}) = E\chi_j(\mathbf{R}). \tag{1.15}$$

In the previous equation, the terms inside the square brackets give the effective Hamiltonian for the nuclear motion, $\hat{H}_{Nuc}(\mathbf{R})$, given by,

$$\hat{H}_{Nuc}(\mathbf{R}) = \hat{T}_{Nuc} + U(\mathbf{R}). \tag{1.16}$$

Summarizing, Equation 1.9, which approaches the total wave function as a product of electronic and nuclear wave functions, and Equation 1.14, which discards the coupling between the different electronic states, constitute the foundations of the Born-Oppenheimer approximation. Within this approximation, the application of the Schrödinger equation is considerably simplified by allowing us to obtain the eigenvalues of the electronic Hamiltonian operator, followed by their replacement in Equation 1.15 to attain, by addition to the nuclear energy contribution, the total energy of the system.

In everything that follows it shall be deemed as valid the Born-Oppenheimer approximation and all the work will be based on this assumption.

1.2 Coordinates for the description of Potential Energy Surfaces

To describe a system each atom must be uniquely represented. One common representation is to consider the Cartesian coordinates for all the atoms. The movement of the N atoms within a molecule is then totally described by $3N$ coordinates. However, not all coordinates are important to define the PES since the potential energy is invariant concerning the translation and the rotation of a molecule. Excluding the degrees of freedom that describe the molecular translational and rotational symmetry, the PES becomes only a function of the internal movements of a system which are defined by $3N-6$ coordinates for a nonlinear N-atom molecule or $3N-5$ for a linear molecule [7].

A typical approach to represent a molecule, neglecting its position and orientation in space, is to adopt a system of internal coordinates. There are a few possibilities for the definition of these internal coordinates. For instance, in the case of a three-atom system, such as the water molecule, the potential surface will depend on a total of three independent internal coordinates which may be defined by three internuclear distances, two distances and the angle between both, one distance and two angles, among others.

The choice of an internal coordinate system over another must rely on the one that better suits the problem that we are addressing to [8].

There is, however, an important issue that one must account for when potentials are formulated based on internal coordinates. It exists a set of $N(N-1)/2$ internuclear distances for an N-atom molecule but only $3N-6$ of these are independent. This is not problematic when we are dealing with small molecules where $N \leq 4$, since for these cases the degrees of freedom of the molecule equals the number of internuclear coordinates and so, these last ones can be used directly as internal coordinates. However, for molecules with more than four atoms, the number of internal coordinates exceeds the independent internal coordinates being the divergence between both more pronounced as the number of atoms increases. This redundancy leads to some dubiety in the choice of a set of internuclear coordinates [7] but, nevertheless, they still remain the most common choice in the description of PESs.

Thus, the internuclear distance representation is a valid choice of coordinates for small molecules (with $N \leq 4$) such as the hydrogen peroxide, the target molecule of our study, where there are as many internuclear distances (total of six) as there are independent internal coordinates. Yet, one must be aware that when the internal coordinates are chosen to be the bond lengths within a molecule, these internuclear distances are restricted to the triangle inequalities,

$$R_i \leq R_j + R_k, \qquad i,j,k = AB, AC, BC. \tag{1.17}$$

The triangle inequality theorem states that the sum of the lengths of any two sides of a triangle must be greater or equal to the length of the third side. This means that the configuration space of a molecule is limited to geometries that are physically real, i.e., geometries for which the above restrictions are satisfied [7]. We must be aware that when we are dealing with a triatomic system the restriction of the bond length distances to the triangle inequalities are enough, but when it comes to tetratomic molecules, such as the hydrogen peroxide, additional constrains need to be introduced. For four-atom systems we should also consider that the sum of the angles of the planes conjoint in each atom of the molecule must be inferior or equal to 360°, and one of these angles must be inferior to the sum of the other two.

Another important aspect that needs special consideration is that, for a molecular system composed of two or more identical atoms, such as the hydrogen peroxide, the PES must be invariant in what concerns the interchange of these indistinguishable nuclei, becoming useful to use coordinates with the appropriate permutation symmetry. The construction of these symmetry coordinates and its application in the development of

invariant polynomial functional forms for the representation of the hydrogen peroxide molecule PES will be further discussed in Subsection 5.1.3.

1.3 Topological features of Potential Energy Surfaces

Potential Energy Surfaces afford basic information for the theoretical description of molecular structures, molecular properties and chemical reactivity. Being a global function of $3N-6$ internal coordinates ($\mathbf{R}$), the construction of a potential surface is a challenging task since *ab initio* calculations must be performed for all the nuclear configurations that are crucial to the conception of an adequate representation of the surface. Taking the hydrogen peroxide as an example, the number of *ab initio* points required to represent a one-dimensional PES cut range between 10 to 30 points for bond distances and 5 to 20 for bond angles, approximately [9]. If we take as a benchmark 10 points for each dimensional cut and considering the six-dimensional configuration space of the system, a total of 10^6 *ab initio* points must be computed to warrant a proper representation of the H_2O_2 PES.

In this context, a topological analysis provides a global overview of the regions of higher importance through the location and characterization of the critical points on a PES as well as helping us on the determination of the paths connecting them.

Critical points, or stationary points, are points at which the gradient, or the first derivatives, of the potential energy function, $V(\mathbf{R})$, with respect to each geometric variable are zero,

$$\frac{\partial V}{\partial R_i} = 0 \qquad \forall_{i=1,\cdots,3N-6}. \tag{1.18}$$

Stationary points may be a minimum, a maximum or a saddle point. The characterization of these points is accomplished by computing the matrix of the second derivatives of the potential energy function. This matrix, also known as the force constant matrix or the Hessian, comprises the elements,

$$\frac{\partial^2 V}{\partial R_i \partial R_j} \qquad i,j = 1,\cdots,3N-6. \tag{1.19}$$

A minimum on a surface is a point where the eigenvalues of the Hessian matrix of the second derivatives are all positive. Physically it means that at a minimum, the potential function is upper convex in all dimensions. There are two types of minima that must be distinguished. A global minimum is unique and corresponds to the point on the PES that has the lowest value of energy, i.e., it is the energetically most stable conformation of the molecule. Local or relative minima are points with the lowest value of energy in a particular section or region of the PES. In the opposite case, there will be a maximum on the PES

if the eigenvalues of the Hessian matrix are all negative and the function is everywhere concave downward. If the Hessian matrix has both positive and negative eigenvalues then we are in the presence of a saddle point. The number of negative eigenvalues characterizes the order of the saddle point. Thus, a saddle point of order k is a maximum in k directions and a minimum in all the other $3N-6-k$ directions. Let's consider, as an example, a first order ($k=1$) saddle point. For this structure there is one and only one negative eigenvalue for the Hessian matrix. This means that a first order saddle point is a maximum along one direction and a minimum for all the other $3N-7$ directions. A saddle point of first order is particularly important in chemical kinetics because only this kind of structure may represent the transition state that connects the reactant valley to the product valleys [7]. Since the PES is a contiguous function, a surface that encloses a total of η minima must have at least $(\eta-1)$ saddle points of first order [10].

The process of finding a stationary point on a PES is called geometry optimization. A wide range for optimization methods for finding stationary points are presently available and well described in literature [6, 10, 11]. Usually, geometry optimizations are nonlinear processes performed by starting with an initial guess for a specific molecular arrangement that it is believed to be as close as possible to a plausible stationary structure, and submit it to an optimization algorithm that continuously changes the geometrical parameters until it has located a stationary point with a finite precision [12].

Chapter 2

Construction of Potential Energy Surfaces

The introduction of the Born-Oppenheimer approximation was indeed a major breakthrough in quantum mechanics by allowing to solve the electronic part of the Schrödinger equation in the presence of a static potential arising from the nuclei in a specific geometry arrangement. Thus, for each nuclear configuration we can compute the correspondent electronic energy and the set of data so obtained is then used to build the so wanted PES.

Nevertheless, even within the Born-Oppenheimer approximation, we are unable to find an exact solution for the electronic Schrödinger equation (except for quite simple systems like H_2^+ and similar one-electron systems). This drives us to the introduction of further approximations able to provide reliable outcomes. A wide range of standard models for the construction of approximate electronic wave functions has been developed since the beginning of Quantum Mechanics. These methods are broadly referred to as electronic structure calculations and can be classified by levels of specialization as *ab initio*, semitheoretical, semiempirical and empirical calculations [7].

Ab initio methods are derived from the resolution of the electronic Schrödinger equation and so, the only approximations introduced are the ones related to the theoretical method adopted and to the level of representation of the wave function. These calculations are the ones that give more reliable results, since their approximations come directly from theoretical principles, without considering any experimental information except for fundamental physical constants, within errors which are known beforehand. Irrespective of their advantages, PESs based on *ab initio* calculations are only achievable for systems with a small number of electrons.

The semitheoretical methods are essentially *ab initio* calculations, used to carry out calculations for systems with a large number of electrons, where specific approximations

are made namely by replacing the inner shell electrons (core electrons) with effective potentials.

Semiempirical quantum chemistry calculations stem from the *ab initio* formalism and use available information, both theoretical and experimental, to construct an analytical functional form for the potential.

Different from the previous approaches, the empirical methods use exclusively experimental reference data to calibrate the functional form.

In the present work, *ab initio* calculations were used exclusively to map the PES for the hydrogen peroxide and so, a synopsis of the basal concepts that underlie the main theoretical methods are presented in the remainder of this chapter.

2.1 *Ab initio* calculations

In general, there are two approximation methods to estimate the electronic energy based on quantum mechanical techniques, the variational methods or supermolecular and the methods based on perturbation theory.

2.1.1 Variational methods

The Rayleigh-Ritz variational method is based on the theorem according to which *given a system whose Hamiltonian operator, $\hat{H}$, is time independent and whose lowest-energy eigenvalue is E_{exact}, if ϕ is any normalized, well-behaved function of the coordinates of the system's particles that satisfies the boundary conditions of the problem, then [13]*

$$E = \frac{\left\langle \Phi | \hat{H} | \Phi \right\rangle}{\langle \Phi | \Phi \rangle} \geq E_{exact}. \tag{2.1}$$

This signifies that within the variational framework the expectation value of the Hamiltonian is an upper bound to the exact ground state energy [14]. *Ab initio* methods based on variational principles differ among themselves on how they define the wave function, ϕ. In order to obtain a good approximation to the exact energy, several parameters of the wave function are optimized so as to minimize the variational integral.

2.1.1.1 The Hartree-Fock Self-Consistent Field method

From all the variational methods, the Hartree-Fock (HF) approach is of particular importance being often the first step toward more accurate methods which include the effect

of electron correlation. Before the description of the basic theory and the mathematical treatment inherent to this method, we must first introduce some important concepts.

As mentioned before, our interest lies on the solution of the electronic part of the Schrödinger equation obtained from the time-independent Schrödinger equation, after the introduction of the Born-Oppenheimer approximation. This electronic hamiltonian in Equation 1.6 depends only on the spatial coordinates of the electrons although, to accurately describe an electron, we must also account for an intrinsic spin coordinate. There are two orthonormal spin functions conventionally labelled as α and β, meaning spin up and down, respectively. These functions constitute a complete set for describing the spin of an electron. Denoting a generic spin coordinate by ω, we can combine the spatial and spin coordinates of an electron and simply represent them as $x = \{r, \omega\}$. This wave function is commonly referred to as a spin orbital. By doing so, the wave function for an N-electron system that describes, for each electron, both the spatial and spin coordinates can be written as $\Psi(x_1, x_2, \cdots, x_N)$.

The existence in the electronic Hamiltonian of an electron-electron repulsion term, $\frac{1}{r_{ij}}$, makes the electronic Schrödinger equation inseparable for each electron without further approximations. One way to approach this problem is to consider, as a first step, that the electron-electron repulsion term can be neglected or substituted by any function depending only on the coordinates of one of the electrons. By starting with this assumption, the Hamiltonian would be separable and the total electronic wave function, describing the motion of an N-atom system, would be composed by a simple product of one-electron spin orbital wave functions,

$$\Psi^{HP}(x_1, x_2, \cdots, x_N) = \prod_{k=1}^{N} \chi_k(x_k) = \chi_1(x_1)\chi_2(x_2)\cdots\chi_k(x_N). \tag{2.2}$$

This product is known as the Hartree product, where each electron is independently described by its own spin orbital. Despite being convenient, the Hartree product holds one important drawback, it doesn't account for the indistinguishability of electrons. By giving electrons an identity, it fails to satisfy the antisymmetry principle, which states that *a many electron wave function must be antisymmetric with respect to the interchange of any set of space-spin coordinates of any two electrons*,

$$\Psi(x_1, x_2, x_3, \cdots, x_N) = -\Psi(x_1, x_3, x_2 \cdots, x_N). \tag{2.3}$$

The minus sign of the equation forces the wave function to vanish if both electrons have the same values for all their quantum numbers. A direct consequence of the antisymmetry

principle requirement is the Pauli exclusion principle that states that *no two electrons can occupy the same spin orbital*. One way to enforce the antisymmetric principle is to write the overall wave function as linear combinations of Hartree products, which for a many electron system can be represented as,

$$\Phi(x_1, x_2, \cdots, x_N) = (N!)^{-\frac{1}{2}} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \cdots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \cdots & \chi_k(x_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \cdots & \chi_k(x_N) \end{vmatrix}, \tag{2.4}$$

also known as a Slater determinant, where the factor $\frac{1}{\sqrt{N!}}$ is the normalization constant. In this determinant all the elements of a given line are related to the same electron while the elements of a specific column are all associated to the same spin orbital.

Exploiting the properties of determinants we may state that switching two lines of the determinant, which is equivalent to exchange the coordinates of two electrons, leads to a change of sign of the determinant. This fulfils the requirement of the antisymmetry principle. On the other hand, if two columns are identical the determinant will be zero. Thus, we may say that if two electrons occupy the same spin orbital, the wave function will vanish. This representation of the wave function also satisfies the Pauli exclusion principle.

At this point, we are prepared to underline the foundations of the Hartree-Fock method. In this approach the total electronic wave function for a many electron system is characterized by a single, antisymmetric Slater determinant made of one spin orbital per electron of the form of Equation 2.4, and the electron-electron repulsion is treated in an average way. Within this method the "best" wave function corresponds to the set of spin orbitals that minimize the variational value of the electronic energy,

$$E_{elec} = \frac{\left\langle \Phi_{elec} | \hat{H}_{elec} | \Phi_{elec} \right\rangle}{\langle \Phi_{elec} | \Phi_{elec} \rangle}. \tag{2.5}$$

The variational optimization is performed through the application of Lagrange multipliers constrained to the orthonormality of the spin orbitals, $\{\chi_i | \chi_j\} = \delta_{ij}$. This yields the Hartree-Fock equations, which determine the individual optimal spin orbitals,

$$f(i)\chi(x_i) = \varepsilon_i \chi(x_i) \tag{2.6}$$

where $f(i)$ is the Fock operator,

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N} \frac{Z_A}{r_{Ai}} + \upsilon(i)^{HF}. \tag{2.7}$$

The two first terms of the Fock operator are the kinetic energy and potential energy for attraction to the nuclei experienced by the single electron i. The third term represents the average potential energy of ith electron due to the presence of the other $n-1$ electrons, named the Hartree-Fock potential, and is represented by,

$$\upsilon(i)^{HF} = \sum_{u} \mathcal{J}_u(i) - \mathcal{K}_u(i), \tag{2.8}$$

where the sum is over all spin orbitals $u = a, b, c, \cdots$. Considering that electron 1 is arbitrary assigned to spin orbital χ_a and electron 2 to spin orbital χ_u, the Coulomb ($\mathcal{J}_u$) and the Exchange operators ($\mathcal{K}_u$) are defined as follows,

$$\mathcal{J}_u(1)\chi_a(1) = \left\{ \int \chi_u^*(2)(r_{12})^{-1}\chi_u(2)dx_2 \right\} \chi_a(1) \tag{2.9}$$

$$\mathcal{K}_u(1)\chi_a(1) = \left\{ \int \chi_u^*(2)(r_{12})^{-1}\chi_a(2)dx_2 \right\} \chi_u(1). \tag{2.10}$$

Since the Hartree-Fock operator $f(i)$, which generates the one-electron spin orbital $\chi(x_i)$ itself, depends on the spin orbitals of all the other $n-1$ electrons, an iterative method must be adopted to solve the Hartree-Fock equations. Therefore, the algorithm starts with an initial guess for the spin orbitals and use them to compute the average field. The value of the average field is then used to solve Equation 2.6 to yield a new set of spin orbitals which in turn will be utilized to construct an improved Fock operator. This procedure is repeated until self-consistence is reached, hence the reason that Hartree-Fock is also named the Self-Consistent Field (SCF) approach.

The orthonormal and optimized spin orbitals gathered from the Hartree-Fock SCF procedure are arranged in order of increasing energy value ε_i. The lowest energy spin orbitals are called the occupied orbitals and the Slater determinant formed by these orbitals corresponds to the Hartree-Fock ground state wave function for the molecule. The other unoccupied spin orbitals are named virtual orbitals [14].

In order to make the Hartree-Fock equations computationally more easily tractable, Roothaan [15] proposed the application of a known set of M basis functions, θ_j, to expand

each spin orbital χ_i by means of a linear combination of these functions,

$$\chi_i = \sum_{j=1}^{M} c_{ij}\theta_j, \tag{2.11}$$

being the linear coefficients, c_{ij}, determined in the self-consistent process. The introduction of this basis set expansion into Equation 2.6 transforms the Hartree-Fock equations into the Hartree-Fock-Roothan equations and the calculation of the expansion coefficients can be accomplished by means of the matrix equation $Fc = Sc\varepsilon$, where F is the Fock matrix resulting from the representation of the Fock operator in the finite basis functions and S is the overlap matrix resulting from the overlap between the basis functions [6].

Due to the approximations inherent to the Hartree-Fock method, if a complete set of basis functions (infinite large basis set) were employed, the energy obtained would not correspond to the exact energy of the system but to an upper energy value denominated as the Hartree-Fock limit. In practice, we are constrained to the use of finite bases and the more flexible and complete these bases are, the closer we will be to the Hartree-Fock limit. The difference between the exact solution of the nonrelativistic Schrödinger equation and the limiting Hartree-Fock energy is the correlation energy,

$$E_{\text{corr}} = E_{\text{exact}} + E_{\text{HFlimit}}. \tag{2.12}$$

However, it should be pointed out that by using an antisymmetrized wave function to describe the system, the Hartree-Fock method introduces the contribution of the exchange correlation or Fermi correlation between electrons of parallel spin. Since the motion of electrons with different spins remains uncorrelated, the single determinantal wave function is simply referred to as an uncorrelated wave function [14].

The Hartree-Fock calculations performed with sufficient large basis are capable of recovering approximately 99% of the exact total energy of a system. However, the remaining error, which represents the correlation energy, is still very large for the accurate description of chemical phenomena [3].

This is more easily seen if we pick up, as an example, the energies computed in this work for the equilibrium geometry of the H_2O_2 molecule. The total energy for the system computed with the Hartree-Fock method gives $E_{\text{HF}} = -150.85089\,E_{\text{h}}$ compared to $E_{\text{CCSD(T)}} = -151.42920\,E_{\text{h}}$ obtained with the CCSD(T) method, both with an aug-cc-pV5Z basis set (a detailed discussion of this correlation method and basis set will be given in 2.1.3 and 2.2.1, respectively), which gives a correlation energy of $-0.57831\,E_{\text{h}}$. From these results, we may deduce that the correlation energy is approximately 0.38% of the

CCSD(T) energy.

Now, let us focus on the interaction energy. At the Hartree-Fock level, the interaction energy, given by the difference between the total energy of the supermolecule and the total energy of the correspondent fragments (two oxygen and two hydrogen atoms), is $-0.22631\,E_\mathrm{h}$ and at the CCSD(T) level is $-0.42578\,E_\mathrm{h}$. The difference between both energies is $-0.19948\,E_\mathrm{h}$, representing about 46.85% of the CCSD(T) interaction energy, which corresponds to an error of $125.17\,\mathrm{kcal\,mol^{-1}}$. Since we aim to achieve chemical accuracy in our surface, which has been established to be a root mean square deviation of $1\,\mathrm{kcal\,mol^{-1}}$, the inclusion of electron correlation is fundamental.

2.1.1.2 Configuration Interaction method

The Hartree-Fock method is based on the approximation that a poly-electronic wave function is represented by an antisymmetrized product, a single Slater determinant, of mono-electronic spin orbitals. However, there are endless ways to represent a Slater determinant since we have a all set of possible configurations for an N-electron wave function. Thus, at the conceptual level, the simplest way to introduce the correlation into the wave function is to represent it as a linear combination of Slater determinants. The basic idea is to use the virtual orbitals of the Hartree-Fock wave function to represent the excited configurations of the molecular system, followed by the optimization, by means of the variational method, of the coefficients of all the Slater determinants included in the representation of the wave function. This *ab initio* method based on the Hartree-Fock model is known as the Configuration Interaction (CI) method.

By taking the Hartree-Fock wave function ϕ_0 that is given by the lowest occupied orbitals as a reference, it is then possible to specify the other determinants according to the number of electrons that are promoted from the occupied orbitals to the virtual ones. Therefore, the total molecular wave function Ψ may be expressed as,

$$\Psi = c_0\phi_0 + \sum_i c_{Si}\phi_{Si} + \sum_i c_{Di}\phi_{Di} + \sum_i c_{Ti}\phi_{Ti} + \sum_i c_{Qi}\phi_{Qi} + \cdots\,, \tag{2.13}$$

where the c's are the expansion coefficients. The subscripts S, D, T, Q stand for single, double, triple, quadruple excited determinants respectively.

In the CI approach, the expansion of the wave function as a linear combination of the configuration functions only include those functions that have the same symmetry as the fundamental state, this way leading to a significant reduction of the number of determinants to be computed. A configuration interaction calculation that includes all the terms of Equation 2.13 with the proper symmetry for a given finite basis set is called

a full CI. For a specific basis set, the full CI approach is the best that we can achieve, constituting itself a standard with which all the other approximation methods should be compared with. The difference between the energy provided by a full CI and the Hartree-Fock energy, both obtained within the same basis set, gives the basis set correlation energy. As the basis set approaches completeness, the basis set correlation energy gets closer to the exact correlation energy [14].

Even with the proper symmetry restriction, the number of configuration state functions (CSF) increases considerably with the number of electrons as well as the size of the basis set which makes a full CI computationally impractical except for small molecules and basis sets. To make this theoretical method feasible one must resort to the truncation of the CI expansion for the wave function, which is referred to a limited CI, restricting the calculation to a limited number of excitations. A common approach is to confine the calculations by considering only the contributions of the single and double excitations out of the Hartree-Fock reference configuration as they are expected to be the ones to make the largest contributions, although with different weights, to the wave function. This level of calculation is typically referred to as Configuration Interaction with Single and Double excitation (CISD).

One must have in mind that the selection of determinants to be used in Equation 2.13 in order to keep the CSFs at a workable size has a major drawback, truncated CI calculations do not have the property of size consistency. This may lead to a wrong description of the PES since the energy of a system AB, when the distance between the subsystems A and B tends to infinity, may not coincide with the sum of the energies of A and B separately computed with the same approximation. This problem of lack of size consistency does not apply to a full CI calculation which is size consistent [16].

2.1.1.3 Multiconfiguration SCF method

To recover a considerable amount of the correlation energy by means of a configuration interaction method, a high number of configurations must be considered, which is extremely expensive from the computational point of view. An alternative *ab initio* approach is the Multiconfiguration Self-Consistent-Field (MCSCF) method. In this method the wave function is constructed as a linear combination of configuration state functions. These are followed by the optimization not only of the expansion coefficients in Equation 2.13, but also of the forms of all the molecular orbitals in the CSFs through the variation of the expansion coefficients of Equation 2.11 in a double iterative process of calculation. The simultaneous optimization restricts considerably the length of the MCSCF expansion, but brings an important advantage comparatively to the configuration interaction

method since it allows to obtain a similar level of accuracy using a smaller number of configuration state functions [16].

The major difficulty with the MCSCF methods is how to construct the wave function in order to make it small enough to be computationally tractable. A successful approach was to consider a specific orbital space and perform a full CI in it, being this the concept of the Full Optimized Reaction Space (FORS), most commonly known as Complete Active Space Self-consistent Field (CASSCF) [6].

The idea is to divide the orbital space into three subspaces, inactive, active and virtual orbitals [17]. The inactive orbitals are the spin orbitals that are doubly occupied in all configurations; the set of virtual orbitals are composed by the spin orbitals that are unoccupied in all configurations, and the active orbitals are the spin orbitals that do not have restrictions on their occupations. After this partition of the molecular orbitals, the wave function is constructed by performing a full CI in the set of active orbitals with all the proper symmetry adapted configurations included in the optimization. The space spanned is denominated as the reference space [1, 18].

The various multireference methods differ in the way used to partition the molecular orbital space as well as by the level of occupation of the active orbitals. An important question that arises is which orbitals to include in the active space since the number of configurations increases rapidly with the number of active orbitals. It takes particular importance the CASSCF method, in which the active space matches the valence orbitals also known by FV CASSCF (Full Valence Complete Active Space SCF). The major difficulty of these methods is to properly choose the active space, due to size restrictions. A large active space increases severely the computational time making the calculations in some cases impossible to carry out.

An alternative approach widely accepted as an *ab initio* method that yields the most accurate potential surfaces is the Multireference Configuration Interaction (MRCI) method. This method combines the MCSCF and the CI methods by generating a reference space optimized at the MCSCF level and by promoting electrons from that reference space into the virtual space to produce further configurations. The most common method of this type is the MRCISD (Multireference Configuration Interaction with Single and Double excitations), in which the reference wave function includes single and double excitations CSFs. Thus, if the reference wave function contains single and double CSFs and further single and double excitations from the reference CSFs are performed, the final MRCISD wave function will also include some triple and quadruple excitations [13].

Despite the ability to recover a considerable amount of the correlation energy, the number of configurations and coefficients to optimize in a MRCI calculation grows very fast with the number of reference configurations. This bottleneck has been the spur of the

development of different contraction schemes in order to reduce the number of coefficients to optimize, making the MRCI calculation less demanding [19].

2.1.1.4 Partition of the correlation energy

At this point it becomes important to improve our understanding regarding the nature of the electron correlation. As already mentioned, the electron correlation is defined by the difference between the exact nonrelativistic energy of the electronic Schrödinger equation, E_{exact}, and the basis limit energy of the Hartree-Fock energy, E_{HF}, as expressed in Equation 2.12.

In turn, two types of effects contribute to the correlation energy, the dynamical correlation E_{dyn}, and the static correlation E_{stat},

$$E_{\text{corr}} = E_{\text{dyn}} + E_{\text{stat}}. \tag{2.14}$$

The static correlation, also called nondynamical correlation energy, arises from near-degeneracies of the bonding and antibonding configurations while the dynamic correlation arises from the Coulomb repulsion between electrons. Systems where the static correlation energy covers a considerable fraction of the total correlation energy are said to have a multireference character.

An illustrative example, widely referred to in literature, in which the static correlation energy takes particular importance, is the dissociation of the H_2 molecule. For this system, as the bond length increases the HOMO (σ_g) and the LUMO (σ_u) orbitals become degenerate, the HF method fails to provide a good description of the homolytic H-H bond dissociation [20]. The incorrect behaviour of this approach is consistent for any bond breaking process that involves the evolution into open-shell fragments. To attain the static component of the correlation energy it is necessary to resort to an *ab initio* method which yields a wave function composed of all configuration state functions obtained from all possible occupancies of the valence molecular orbitals. As we have already seen previously, FV CASSCF fulfil this requirement and so it is the method of choice for evaluating the nature of the nondynamical correlation energy.

The dynamical correlation is subsequently treated by constructing further configurations resulting from the promotion of electrons out of the reference space up to a given excitation level (e.g. all single and double excitations, and so on) and it is defined as,

$$E_{\text{dyn}} = E_{\text{exact}} - E_{\text{FV CASSCF}}, \tag{2.15}$$

where the E_{exact} is the exact solution of the nonrelativistic Hamiltonian and $E_{\text{FV CASSCF}}$

is the energy obtained with the FV CASSCF method.

2.1.2 Many-Body Perturbation theory

A different systematic approach for finding the correlation energy of molecular systems is the Perturbation Theory (PT). The basic idea of the perturbation methods is to divide the Hamiltonian operator into two parts, an unperturbed zero-order part symbolized as $\hat{H}^{(0)}$, from which it is known the eigenfunctions and eigenvalues, and a perturbation part connoted as $\hat{H}'$ leading to the following expression for the total Hamiltonian,

$$\hat{H} = \hat{H}^{(0)} + \hat{H}'. \tag{2.16}$$

The exact solution to the Schrödinger equation is attained by the sum of an infinite number of contributions resulting from an order-by-order expansion of the wave function and the energy. The basis premise behind this method is that the contributions from the perturbed system are significantly small comparing to the soluble part of the Hamiltonian and they can be seen as mild corrections gradually applied to the unperturbed system. Mathematically, this can be perceptible by the introduction of a parameter λ into the Hamiltonean,

$$\hat{H} = \hat{H}^{(0)} + \lambda\hat{H}', \tag{2.17}$$

which gives information related to the magnitude of the perturbation.

The most common approach to the perturbation theory, widely referred to in the literature, is the Rayleigh-Schrödinger perturbation theory for non-degenerate stationary states. Considering $\Psi_n^{(0)}$ as the wave function of the unperturbed system and Ψ_n the perturbed wave function obtained after the application of the perturbation to the reference (unperturbed) system, and letting the $E_n^{(0)}$ and E_n be the correspondent energies associated to the eigenfunctions $\Psi_n^{(0)}$ and Ψ_n, respectively, the perturbed Schrödinger equation becomes

$$\hat{H}\Psi_n = (\hat{H}^{(0)} + \lambda\hat{H}')\Psi_n = E_n\Psi_n. \tag{2.18}$$

If $\lambda = 0$, then $\hat{H} = \hat{H}^{(0)}$, $\Psi_n = \Psi_n^{(0)}$ and $E_n = E_n^{(0)}$ and the previous equation is reduced to a zeroth order equation. As the value of λ increases, a perturbation is introduced to the reference wave function and energy and then, both the wave function and the energy can be expressed as an expansion in powers of the perturbation parameter

λ,

$$\begin{aligned} E_n &= E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \lambda^3 E_n^{(3)} + \cdots \\ \Psi_n &= \Psi_n^{(0)} + \lambda \Psi_n^{(1)} + \lambda^2 \Psi_n^{(2)} + \lambda^3 \Psi_n^{(3)} + \cdots . \end{aligned} \tag{2.19}$$

The $\Psi_n^{(1)}$, $\Psi_n^{(2)}$, $\cdots$ and $E_n^{(1)}$, $E_n^{(2)}$, $\cdots$ are the first-, second-, *etc.* order corrections to the wave function and energy, respectively. Introducing the previous expression for the E_n and Ψ_n into Equation 2.18 and equating coefficients of like powers of λ, we obtain the following set of equations,

$$\hat{H}^{(0)}\Psi_n^{(0)} = E_n^{(0)}\Psi_n^{(0)} \tag{2.20}$$

$$\hat{H}^{(0)}\Psi_n^{(1)} + \hat{H}'\Psi_n^{(0)} = E_n^{(0)}\Psi_n^{(1)} + E_n^{(1)}\Psi_n^{(0)} \tag{2.21}$$

$$\hat{H}^{(0)}\Psi_n^{(2)} + \hat{H}'\Psi_n^{(1)} = E_n^{(0)}\Psi_n^{(2)} + E_n^{(1)}\Psi_n^{(1)} + E_n^{(2)}\Psi_n^{(0)}, \tag{2.22}$$

and so on. The first-order energy correction $E_n^{(1)}$ is obtained by multiplying Equation 2.21 by $\Psi_m^{(0)*}$ followed by integration over all space, which yields,

$$\left\langle \Psi_m^{(0)} | \hat{H}^{(0)} | \Psi_n^{(1)} \right\rangle - E_n^{(0)} \left\langle \Psi_m^{(0)} | \Psi_n^{(1)} \right\rangle = E_n^{(1)} \left\langle \Psi_m^{(0)} | \Psi_n^{(0)} \right\rangle - \left\langle \Psi_m^{(0)} | \hat{H}' | \Psi_n^{(0)} \right\rangle . \tag{2.23}$$

Using the fact that the unperturbed Hamiltonean $\hat{H}^{(0)}$ is Hermitian,

$$\left(E_m^{(0)} - E_n^{(0)}\right) \left\langle \Psi_m^{(0)} | \Psi_n^{(1)} \right\rangle = E_n^{(1)} \left\langle \Psi_m^{(0)} | \Psi_n^{(0)} \right\rangle - \left\langle \Psi_m^{(0)} | \hat{H}' | \Psi_n^{(0)} \right\rangle . \tag{2.24}$$

Since the unperturbed eigenfunctions are orthogonal and normalized, $\left\langle \Psi_m^{(0)} | \Psi_n^{(0)} \right\rangle = \delta_{mn}$, and assuming the perturbed wave function to be intermediately normalized, $\left\langle \Psi_n^{(0)} | \Psi_n \right\rangle = 1$, we have for $m = n$ that,

$$E_n^{(1)} = \left\langle \Psi_n^{(0)} | \hat{H}' | \Psi_n^{(0)} \right\rangle , \tag{2.25}$$

which expresses the first-order energy correction as an average of the perturbation operator over the unperturbed wave function. To find the first order correction to the wave function, $\Psi_n^{(1)}$ can be expanded in terms of the complete set of unperturbed eigenfunction $\Psi_m^{(0)}$,

$$\Psi_n^{(1)} = \sum_m a_m \Psi_m^{(0)}. \tag{2.26}$$

Using $a_m = \left\langle \Phi_m^{(0)} | \Phi_n^{(1)} \right\rangle$ in Equation 2.23 for $m \neq n$, we obtain the following expression

for the coefficients

$$a_m = \frac{\left\langle \Psi_m^{(0)} | \hat{H}' | \Psi_n^{(0)} \right\rangle}{E_n^{(0)} - E_m^{(0)}}, \tag{2.27}$$

where the division by $E_n^{(0)} - E_m^{(0)}$ is only possible because of the non-degeneracy of the energy levels, i.e., $E_n^{(0)} - E_m^{(0)} \neq 0$. Combining Equation 2.26 and Equation 2.27 we get for the first-order wave function correction,

$$\Psi_n^{(1)} = \sum_{m \neq n} \frac{\left\langle \Psi_m^{(0)} | \hat{H}' | \Psi_n^{(0)} \right\rangle}{E_n^{(0)} - E_m^{(0)}} \Psi_m^{(0)}. \tag{2.28}$$

Similarly, one can obtain analogous equations for the second-order corrections to the energy and wave function starting from the expression which equates the coefficients of λ^2 terms in Equation 2.22. The complexity of these expressions gradually increases for higher-order corrections [1, 6].

To obtain the correlation energy by means of perturbation theory we need first to select the zero-order Hamiltonian operator being the most common choice the Hartree-Fock Hamiltonian. This partition of the Hamiltonian constitutes the foundations of the Møller-Plesset (MP) perturbation theory. The extension of the perturbation theory to multiconfigurational electronic systems led to the adoption of an MCSCF wave function as the zero-order wave function, being the most commonly used for this purpose the CASSCF wave function. The inclusion of energy corrections up until $E_n^{(2)}$ provides the CASPT2 (Complete Active Space Second-order Perturbation Theory) method [18].

The MP calculations, truncated at any order, have the advantage of being size-consistent. However, these calculations are not variational and, therefore, they may not yield energies that are upper bounds to the exact energy [1].

2.1.3 Coupled-Cluster theory

The Coupled-Cluster (CC) theory is widely recognized in the area of quantum chemistry as one of the most successful theoretical formulations to construct accurate many electron molecular wave functions, increasingly becoming an efficient strategy to address a diversity of chemical problems. The coupled-cluster fundamental equation represents a nonlinear exponential parametrization of the wave function,

$$\Psi_{CC} = e^{\mathbf{T}} \phi_0, \tag{2.29}$$

where ϕ_0 is the ground state Hartree-Fock wave function and the exponential operator $e^{\mathbf{T}}$ is expanded in a power series as,

$$e^{\mathbf{T}} = 1 + \mathbf{T} + \frac{\mathbf{T}^2}{2!} + \frac{\mathbf{T}^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{\mathbf{T}^k}{k!}, \tag{2.30}$$

being the cluster operator $\mathbf{T}$ composed of a series of connected operators,

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_n, \tag{2.31}$$

where n represents the number of electrons in the molecule. Each of the previous operators acts on the HF reference wave function ϕ_0 introducing single ϕ_i^a, double ϕ_{ij}^{ab}, triple ϕ_{ijk}^{abc}, etc. excitations into the CC wave function. Generally speaking the $e^{\mathbf{T_i}}$ operator acts on ϕ_0 generating all ith excited Slater determinants which, for the case of the one-particle excitation operator $\mathbf{T_1}$ and the two-particle excitation operator $\mathbf{T_2}$, are represented as,

$$\mathbf{T}_1\phi_0 = \sum_i \sum_a t_i^a \phi_i^a \quad \textit{and} \quad \mathbf{T}_2\phi_0 = \sum_{i<j} \sum_{a<b} t_{ij}^{ab} \phi_{ij}^{ab}, \tag{2.32}$$

where ϕ_i^a corresponds to single excitations from the occupied spin orbitals, represented by the indices i and j, to the unoccupied spin orbitals, represented by the indices a and b; being the other operators defined in a analogous way. For single excitations, an amplitude, designated by t_i^a, is associated with each excitation. The excitation operators, $\mathbf{T_1}$, $\mathbf{T_2}$, $\cdots$, always promote electrons from the set of occupied Hartree-Fock spin orbitals to virtual ones; being the result equal to zero when the excitation operation is performed on determinants containing only virtual spin orbitals. In the exponential expansion of the CC wave function also appear the so called disconnected terms such as $\frac{\mathbf{T}_1^2}{2}$, $\frac{\mathbf{T}_2^2}{2}$, $\mathbf{T}_1\mathbf{T}_2$ and so on, which are defined as,

$$\frac{1}{2}\mathbf{T}_1^2\phi_0 = \sum_{\substack{i,a \\ j,b}} t_i^a t_j^b \phi_{ij}^{ab} \tag{2.33}$$

$$\frac{1}{2}\mathbf{T}_2^2\phi_0 = \sum_{\substack{i>j,a>b \\ k>l,c>d}} t_{ij}^{ab} t_{kl}^{cd} \phi_{ijkl}^{abcd}$$

$$\mathbf{T}_1\mathbf{T}_2\phi_0 = \sum_{\substack{i,a,a>b \\ k>l,c>d}} t_i^a t_{kl}^{cd} \phi_{ijk}^{abc} \tag{2.34}$$

$$\cdots . \tag{2.35}$$

When we look more carefully into the previous expressions one may note that, for instance, the disconnected term $\frac{\mathbf{T}_2^2}{2}$ introduces into the CC wave function quadruple excitations with amplitudes determined as products of just double excitation coefficients. This brings a considerable simplification into the calculation of the excitation coefficients when compared with the connected $\mathbf{T_4}$, with a single excitation coefficient describing a simultaneous interaction of four electrons, since in the first case we have $\approx n_0^2 n_u^2$ coefficients (n_0 and n_u represent the numbers of occupied and unoccupied orbitals in a molecular basis set) instead of $\approx n_0^4 n_u^4$ associated with $\mathbf{T_4}$ [21].

The goal of the CC wave function is to determine the values of the amplitudes associated with the respective excitations. This requires for the coupled-cluster wave function to satisfy the Schrödinger equation,

$$\hat{H} e^{\mathbf{T}} \phi_0 = E e^{\mathbf{T}} \phi_0, \tag{2.36}$$

by projecting the coupled-cluster Schrödinger equation onto the Hartree-Fock reference wave function. The multiplication of this equation by ϕ_0, and by the ϕ_μ (generic notation for single ϕ_i^a, double ϕ_{ij}^{ab}, and so on, excitations) followed by integration gives respectively [18],

$$\langle \phi_0 | \hat{H} | e^{\hat{T}} \phi_0 \rangle = E \langle \phi_0 | e^{\hat{T}} \phi_0 \rangle = E \tag{2.37}$$

and

$$\langle \phi_\mu | \hat{H} | e^{\hat{T}} \phi_0 \rangle = E \langle \phi_\mu | e^{\hat{T}} \phi_0 \rangle. \tag{2.38}$$

Up until now the solution is exact and the consideration of all cluster operators in T yields all possible excited determinants, which recover the FCI wave function. When a truncation into the CC wave function is performed, for instance, considering the CCD (Coupled-Cluster Doubles method) approximation where $\hat{T} \approx \hat{T}_2$, the CCD energy is obtained from,

$$\langle \phi_0 | \hat{H} | e^{\hat{T}_2} \phi_0 \rangle = E_{CCD} \langle \phi_0 | e^{\hat{T}_2} \phi_0 \rangle \tag{2.39}$$

and Equation 2.38 becomes

$$\langle \phi_{ij}^{ab} | \hat{H} | e^{\hat{T}_2} \phi_0 \rangle = \langle \phi_0 | \hat{H} | e^{\hat{T}_2} \phi_0 \rangle \langle \phi_{ij}^{ab} | e^{\hat{T}_2} \phi_0 \rangle. \tag{2.40}$$

For each unknown amplitude t_{ij}^{ab} there is one Equation 2.38 and the whole set of these nonlinear equations are solved iteratively [1].

Several approximations can be introduced with the CC theory depending on how many terms are included in $\hat{T}$. The enclosure of only $\hat{T}_2$ excitations gives the Coupled-Cluster Doubles (CCD) method. The truncation at this level contains all the even ordered

excitation determinants (double, quadruple, hextuple, etc) by virtue of the successive application of the pair operators. The contribution of the single excitations operator $\hat{T}_1$ yields the Coupled-Cluster Singles and Doubles (CCSD) method. The CCSD wave function contains the contribution from all the determinants of the full CI wave function, being the coefficients associated with the highly excited determinants obtained as products of the coefficients of the lower order clusters [18].

More accurate results can be obtained by the gradual introduction of higher order excitations operator in $e^{\hat{T}}$ as in the case of the CCSDT model, which includes the contribution of triple excitations $\hat{T}_3$. Although highly accurate, coupled-cluster models that include excitations higher that the double ones are only affordable for small molecular systems due to the computational cost involved. A very common way to address this problem consists in the application of hybrid methods in which perturbation corrections are introduced into the coupled-cluster approach. An example of such methods is the CCSD(T), where the contribution of triple excitations is calculated using perturbation theory.

The CC methods have the advantage of being size-extensive but their application fails to provide good results when the reference function, ϕ_0, fails to provide a good representation of the wave function as in the case of systems with degenerate or nearly degenerate electronic configurations [18].

2.2 Basis sets

The vast majority of quantum chemistry calculations start with the choice of a finite set of basis functions used to build the molecular orbitals for a system. Commonly, these analytical functions are centred on each atom within a molecule being this the reason why they are also called atomic basis functions. The molecular orbitals ψ_i are then generated as a linear combination of these atomic orbitals (LCAO-MO),

$$\psi_i = \sum_{\mu}^{n} c_{\mu i}\chi_{\mu}, \tag{2.41}$$

where ψ_i is the ith molecular orbital, $c_{\mu i}$ are the coefficients of the expansion whose optimal values are settled by applying the variational principle; χ_μ are the μth atomic orbital and n is the number of the atomic orbitals. Thus, the quality of a molecular calculation is intimately related to the basis function chosen to perform it. Ideally, three criteria should be considered in seeking for a suitable set of basis functions: systematic extension towards completeness, rapid convergence, and the analytical form should be

easy to integrate. The fulfilment of all of these requirements are difficult to achieve and compromises must be made [18].

One of the two types of basis functions used to carry through electronic structure calculations are the Slater-type orbitals (STOs) [22] with general definition as follows,

$$\psi_{nlm}^{STO}(r,\theta,\phi) = Nr^{n-1}e^{-\zeta r}Y_{lm}(\theta,\phi), \tag{2.42}$$

where N is the normalization constant, n represents the principal quantum number, r is the distance of the electron to the atomic nucleus and ζ is a constant related to the effective charge of the nucleus. Despite showing the correct shape to perform the expansion of the molecular orbitals, since they have a cusp behaviour when $r \to 0$ and decay with $e^{-\zeta r}$, the STOs have a major drawback. The evaluation of the two-electron integrals on three and four centres leads to a severe increase of the computational time when we perform *ab initio* calculations on polyatomic molecules. To speed up the computation of the molecular integrals, Boys [23] proposed the introduction of a basis function with a Gaussian dependence $e^{-\alpha r^2}$ instead of an exponential form. The GTOs spherical harmonic functions assume the form

$$\psi_{\alpha lm}^{GTO}(r,\theta,\phi) = N_{\alpha l}^{GTO}r^{2n-2-l}e^{-\alpha r^2}Y_{lm}(\theta,\phi), \tag{2.43}$$

where α is the orbital exponent, $N_{\alpha l}^{GTO}$ is the normalization constant and l is the angular momentum quantum number. The suggestion of substituting the STOs by GTOs functions in the basis set expansion comes from the Gaussian product rule. This rule states that the product of two Gaussian, each located at different centres, may be written as a single Gaussian function centred at a point on the line segment that joins the two centres [24]. This important property of Gaussian functions simplify significantly the calculations since the complicated bi-electronic integrals on three or four different atomic centres are reduced to integrals over two different centres. Despite this enormous advantage, from the analytical point of view, GTOs have two important handicaps that have a marked influence on the results. First, a Gaussian function does not have the desired cusp when $r = 0$ and second, it shows a more abrupt decay at large distances by comparison with a Slater function. To minimize the incorrect behaviour of the GTOs orbitals and to achieve results comparable, in terms of accuracy, to those obtained with the STOs functions, Contracted Gaussian-type Orbitals (CGTOs) [25, 26] were developed. In this approach a normalized linear combination of a few GTO's, G_α centred on the same atom and with different values of α's is used instead of an individual one. Each GTO, g_ν, used

in the contraction is called a primitive GTO's,

$$G_\alpha = \sum_{\nu=1}^{\alpha} d_\nu g_\nu, \tag{2.44}$$

where d_ν are the contraction coefficients. For the construction of a CGTO basis, it is then necessary to specify the set of primitive Gaussian exponents and their corresponding contraction coefficients so as to lead to Gaussian basis functions that have the desired quality. For instance the contracted basis functions may attempt to approximate the shape of the Slater-type orbitals. Once optimized the exponents and consequently the contraction coefficients, their values remain unchanged when the basis is applied to molecular calculations [14].

The simplest type of basis sets composed of contracted Gaussian are the STO-nG, where n is the number of primitive GTOs used in the contraction procedure. The most common choice is to consider $n = 3$, which leads to a CGTOs named STO-3G. In this typical basis set a STO is approximated as a linear combination of 3 Gaussian functions, where the exponents and coefficients of the primitive Gaussians are chosen so that the basis functions approximate, by a least-squares fit, to Slater functions [27].

The smallest number of functions per atom needed to express the occupied atomic orbitals of that atom is called minimal basis set. Considering as an example the H_2O molecule, a minimal basis set will be formed by seven functions, one $1s$ basis function on each hydrogen and one basis function each for the $1s$, $2s$, $2p_x$, $2p_y$, $2p_z$ orbitals of the oxygen atom [16]. The STO-nG bases are an example of such minimal basis sets. One important issue to take into consideration is that these minimal basis sets are very small if we aim at accurate results, meaning that more extensive basis sets are demanded.

An obvious improvement upon the minimal basis set is to replace each basis function in the minimal set by two basis functions differing in their orbital exponents, which is called a double-zeta (DZ) basis set. The number of basis functions doubled and consequently the number of independently determined variational coefficients, $c_{\mu i}$, in Equation 2.41 is twice the number of coefficients in a minimal basis set wave function. Considering the water molecule example, using a DZ basis set we will have a total of 14 basis functions, two $1s$ functions on each hydrogen atom, two $1s$ functions, $2s$ functions, two $2p_x$ functions, two $2p_y$ functions and two $2p_z$ functions on the oxygen atom. Further improvements can be achieved by adding continuously more basis functions with different orbital exponents to the minimal basis set leading to triple-zeta (TZ), quadruple-zeta (QZ), quintuple-zeta (5Z) basis sets and so on.

As the number of basis functions increase the computational calculations become more

demanding. In recognition to the fact that the contribution from the inner shell electrons is not expressive to most of the chemical properties of a molecule, a single function may be used to represent each core atomic orbital and only the valence atomic orbitals are represented by two (or more) basis functions. A basis set of this type where a minimal basis set is used to represent the inner shells and the outer orbitals are expressed by a double-zeta (or triple, or quadruple, or $\cdots$) basis set is called a Split-valence (SV) basis set. The $4-31$G [28] is an example of SV basis set. The acronym means that the core orbitals are described by a contraction of 4 primitive GTO, while the valence orbitals are split into two functions, one contracted Gaussian of three primitive gaussians and another one given by a unique primitive GTO. In accordance with the number of functions used to describe each valence orbital, the SV basis sets may be valence double-zeta (VDZ), valence triple-zeta (VTZ), etc. In the case of $4-31$G, it is define as a split-valence double-zeta basis set.

The next step to upgrade a basis set is to introduce polarization functions. As atoms are brought together upon molecule formation their centres of charge are shifted originating a distortion in the shape of the atomic orbitals. The atoms are said to be polarized. To account for the polarization effect one adds basis functions which accommodate orbitals with angular momentum, l, higher than the valence shell of the ground state atom. Examples of polarization functions are p-type functions in the case of the hydrogen atom and d-type functions for the first row atoms Li-F. In the $6-31$G* basis set the asterisk stands for the addition of d-type functions on each first row atom into the valence double-zeta $6-31$G basis set. A second asterisk to yield $6-31$G** means the introduction of one set of three p-type functions on each hydrogen atom [29].

Further improvements in accuracy of calculations for compounds that enclose an eminent electron density in regions far away from the nuclei (e.g. anions, hydrogen-bonded compounds, etc.) can be accomplished by the addition into the basis sets of diffuse functions. These functions are functions with very small orbital exponents and are usually referred to by the notation "+". One single "+" sign such as in $6-31+$G signifies the introduction of a set of s and p-type functions to the heavy atoms and a "++" means that a highly diffuse s-type function is also added to each hydrogen atom [30].

2.2.1 Correlation-consistent basis functions

For correlated calculations such as CI, it is mandatory to introduce more stringent requirements into the basis sets mainly because the virtual orbital space, needed to describe the excitations, must be capable of recovering a large amount of the correlation energy.

Nowadays, the state of the art for correlated calculations is the application of the

so called correlation-consistent (cc) basis sets proposed by Dunning [31]. In these basis sets each correlating orbital is represented by a single primitive Gaussian, in which the exponent is optimized in order to maximize its contribution to the correlation energy. A hierarchy of correlation-consistent atomic basis sets can then be established by adding groups of correlating orbitals that have similar contributions to the correlation energy in a simultaneous way. These contracted Gaussian basis functions are collectively denoted as cc-pVXZ, meaning correlation-consistent polarized valence basis sets, where the X is the cardinal number which represents the hierarchy of the basis set. If X = D, T, Q, $\cdots$ the basis sets will be of a double-, triple- or quadruple-zeta quality, respectively.

Considering the first row atoms, the smallest set of split-valence plus polarization type cc-pVDZ, contains the functions [3s2p1d]; the triple-zeta basis cc-pVTZ contains the functions [4s3p2d1f]; the cc-pVQZ contains the functions [5s4p3d2f1g], and so on. So, we may say that each basis supplements the previous one by successively adding one set of functions with maximum angular momentum $l_{max} = X$ (except for the H and He atoms). As an example, the largest angular momentum in a cc-pVQZ basis function is g and so, $l_{max} = 4$ and $X = 4$. One must note that these basis functions are split-valence basis sets especially designed to perform calculations of the valence correlation energies. These bases neither do have enough flexibility to describe the correlation energy of the core electrons, nor do have to describe species with diffuse electronic structures [18]. To account for the correlation concerning the core electrons, further functions with higher exponents are added yielding the correlation-consistent polarized core-valence basis sets, cc-pCVXZ. By doing so, the standard cc-pVDZ basis set is extended by a set of [1s1p] core-correlated orbitals leading to the formation of the cc-pCVDZ set with the structure [4s3p1d] [32]. The addition of diffuse functions to the cc-pVDZ set originate the augmented correlation-consistent polarized valence basis sets, aug-cc-pVXZ, adequate for correlation calculations of systems with diffuse electron distributions such as anions [33].

In the present work, the *ab initio* calculations were carried out using Dunning's basis sets of the aug-cc-pVXZ family with X = T(3), Q(4) and 5 levels of quality. This choice allowed us to adopt a extrapolation/scaling scheme which will be explained in detail in Section 4.3.

2.2.2 Basis Set Superposition Error

The interaction energy of a molecular system, obtained through electronic structure calculations, is affected by a spurious contribution named Basis Set Superposition Error (BSSE). This error arises from the fact that the energy of each constituent fragment is improved by the presence of the basis set of the other fragment within the supermolecule,

when compared with the energy computed from the single fragments using its own basis set. In other words, this effect is solely assigned to the incomplete description of the fragments due to the use of finite basis sets. It is more pronounced for smaller basis sets and vanishes asymptotically when the complete basis set is approached. It becomes apparent that without correction, the value of the interaction energy will suffer from an artificial increase, which may lead to an unrealistic characterization of the PES [34].

There are a few possible ways to overcome the BSSE. The most evident way is to use a saturated basis set. However this approach is too much expensive in terms of computational time mainly when we aim to perform an accurate study of a PES for a considerable large system. Another alternative treatment is the counterpoise (CP) correction scheme proposed by Boys and Bernardi [35]. The basic idea is to compute the energies of the isolated fragments in the full basis of the supermolecule and then subtract them from the energy of the entire system. Thus, if we consider the complex AB, the counterpoise corrected interaction energy is given by,

$$\Delta E_{\mathrm{AB}}^{CP} = E_{\mathrm{AB}} - E_{\mathrm{A}}^{\mathrm{CP}} - E_{\mathrm{B}}^{\mathrm{CP}} \tag{2.45}$$

where E_A^{CP} is the energy of the fragment A computed in the presence of ghosts orbital from fragment B, and similarly for E_B^{CP}. Considering that the uncorrected interaction energy of the AB complex is,

$$\Delta E_{\mathrm{AB}} = E_{\mathrm{AB}} - E_{\mathrm{A}} - E_{\mathrm{B}}, \tag{2.46}$$

the counterpoise correction to the interaction energy is,

$$\Delta E_{\mathrm{corr}}^{\mathrm{CP}} = \Delta E_{\mathrm{AB}}^{\mathrm{CP}} - \Delta E_{\mathrm{AB}}. \tag{2.47}$$

Substituting in this equation the expressions for ΔE_{AB}^{CP} and ΔE_{AB} we obtain:

$$\Delta E_{\mathrm{corr}}^{CP} = (E_{\mathrm{A}} - E_{\mathrm{A}}^{\mathrm{CP}}) + (E_{\mathrm{B}} - E_{\mathrm{B}}^{\mathrm{CP}}). \tag{2.48}$$

If we use the variational method to construct the wave functions, the value of the $\Delta E_{\mathrm{corr}}^{\mathrm{CP}}$, which gives an estimate of the BSSE effect, will always be positive since $E_{\mathrm{A}} > E_{\mathrm{A}}^{\mathrm{CP}}$ and $E_{\mathrm{B}} > E_{\mathrm{B}}^{\mathrm{CP}}$ [18].

Despite the clear advantage of using the counterpoise method to correct for the BSSE, we must be aware that it is believed that this correction may overestimate the BSSE due to the fact that a fragment uses orbitals of the other fragments that should be unavailable since they are occupied. This overcorrection is particularly important for small basis sets

and disappears more rapidly than the BSSE for larger basis sets. Since we are using a medium to large basis sets we thing that the application of the CP correction is the best way to correct for the BSSE.

Chapter 3

Representation of Potential Energy Surfaces

In the previous chapters the concept of Potential Energy Surface was introduced as well as the methods used to calculate it. Once the calculations of the electronic energies for a broad and representative range of geometric configurations of the nuclei are collected, the next step is to develop a global analytical function. The great challenge lies on the construction of a functional form capable of matching the data set of *ab initio* energies within chemical accuracy, in order to carry out dynamics studies. To accomplish these requirements, several criteria should be taken into account.

The analytical function must be consistent with the symmetry properties of the molecular system under consideration, i.e., it should reproduce exactly the same amount of energy when two identical atoms are permuted; it should appropriately define the asymptotic regions and smoothly connect them with the interaction regions; it should describe the potential accurately in regions where experimental and theoretical information is available; and it should present a reasonable physical behaviour in regions of the configuration space where no potential information is available. These prerequisites are mandatory to achieve an accurate analytic representation of the PES, although we can find in the literature the allusion to many other desirable criteria that a PES must fulfil [36].

When we impose the condition that the functional form must have the ability to represent distinct characteristics for different regions of the surface, another important issue that emerges is the flexibility of the functional form used to represent the PES. The higher the flexibility the higher the number of parameters to be optimized, which may lead to convergence problems. Furthermore, if complex functions are used, additional problems may arise in the derivation process since it may imply the resolution of very complicated algorithms. We should reach a compromise concerning the structure of the

functional form. On the one hand it should be simple enough, so that it can be used in the resolution of dynamical problems and, on the other hand, it must have enough flexibility to reproduce accurate results.

Diverse methods for representing PESs have been developed for many years. Different approaches are reviewed in literature [37–39]. Among the most popular ones is the Many-Body Expansion (MBE) method proposed by Murrell and co-workers [40], later extended by Varandas in the Double Many-Body Expansion (DMBE) method [7, 41, 42]. In the present work the construction of the analytical potential for the ground state of the hydrogen peroxide is based on the DMBE framework, thereby a more detailed discussion of the MBE and DMBE formalisms is given in the following sections.

3.1 Many-Body Expansion method

The MBE method offers a general strategy for the global representation of PESs. It represents the PES in terms of a summation of all the potential energy contributions of each sub-part of the total polyatomic system. This approach, first introduced by Murrell and co-workers [40], states that the total potential for the interaction of N-atoms can be written in the form,

$$V_{A,B,\ldots,N(\mathbf{R})} = \sum_{A} V_A^{(1)} + \sum_{AB} V_{AB}^{(2)}(R_{AB}) + \sum_{ABC} V_{ABC}^{(3)}(R_{AB}, R_{AC}, R_{BC}) + \cdots + V_{A,B,\ldots,N(R)}^{n}(\mathbf{R}), \tag{3.1}$$

where R are the interatomic distances. The summations of all one-body terms represented by $V_A^{(1)}$ correspond to the energy of the isolated atoms. The first sum of Equation 3.1 is performed over the energy of all n-atoms, being omitted if the dissociation limits of the surface represents the atoms in their ground state. $V_{AB}^{(2)}(R_{AB})$ represents the two-body interactions related with the diatomic AB isolated from the other atoms. In many PESs these two-body potentials have been represented by extended Rydberg potentials. These two-body terms will tend to zero as the internuclear distance between the diatomic AB approaches infinity.

The three-body potential, $V_{ABC}^{(3)}(R_{AB}, R_{AC}, R_{BC})$, depends on the shape and size of the triangle ABC and its energy is achieved by computing, for each nuclear configuration, the total electronic energy from which are subtracted the energies of the lower order terms. The analytical function that represents a three-body potential is fitted, usually by a least-squares technique, to reproduce the energies obtained from that subtraction. The same procedure is applied to higher order terms.

The asymptotic condition pointed out for the two-body potentials also applies to all

the n-body terms presented in Equation 3.1, i.e., if one of the internuclear coordinates in a specific n-body species is removed to infinity, then the value of the potential energy function of that fragment tends asymptotically to zero. A functional form that fulfils the previous imposition, widely used to describe the three-body and higher terms, is expressed as a product between a polynomial and a switching function. The polynomial represents the short-range interactions of the potential and the switching function guarantees that the polynomial vanishes at longer distances where the potentials of the fragments are the relevant ones.

We can highlight some important advantages on describing PESs through the MBE method. First, the lower order terms of the expansion can be approximately described by means of simpler functional forms and second, if these terms are accurately determined, the resulting PES will include all the correct dissociation limits, no matter how poorly the nth term is described, which is an essential requirement for molecular dynamics studies [43]. On the other hand, the expansion in energy terms of subclusters of atoms has the inconvenience that, while the terms of the expansion increase, the weight of the higher order terms also increases.

3.2 Double Many-Body Expansion method

Within the Double Many-Body Expansion (DMBE), proposed by Varandas [44], the terms of the many-body expansion defined in equation 3.1 are split into *extended* Hartree-Fock (EHF), which includes the nondynamical correlation related to rearrangements of the electrons by degenerate or nearly degenerate valence orbitals, and the dynamic correlation (dc) which results from the correlation of the electronic motion. For a N-atom system, the PES has the form,

$$\begin{aligned} V_{A,B,\ldots,N(\mathbf{R})} &= \sum_{A}\left[V^{(1)}_{A;EHF}+V^{(1)}_{A;dc}\right]+\sum_{AB}\left[V^{(2)}_{AB;EHF}(R_{AB})+V^{(2)}_{AB;dc}(R_{AB})\right]+ \\ &\quad \sum_{ABC}\left[V^{(3)}_{ABC;EHF}(R_{AB},R_{AC},R_{BC})+V^{(3)}_{ABC;dc}(R_{AB},R_{AC},R_{BC})\right]+ \\ &\quad V^{n}_{A,B,\ldots,N(R);EHF}(\mathbf{R})+V^{n}_{A,B,\ldots,N(R);dc}(\mathbf{R}). \end{aligned} \tag{3.2}$$

In the same way as for the MBE method, also for the DMBE approach each term of Equation 3.3 with order higher than one must vanish asymptotically when any one of the atoms belonging to a specific cluster is removed to infinity. The partition of the interaction energy into EHF and dynamic correlation contributions offers an important advantage over the MBE method. It enables to model, separately, each contribution to

the total energy and to describe, through different functional forms, the correct asymptotic behaviour of the EHF and *dc* terms. The DMBE formalism generally uses *ab initio* methods, such as SCF calculations, or a function parameterized from available spectroscopic data, to define the $V_{EHF}^{(n)}$ terms that describe the short-range interactions. On the other hand, the $V_{dc}^{(n)}$ terms, which are dominant at long distances, are semiempirically approximated from the dispersion energy coefficients, employing dumping functions to account for orbital overlap and electron exchange effects [7, 44].

3.3 Molecular PES Least-Squares Fitting and Single Value Decomposition

After the collection of a considerably large amount of discrete highly accurate potential energy points, the next step is the formulation of a mathematical model that better fit the data set in order to obtain a good description of its shape. Among the methods for fitting potential energy functions, the method of least-squares is one of the most commonly used. Let us consider a general system composed of n-atoms. The mathematical model that we are looking for, that is, the PES, is simply an analytical function that gives the relationship between the energy of the n-atoms system and its geometry. In other words, for each nuclear configuration denoted by $\mathbf{z_i}$, there is a potential energy value, E_i, associated with it. Depending on the number of measurements a total of np sets of observations ($\mathbf{z_1}$, E_1), ($\mathbf{z_2}$, E_2), $\cdots$,($\mathbf{z_{np}}$, E_{np}) will be available to perform the fitting procedure, where $\mathbf{z_i}$ is the independent variable and E_i is the correspondent dependent variable. If we use the bond length representation to represent uniquely our n-atoms molecule, the potential fitting function, $V(\mathbf{z})$, will depend on the nuclear coordinates $V = V(\mathbf{z})$, where $\mathbf{z} = (r_1; r_2; r_3; \cdots; r_N)$, being N the number of degrees of freedom of the bond length representation, $N = \frac{n(n-1)}{2}$. The model function of the potential may assume a wide variety of algebraic forms. Polynomial functions are one of the most commonly adopted functions for the representation of $V(\mathbf{z})$ mainly due to simplicity in the resolution of the derivatives, which is important for the purpose of dynamical calculations but also because it can incorporate sufficient complexity in order to obtain a reliable representation of the PES. Therefore, a PES can be represented by a nth degree multivariate polynomial functions with the form,

$$f(\mathbf{z}, \beta) = \sum_{a=0,b=0,c=0,\cdots}^{a+b+c+\cdots \leq n} \beta_{abc\cdots} {r_1}^a {r_2}^b {r_3}^c \cdots {r_N}^m = \sum_k \beta_k P^k(z_i), \tag{3.3}$$

where $\beta_k P^k(z_i)$ represents each monomial of the polynomial functions and k is a cumulative variable and gives the order of the monomial in the polynomial.

The method of the least-squares fitting defines the "best" fit function through the minimization of the sum, S, of the deviations squares, known as the residuals. A residual is defined as the difference between the values of the energy of the *ab initio* point and the analytical function model, $d = E_i - f(z_i;\beta)$. Thus, the sum of squares of the deviation from the true polynomial function for a given set of data points $\{(\mathbf{z_i},\ E_i)\}_{i=1}^{np}$, can be expressed as,

$$S = d_1^2 + d_2^2 + \cdots + d_{np}^2 = \sum_{i=1}^{np} d_i^2 = \sum_{i=1}^{np} [E_i - f(z_i;\beta)]^2 = \text{a minimum.} \tag{3.4}$$

The goal is then to obtain a function model $f(z_i;\beta)$, parametrically dependent on the β parameters, which holds the values for β that produce the least possible value of S. To do so, we must differentiate Equation 3.4 with respect to each parameter β and set each result equal to zero,

$$\frac{\partial S}{\partial \beta_k} = -2 \sum_{i=1}^{np} [E_i - f(z_i;\beta)] \frac{\partial f(z_i;\beta)}{\partial \beta_k} = 0;\ j = 1, \cdots, nc \tag{3.5}$$

with nc adjustable parameters, being nc less than the number of data values np. Since these minimization conditions are linear combinations of the coefficients,

$$f(z_i;\beta) = \sum_{k=1}^{nc} \beta_k P^k(z_i), \tag{3.6}$$

where $P^k(z_i)$ is the polynomial basis function represented in Equation 3.3, we have that,

$$\frac{\partial f(z_i;\beta)}{\partial \beta_k} = P^k(z_i) \tag{3.7}$$

and so, Equation 3.5 may be expressed as,

$$\sum_{i=1}^{np} [E_i - f(z_i;\beta)] P^k(z_i) = 0. \tag{3.8}$$

Substituting Equation 3.6 in Equation 3.8 we have,

$$\sum_{i=1}^{np} P^k(z_i) E_i = \sum_{i=1}^{np} P^k(z_i) \sum_{i=1}^{nc} \beta_k P^k(z_i). \tag{3.9}$$

These equations are known as the normal equations related to the least-squares problem and it can be expressed in a matrix format as,

$$Z^T E = Z^T Z\beta. \tag{3.10}$$

Rearranging the equation we obtain a unique solution,

$$\hat{\beta} = (Z^T Z)^{-1} Z^T E \tag{3.11}$$

for the estimator $\hat{\beta}$, of the parameters β, which minimizes the sum of squares in Equation 3.4 [45].

One important aspect related to the application of the least-squares method concerns the number of the normal equations. If this number is too large, the matrix of the normal equations tends to be ill-conditioned, i.e., the solution of the system becomes numerically unstable.

Within the normal equation formulation, the columns of $Z^T Z$ must be linearly independent, i.e., nonsingular in order to be an invertible matrix. The inversion of $Z^T Z$ can lead to large numerical problems, due to the limited precision of the computational calculations. To better understand the cause of this numerical instability we must define what the condition number of a matrix is. The condition number of the matrix $Z^T Z$ is given by the ratio,

$$cond(Z) = \frac{\lambda_{Max}}{\lambda_{Min}}, \tag{3.12}$$

where λ_{Max} and λ_{Min} are the largest and the smallest eigenvalues of $Z^T Z$. When the eigenvalues of the matrix have a high discrepancy in magnitude, the difference between the largest order eigenvalue and the lowest order eigenvalue is large and the contribution of the lowest one is so small that $Z^T Z$ becomes "almost" singular. The higher the condition number is, the higher the effect of truncation is on the precision of the computational calculations when we attempt to solve linear systems.

One of the most used method to solve ill-conditioned problems is the Single Value Decomposition (SVD). In the SVD approach a rectangular $m \times n$ real matrix Z, with $m \geq n$ can be factorized into a product $U\Sigma V^T$, where U is a $m \times m$ orthogonal matrix, V is a $n \times n$ orthogonal matrix and Σ is a diagonal matrix with the same dimensions as Z and whose diagonal elements σ satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. This positive nonzero diagonal entries , σ_i, labelled in a decreasing order of magnitude are called single values of Z. The columns of U and V are named the left and the right singular vectors respectively, for Z. Using the SVD method, the solution to the linear system $Z\beta = E$ is given by $\hat{\beta} = V\Sigma^{-1}U^T E$.

By definition the singular values $\sigma_1, \cdots, \sigma_n$ of a $m \times n$ matrix Z are the square roots, $\sigma_i = \sqrt{\lambda_i} > 0$ of the eigenvalues of the symmetric $n \times n$ matrix $Z^T Z$. In terms of condition numbers, this means that the condition number of Z is the square root of the condition number of $Z^T Z$. This definition implies that by applying the SVD approach we obtain a gain in precision which brings more stability in the estimation of the least-squares parameters [46–48].

Part II

Potential Energy Surface of the $H_2O_2(X\ ^1A)$ molecule

Chapter 4

Ab initio energies

4.1 General description

The hydrogen peroxide molecule, H_2O_2, has attracted widespread attention due to their overwhelming importance in combustion [49], atmospheric [50–52] and interstellar [53] chemistry. This system plays an important role in the HO_x chemistry, which in turn participates in the stratospheric ozone depletion cycle [54]. Its thermal dissociation, $H_2O_2 + M \rightarrow 2OH + M$, is one of the reactions responsible for the chain-branching at the third explosion limit of hydrogen-oxygen system [49, 55]. The interest in this molecule has also grown in astrochemistry sustained by its presence in the icy Galilean moons of Jupiter mainly Europa, Ganymede and Callisto [56], on comets, on molecular clouds and on small solar system objects such as Kuiper Belt Objects (KBOs) as a result of the irradiation of the water ice by charged particles and ultra violet (UV) photon irradiation [57, 58].

From the theoretical point of view the hydrogen peroxide has quite few peculiar features. At the equilibrium structure, this chiral nonplanar molecule with C_2 symmetry, is among the simplest molecules presenting a large amplitude motion (internal rotation or torsion) that manifests itself by a rotation of the two O-H moieties around the O-O bond. This stereomutation is described by a torsional potential displaying a small barrier for the *cis-trans* isomerization, which gives rise to interesting vibration-rotation tunneling effects. The floppiness due to internal rotation, which results in large anharmonic vibrational modes, makes the determination of the molecular properties quite difficult to accomplish implying the necessity of an accurate six-dimension potential to carry out spectroscopy studies. Furthermore, the relatively low dissociation energy [59] makes hydrogen peroxide an ideal candidate for experimental research of the dissociation process [60]. Therefore, it is not surprising that, due to its attractiveness, the modelling of *ab initio* potential energy

surfaces for hydrogen peroxide has been the main subject of many studies.

One of the research works that boosted the construction of analytical potentials for the ground state of the hydrogen peroxide was a set of two articles developed by Harding [61,62]. His first paper reported an anharmonic potential for the description of the region close to the hydrogen peroxide minimum determined, based on high-quality GVB+CI calculations performed with a triple-zeta basis set. The topic of the second paper focused on the characterization of the long-range interaction potential between two OH radicals with the recourse to multireference configuration interaction calculations.

Besides the first works of Harding, other authors devoted their studies to the accurate determination of the equilibrium structure as well as to the torsional potential of the H_2O_2 molecule. As an example, we refer to Koput's work [63], in which the *ab initio* molecular parameters for hydrogen peroxide and its two (*cis* and *trans*) conformers were determined using the coupled-cluster level of theory, CCSD(T) and cc-pVXZ Dunning's basis sets, which ranged from double- to quadruple-zeta quality. With the recourse to the same theoretical *ab initio* method combined with a quadruple-zeta basis set, cc-pVQZ, Koput et al. determined the anharmonic quartic force field and used the variational method to calculate the vibrational-rotational energy levels for the molecule [64] and for its various isotopomers [65]. The influence of the core-electron correlation on the molecular parameters and on the torsion potential energy were also studied [64] by employing both cc-pCVTZ and cc-pwCVTZ basis sets. Later, Malyszek and Koput [66] computed an enhanced version of the ground state potential using the conventional CCSD(T) and the explicitly correlated coupled-cluster method, CCSD(T)-F12, in conjunction with the augmented correlation-consistent basis sets, aug-cc-pVnZ, up to septuple-zeta quality. This highly accurate potential was applied to predict the vibrational-rotation energy levels of H_2O_2, D_2O_2 and HOOD molecules [66] as well as highly excited rovibrational energy levels [67].

One of the most extensive studies concerning the electronic ground state potential of hydrogen peroxide was reported by Kuhn et al. [68]. In that work, a semiglobal PES was derived, based on large-scale *ab initio* calculations, conducted with a combination of two computational methods including the density functional theory (DFT) and multiconfiguration second order perturbation theory (CASPT2) calculations, both using correlation consistent cc-pVTZ basis set. The potentials so designed were particularly devoted to the characterization of the regions around the equilibrium structure and the dissociation channel into hydroxyl radicals. This potential surface became a benchmark for a variety of studies namely the dynamics involved in torsional isomerization and stereomutation [69,70], spectroscopic calculations [71–74], dissociation/recombination dynamics of $H_2O_2 \leftrightarrow 2OH$ [75], and parity-violating potential energy hypersurfaces calculations [76].

In a series of two articles Sayós and González modelled a pseudotriatomic analytical representation of the ground potential energy of the $O(^1D)$ + H_2O system, where an OH group of the H_2O molecule was treated as an atom of 17 a.m.u., employing Møller-Plesset methods [77], whose validity was afterwards analysed with CASPT2/CASSCF *ab initio* methods [78]. In both studies, reactants, products, saddle points and minima (including the global minimum of the molecule) were characterized as well as the paths connecting them. The first PES based on perturbation theory served as background to Quasiclassical Trajectory (QCT) dynamics studies of the $O(^1D)$ + H_2O reaction that leads to the formation of 2OH [77, 79] and H + HO_2 [79]. To our interest, we further mention the work of Ge et al. [57], who also studied the PES for the singlet and triplet H_2O_2 system, with special focus on optimized structures for H_2O_2 local isomers and transition states. Several reaction paths and correspondent energy barriers were also evaluated.

The accurate determination of the PES for the hydrogen peroxide, by means of *ab initio* calculations, remains a great challenge since it implies the need not only of high-quality basis sets but also of highly expensive methods capable of recovering a considerable amount of the correlation energy. For this reason, the vast majority of the literature available on this subject, which we have briefly summarized, targets the detail description of selected regions of the PES.

In order to fulfil the need of a global PES for the singlet H_2O_2 system, in the present work we propose a full dimensional analytical potential covering all possible reaction channels with chemical accuracy to achieve trustworthy kinetic data. The constrictions, in terms of computational effort, imposed by the size of the basis set and by the *ab initio* correlation method employed were overcome by the implementation of an extrapolation-scaling procedure which will be explained in detail in section 4.3.

4.2 Electronic structure calculations

Electronic structure calculations were performed after a careful analysis of the theoretical information available for the hydrogen peroxide. These previous studies pointed out a few important aspects. It is well established that for a good description of the equilibrium structure and of the heights of the torsional barriers at the *trans* and *cis* conformations of the hydrogen peroxide, it is important the application of high-quality basis functions which include electron correlation and multiple sets of polarized functions [63, 68]. It was also pointed out that in some regions of the configuration space the system has multireference characteristics and so diffuse functions must be also considered to obtain trustworthy energy barriers [57]. The presence of open shell species of interest highlight

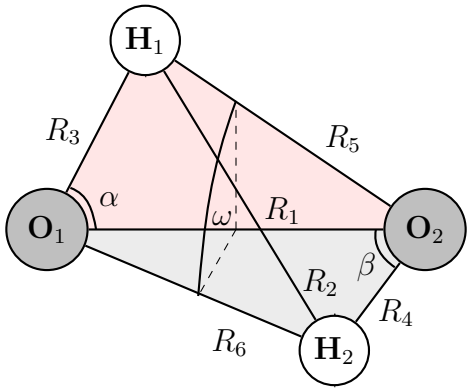


Figure 4.1: Bonds used in the present work for an H_2O_2 system.

the fact that to predict accurate energies for this system multireference methods such as the multiconfigurational second-order perturbation theory (CASPT2) method should be considered [57, 62, 68, 78].

By taking the above features into consideration our first approach to the derivation of the potential hypersurface of the H_2O_2 molecule was to conduct several *ab initio* calculations with different methods using Dunning's aug-cc-pVXZ basis sets. Having in mind that we were looking for the best affordable approach, the test results revealed that a coupled-cluster approach with single and double excitations, including perturbative contributions of connected triple excitations, CCSD(T), performed with an aug-cc-pVTZ basis set, was the one that offered the best compromise between accuracy and cost. Therefore, we decided to map the major part of the configuration space with the CCSD(T)/aug-cc-pVTZ approach. Despite the advantages of this *ab initio* method, CCSD(T) calculations fail to provide a good description of the dissociation processes involving the formation of radical species and, besides, it is difficult to achieve convergence in regions where the system is characterized by degenerate or near degenerate electronic configurations. To overcome the limitation of the CCSD(T) level of theory, a multireference method was used namely, a multiconfiguration second-order perturbation theory, CASPT2, to ensure the accuracy of the results and a suitable description of the asymptotic regions. Yet, even applying the CASPT2 method, we faced difficulties in the computation of the energies in the O approaching H_2O region of the PES.

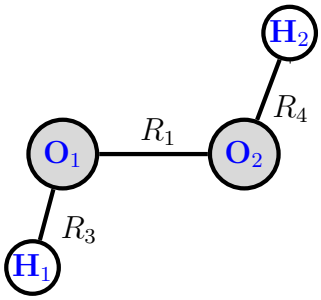
A general description of the PES was obtained by making small variations on all of the six internal coordinates (R_1, R_3, R_4, α, β and ω) represented in Figure 4.1, where ω is the dihedral angle defined as the angle between the two $O_1O_2H_1$ and $O_1O_2H_2$ intersecting planes of the molecule. The six internuclear distances (R_1, R_2, R_3, R_4, R_5 and R_6), which will be further used in this work in section 5.1.3.1, are also represented.

A total of 38440 fixed nuclear configurations were computed. Of these, 29980 points were calculated with the CCSD(T) method and the remaining 8460 points with the CASPT2 method. In the CASPT2 approach the CASSCF is taken as the reference wave function for the calculation of the a second-order perturbation theory. Since the $H_2O_2(X\ ^1A)$ molecular system has a total of 18 electrons and 12 orbitals, the CASSCF zero-order wave function was constructed by considering the lowest two orbitals, the 1s orbitals from the oxygen atoms, as closed shell orbitals, doubly occupied in all configuration state functions (CSFs). This means that the complete active space consists of 14 valence electrons distributed over 10 active orbitals, which yields a total of 4950 CSFs considering a C_1 symmetry. These specifications for the wave function were also considered in the Multireference Rayleigh Schrödinger perturbation theory (RS2) calculations computed to account for the dynamic correlation contributions.

The *ab initio* calculations were obtained through a sequence of steps which allowed us to gradually improve the surface. Initially, we started to obtain optimized geometries for the different minima and transition states previously characterized by González and co-workers [78] at the CASSCF level of calculation with the cc-pVTZ basis set. Our optimizations were conducted at CCSD(T) level of theory with the Dunning's aug-cc-pVTZ basis set, which includes further addition of diffuse functions on the O and H atoms. The optimized structures found for the stationary points are illustrated in Figure 4.2.

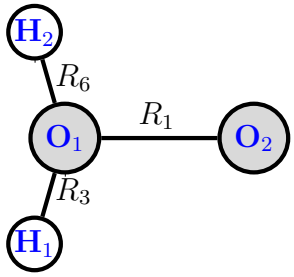
For a better comparison between our results and those from Reference [78], we gave different nomenclatures to the structures but we kept the names adopted by those authors between parenthesis. Thus, in accordance to the previous reference, the H_2O_2 structure corresponds to the absolute minimum of the potential surface; the local minimum corresponds to a minimum structure resulting from a [1,2]-hydrogen shift of hydrogen peroxide via formation of a transition state denoted in Figure 4.2 by isomerization (ts_9); $dihedral_0$ and $dihedral_{180}$ are the *cis* and *trans*- conformers of hydrogen peroxide, respectively; $ts\,H + HO_2$ is a transition state connecting the local minimum to the hydroperoxyl radical; $ts\,O_2 + H_2$ is a transition state representing the evolution from the hydroxyl radical toward the formation of O_2 + H_2; and $ts\,OH + OH$ that corresponds to a transition state leading to the formation of OH + OH from the global minimum.

We point out that we were not able to optimize the stationary point label $ts\,OH + OH$ (ts_2), not even with multireference methods like CASPT2 or MRCI. Due to the importance of this structure in the reaction paths that leads to the formation of two OH radicals, we adopted its geometry from Reference [78]. One way to assess the quality of our *ab initio* calculations is to compare them with experimental data which is only available for the absolute minimum. Table 4.1 lists optimized geometries obtained for the equilibrium



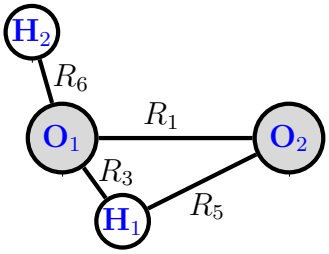
R_1=2.759; R_3=1.826; R_4=1.826;
$\angle O_2O_1H_1$=99.86; $\angle O_1O_2H_2$=99.86;
$\angle H_1O_1O_2H_2$=112.54

absolute minimum (H_2O_2)



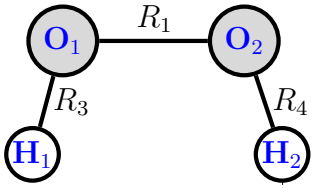
R_1=2.912; R_3=1.832; R_6=1.832;
$\angle O_2O_1H_1$=100.24; $\angle O_2O_1H_2$=100.24;
$\angle H_1O_1O_2H_2$=109.03

local minimum (min_1)



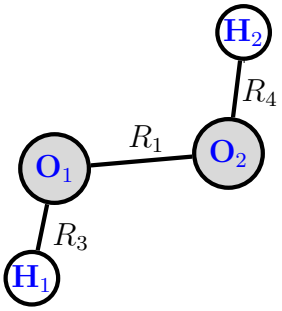
R_1=3.074; R_3=1.962; R_6=1.833
$\angle O_2O_1H_1$=57.87; $\angle O_2O_1H_2$=97.89;
$\angle H_2O_1O_2H_1$=103.80

isomerization (ts_9)



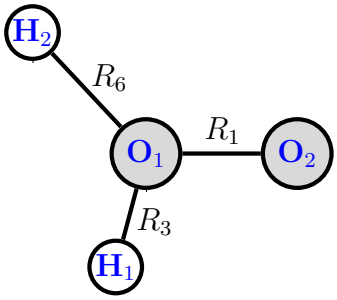
R_1=2.780; R_3=1.827; R_4=1.827
$\angle O_2O_1H_1$=104.13; $\angle O_1O_2H_2$=104.13;
$\angle H_1O_1O_2H_2$=0.00

dihedral$_0$ (ts_4)



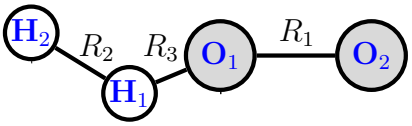
R_1=2.779; R_3=1.825; R_4=1.825;
$\angle O_2O_1H_1$=98.30; $\angle O_1O_2H_2$=98.30;
$\angle H_1O_1O_2H_2$=180.00;

dihedral$_{180}$ (ts_5)



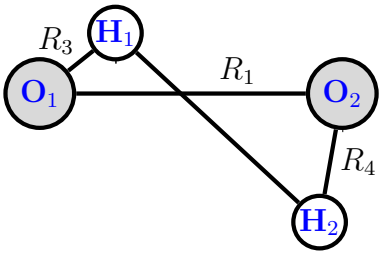
R_1=2.527; R_3=1.846; R_6=3.295;
$\angle O_2O_1H_1$=104.63; $\angle O_2O_1H_2$=122.17;
$\angle H_1O_1O_2H_2$=96.03;

ts H + HO_2 (ts_7)



R_1=2.461; R_2=1.990; R_3=2.083;
$\angle O_2O_1H_1$=111.00; $\angle O_1H_1H_2$=126.94;
$\angle O_2O_1H_1H_2$=104.65;

ts O_2 + H_2 (ts_{13})



R_1=5.503; R_3=1.842; R_4=1.842;
$\angle O_2O_1H_1$=78.49; $\angle O_1O_2H_2$=87.60;
$\angle H_1O_1O_2H_2$=131.11;

ts OH + OH (ts_2)

Figure 4.2: Stationary points (minimum and transition states) of the ground PES of the H_2O_2 system optimized at the CCSD(T) *ab initio* level of theory with an aug-cc-pVTZ basis set. We used as starting points for the optimization process the geometries obtained from Reference [78]. The distances and the angles are given in Bohr and degrees, respectively. Since we were not able to optimize transition state ts_2, the values in the figure were adopted from Reference [78].

Table 4.1: Comparison between theoretically predicted geometries for the equilibrium structure of the hydrogen peroxide and experimental data. Bond distances are in Bohr and angles are in degrees. Bond length and angles are defined in Figure 4.1

.

	CCSD(T)[a]	Koput[b]	Kuhn[c]	González[d]	Ge[e]	Exp. Data[f]
R_1	2.759	2.741	2.744	2.796	2.729	2.751
R_3	1.826	1.820	1.825	1.830	1.822	1.827
α	99.86	100.01	100.40	98.78	100.52	102.32
ω	112.54	112.75	114.00	115.52	111.62	113.70

[a]Optimized geometry calculated in this work at the CCSD(T) level of theory with the aug-cc-pVTZ basis set.

[b]Geometric parameters obtained from: Reference [66] at the CCSD(T) level of theory with the aug-cc-pVQ7Z basis set,

[c]Reference [68] at the DFT level of theory with the cc-pVTZ basis set,

[d]Reference [78] at the CASSCF level of theory with the cc-pVTZ basis set

[e]Reference [57] at the QCISD level of theory with the aug-cc-pVTZ basis set.

[f]Experimental data obtained from Reference [80].

structure of hydrogen peroxide for different *ab initio* methods and experimental data.

The calculations performed in this work demonstrate that the estimated bond distances and angles are in close agreement with those obtained by other authors and with the experimental data. For the bond length between the O-O atoms and the O-H atoms we observe a slight discrepancy, with regard to the experimental results, of about 0.3% and 0.04%, respectively. With regard to the the OOH and the dihedral angles both are underestimated by 2.4% and 1.0%. It is worth noting that the smallest amplitude of variation of the bond lengths in comparison with the variation of the angles are also present for the other theoretical approaches. By comparison with the experimental values, the length of the R_1 distance ranges from 0.3% to 1.6%. Concerning the R_3 bond, the divergence is so small, 0.1% to 0.4%, that we may say that this bond length is well defined. Larger differences can be seen for the α and ω angles which range considerably between $1.8 - 3.5\%$ and $0.3 - 1.8\%$, respectively. Comparing the various *ab initio* calculations, our results are very similar to those obtained by Koput at the CCSD(T) level of theory using a cc-pVQZ basis set and by Ge at the QCISD level of theory using a aug-cc-pVTZ basis set. On the other hand, those estimated by González are those that diverge most from both theoretical and experimental results, which may be explained by the lack of dynamical correlation on those calculations.

After the optimization of the geometries represented in Figure 4.2, several calculations were conducted to map the surrounding configurational space of those structures as well as the different paths that connect them.

Taking these initial points as reference, the next step was to build a preliminary surface in which a whole set of dynamic calculations were performed for all possible reaction channels. These studies enabled us to perform corrections to the PES through the introduction of additional *ab initio* points selected in two ways, first by identification of regions with an abnormal behaviour such as unrealistic wells or energy barriers and second, by randomly choosing configurations adopted by the system when it evolved toward the formation of the products of the chemical reactions that occur on this surface.

Concurrently to this study, a carefully analysis of all the dissociation channels as well as of the torsional potential of the H_2O_2 molecule was also performed by making small displacements of the internal coordinates that intervene the most in the region under consideration. For each of these displacements the value of the potential energy was obtained from the fitted function. Likewise, whenever spurious minima were detected further points were computed and incorporated in the data file to be fitted.

The molecular configurations computed at the CASPT2 *ab initio* level correspond to geometries that cover the regions related to the dissociation channels $OH(X\ ^2\Pi)$ + $OH(X\ ^2\Pi)$, $O(^1D)$ + $H_2O(X\ ^1A_1)$ and $H(^2S)$ + $HO_2(X\ ^2A'')$. The application of the CASPT2 method was limited to specific values of internal coordinates. For geometries positioned along the reaction $OH(X\ ^2\Pi)$ + $OH(X\ ^2\Pi)$ the CASPT2 was confined to $R_{OO} > 4.270\,a_0$. For the O-H bond cleavage leading to the formation of $H(^2S)$ + $HO_2(X\ ^2A'')$, CASPT2 was employed for $R_{OH} > 5.000\,a_0$ and for the last channel, $O(^1D)$ + $H_2O(X\ ^1A_1)$, CASPT2 was circumscribed to $R_{OO} > 4.300\,a_0$. The dissociation process that yields $H_2(X\ ^1\Sigma_g^+)$ + $O_2(a\ ^1\Delta_g)$ was computed with the CCSD(T) method since it does not involve the formation of open-shell species.

The Molpro program package was used to perform all the *ab initio* calculations. All the points were corrected for the Basis Set Superposition Error (BSSE) by means of the counterpoise correction and calculated in C_1 symmetry.

4.3 Extrapolation/scaling procedure

An accurate PES is only achieved through the concomitant application of methods capable of recovering significant amounts of electron correlation and basis sets that approach the complete basis set limit (CBS). Unfortunately, approaching the CBS limit for any particular correlation method is quite difficult. This restriction is based on two important aspects, the first one concerns the fact that the number of bi-electronic molecular integrals in an *ab initio* calculation scales with at least the fourth power of the number of basis functions (n); the second aspect refers to the number of *ab initio* points necessary to

determine the potential energy surface, which grows as X^{3N-6}, where X stands for the number of points required to represent a one-dimensional cut and N is the number of atoms in the $3N-6$ dimensional space of the molecule [7]. Therefore, one can estimate that the total calculation time is proportional to $X^{3N-6} \times n^4$, a very limiting feature for practical applications.

A major breakthrough towards the construction of accurate PESs has become possible after the introduction of the correlation-consistent basis sets of Dunning and co-workers, including the aug-cc-pVXZ basis sets used in this work. These basis sets constitute hierarchical sequences that lead to systematic improvements from level to level, allowing us to approach the basis set limit. This property has been the spur for the development of a wide variety of mathematical expressions to extrapolate the *ab initio* energies toward the complete basis set.

In the present work a multidimensional extrapolation/scaling scheme, combining extrapolation for the basis set and interpolation at the reference geometries, is proposed in order to generate a high-accurate global PES. A more detailed explanation of its construction is given below.

4.3.1 Extrapolation scheme

It is well established that the rate of convergence of the correlation energies is considerably slower than that of the Hartree-Fock energies [18]. The different behaviour of both energies presupposes two important aspects. First, one should treat the HF and correlation parts of the energy separately when using extrapolation techniques and second, the convergence of the correlation energies are the limiting factor in the achievement of accurate results. Therefore, many recent studies on basis set convergence have focused on the convergence acceleration of the correlation energy [81].

Furthermore, as we noted during the derivation of the extrapolation procedure, the variational principle only applies to the total energy of the system. This suggests that we should attempt to extrapolate separately the energy of the supermolecule and fragments instead of the interaction energy.

4.3.1.1 Hartree-Fock energies

The extrapolation of the Hartree-Fock energies towards the CBS limit was based on the following three parameter exponential model [82],

$$E_X^{\mathrm{HF}} = E_\infty^{\mathrm{HF}} + \beta \exp(-\alpha X), \tag{4.1}$$

where the subscript X indicates the cardinal number of the basis sets selected to calculate the *ab initio* points (in our case aug-cc-pVXZ), β and α are parameters to be determined. Hartree-Fock calculations were performed for all nuclear configurations with the aug-cc-pVXZ basis set family for the $X = 3$ and $X = 4$ energy levels. Knowing the values of the HF energies for these two basis sets and considering a value for α, it is possible to solve a linear system of two equations,

$$E_3^{\mathrm{HF}} = E_\infty^{\mathrm{HF}} + \beta \exp(-3\alpha) \tag{4.2}$$

$$E_4^{\mathrm{HF}} = E_\infty^{\mathrm{HF}} + \beta \exp(-4\alpha), \tag{4.3}$$

from which we can obtain the expressions for the extrapolated HF energies and for the linear parameter β:

$$E_\infty^{\mathrm{HF}} = \frac{E_4^{\mathrm{HF}} - E_3^{\mathrm{HF}} \exp(-\alpha)}{1 - \exp(-\alpha)} \tag{4.4}$$

$$\beta = \frac{E_3^{\mathrm{HF}} - E_\infty^{\mathrm{HF}}}{\exp(-3\alpha)}. \tag{4.5}$$

We now refer the value of α for Equation 4.1. Halkier et al. [82] proposed a value of $\alpha = 1.54$. These authors calculated for each combination of $(X-1,\ X)$, which in our case is $(X-1,\ X) = 3,\ 4$, an optimized value of α by minimizing the root-mean-square error of the extrapolated limits for a set of diatomic molecules. Instead of using this value we decided to determine it for each molecular configuration carrying out additional calculations with $X = 5$ at specific reference geometries, minimum, saddle points and asymptotic regions. With these three energies, $X = 3$, $X = 4$ and $X = 5$ in Equation 4.1, a nonlinear system was solved to determine the values of E_∞^{HF}, α and β for each of the chosen geometries. To attain the solution to the nonlinear system, we used, as estimate, the values of E_∞^{HF} and β obtained by the resolution of the linear system, Equations 4.4 and 4.5, and considering the initial value of $\alpha = 1.54$ proposed by Halkier et al. [82].

As mentioned before, to account for the correlation energies two different methods of calculation were applied. The first, CCSD(T), implies a prior HF calculation which is used as the reference wave function and in the second, CASPT2, the reference wave function is given by the precedent CASSCF calculation. As pointed out by Varandas [83], the CASSCF energies, which include the nondynamical part of the correlation energy, can be extrapolated in a similar way as the uncorrelated single-configuration HF energies. Thus, we employed the same expressions, Equations 4.1, Equation 4.4 and Equation 4.5, to estimate the complete basis set limit for our active space energies.

Table 4.2 displays the values of α of the extrapolated basis set limit for HF and

CAS energies of the 16 configurations used as reference geometries. We use as reference geometries the minimum and transition states referred to in Figure 4.2, page 48. In addition we included reference geometries in the dissociation channels $O(^1D) + H_2O(X\ ^1A_1)$, $OH(X\ ^2\Pi) + OH(X\ ^2\Pi)$, $O_2(a\ ^1\Delta_g) + H_2(X\ ^1\Sigma_g^+)$ e $H(^2S) + HO_2(X\ ^2A'')$. Concerning the $O(^1D) + H_2O(X\ ^1A_1)$ dissociation channel, we decided to introduce two reference geometries in the asymptotic limits, one on the oxygen side of the H_2O molecule and another on the hydrogen's side, which are starting points to the oxygen insertion and the hydrogen abstraction mechanisms for this reaction. To better reproduce these mechanisms further intermediate points have been included. These four extra references are labelled in Table 4.2 as H_abstraction intermediate, H_abstraction asymptotic, O_insertion intermediate and O_insertion asymptotic. Similarly, for the $H(^2S) + HO_2(X\ ^2A'')$ reaction channel we introduce one extra intermediate point between the ts $H + HO_2$ reference and this asymptotic limit. This new reference is designated by H_approach.

The values of α, for the supermolecule and for the oxygen fragments, are listed. The energies of the hydrogen atoms were not extrapolated; we used the exact value instead. Due to the small active space of the oxygen atoms, the α values for these fragments are similar for both the HF and CASSCF energies. Then we only used one single column to represent them. Analysing the values of Table 4.2 for the supermolecule, we note that, α exhibits a very slight variation with mean values and standard deviation of 1.4436 ± 0.0037 for the HF and 1.4806 ± 0.0089 and for the CASSCF energies.

The small variation of α suggests that we should attempt to use a fixed value of α by using a mean value from all the 16 reference geometries. This approach simplifies the extrapolation to the complete basis set limit significantly, resulting from the conversion of a nonlinear three parameter extrapolation into a linear two parameter extrapolation scheme, as mentioned by Halkier et al. [82]. However, despite being similar, the α values vary with regard to the molecular configuration and this variation, at some reference geometries, are more pronounced. This aspect dissuaded us from using the mean value since it is not a robust measure being strongly affected by the presence of outliers. Instead, we chose to interpolate the values of α.

Thus, for each *ab initio* point of the PES the correspondent value of E_∞^{HF} was determined by solving the linear system in which the α was obtained by interpolation between the α values of the various reference geometries.

Considering the six-dimension of the PES and the small number of reference geometries, the interpolation formula is given by the geometric average of the α_i weighted by the inverse of the distance between the point and the references according to the following

Table 4.2: Values of α for the various references geometries.

	α_{SCF}	α_{CASSCF}	$\alpha_{\mathrm{SCF/CASSCF}}$
Reference	Supermolecule	Supermolecule	O Fragments
absolute minimum	1.4449	1.4767	1.4751
local minimum	1.4363	1.5824	1.4846
isomerization	1.4334	1.4646	1.4789
dihedral_0	1.4381	1.4687	1.4651
dihedral_180	1.4367	1.4677	1.4813
ts $OH + OH$	1.4423	1.4745	1.4560
ts $O_2 + H_2$	1.4527	1.4635	1.4661
ts $H + HO_2$	1.4516	1.4705	1.4702
H_approach	1.3991	1.4223	1.4618
H_2O + O H_abstraction int.	1.4571	1.4897	1.4686
H_2O + O H_abstraction asy.	1.4500	1.4826	1.4588
H_2O + O O_insertion int.	1.4581	1.4861	1.4645
H_2O + O O_insertion asy.	1.4477	1.4816	1.4578
HO_2 + H	1.4421	1.4591	1.4709
OH + OH	1.4470	1.4718	1.4482
O_2 + H_2	1.4605	1.5277	1.4471

expression,

$$\alpha = \frac{\sum_{i=1}^{n_{ref}} \frac{1}{d_i}\alpha_i}{\sum_{i=1}^{n_{ref}} \frac{1}{d_i}} = \frac{\sum_{i=1}^{n_{ref}} \alpha_i \prod_{j\neq i} d_j}{\sum_{i=1}^{n_{ref}} \prod_{j\neq i} d_j}, \tag{4.6}$$

where d_i is a "distance" between the point and the reference point i, computed using symmetry coordinates explained in the section 5.1.3.1, page 5.1.3.1. Although similar (see Table 4.2), the value of α_i of a reference geometry far from a point still influences the interpolated value. To overcome this problem, we found convenient to expand the "cartesian" distance between the symmetric coordinates of the points. This is accomplished defining the "distances" as

$$d_i = exp\sqrt{(\sum_j (\varphi_{j,i})^2)} - 1, \tag{4.7}$$

where $\varphi_{j,i}$ are "normalized" distances in the symmetric coordinates of the point and the i reference. The subtraction by 1 assures that d_i approaches 0 when a point gets close to a reference geometry i. As it is explained in section 5.1.3.1, the symmetry coordinates used to build invariant polynomials are combinations of interatomic distances with different degrees. To give them equal weight it was necessary to "normalized" those coordinates.

This was accomplished by raising them to the inverse of their degree, so all become terms of first degree, and dividing by the number of bond distances used in their definition.

4.3.1.2 Correlation energies

The method of extrapolation described above is quite useful since HF and CAS energies can be accurately extrapolated to the CBS limit with a significant reduction in computational time. Nevertheless, since a wide range of *ab initio* points must be computed to cover a considerable area of the configuration space and being aware that the correlation energies are the most time consuming part of the calculations, a further reduction in the calculation time can be achieved if the correlation energies are scaled rather than extrapolated [9, 83, 84].

In this work, the correlation energies are treated with a simple scaling procedure from the correlation energies computed at the triple-zeta basis set level of accuracy multiplied by a scaling factor γ according to the following expression,

$$E_{\mathrm{X}}^{\mathrm{corr}} = \gamma E_{\mathrm{X=3}}^{\mathrm{corr}}, \tag{4.8}$$

where, $E_{\mathrm{X=3}}^{\mathrm{corr}}$ can be determined by the difference between the CCSD(T) and HF energies,

$$E_{\mathrm{X=3}}^{\mathrm{corr}} = E_{\mathrm{X=3}}^{\mathrm{CCSD(T)}} - E_{\mathrm{X=3}}^{\mathrm{HF}} \tag{4.9}$$

or, between the CASPT2 and the CASSCF energies,

$$E_{\mathrm{X=3}}^{\mathrm{corr,d}} = E_{\mathrm{X=3}}^{\mathrm{CASPT2}} - E_{\mathrm{X=3}}^{\mathrm{CASSCF}}. \tag{4.10}$$

It should be noted that for the CASPT2 calculations only the dynamical part of the correlation energy is scaled since the static part is included in the CASSCF energy treatment. This dynamical energy is represented in the previous expression by the letter "d" in superscript.

To obtain the values of the scaling factor γ in Equation 4.8, first we must determine the values of γ_i for each reference geometries based on the energies computed for the triple and five zeta basis sets,

$$\gamma_i = \frac{E_{\mathrm{X=5},i}^{\mathrm{corr}}}{E_{\mathrm{X=3},i}^{\mathrm{corr}}}. \tag{4.11}$$

By knowing the γ_i values for the pivot geometries, the γ used to correct the triple-zeta correlation energies at each molecular configuration can be easily estimated using an interpolation formula similar to the one described in Equation 4.6. The application of

this scaling factor determined from basis sets levels of quality $X = 3, 5$ has an important consequence: the scaled correlation energies have a five-zeta quality as opposed to the HF and CASSCF energies that extrapolate to $X = \infty$.

In the dissociation regions of the PES, where experimental data concerning the energy of the fragments is known, a different treatment was applied to perform the interpolation of the γ values. The goal is to warrant the dissociation of the global PES to the known experimental energy of the fragments at their equilibrium geometries due to the use of the many-body expansion formalism explained in the section 3.1. In other words, the total energy of the fragments should be equivalent to a CBS result and the Equation 4.11 may be adjusted to

$$\gamma_i = \frac{E^{\mathrm{corr}}_{\mathrm{X}=\infty,i}}{E^{\mathrm{corr}}_{\mathrm{X}=3,i}}, \tag{4.12}$$

in which $E^{\mathrm{corr}}_{\mathrm{X}=\infty,i}$ is defined by

$$E^{\mathrm{corr}}_{\infty,i} = E^{\mathrm{tot}}_{\mathrm{exp,i}} - E^{\mathrm{M}}_{\infty,i}. \tag{4.13}$$

The letter M stands for HF or CASSCF energies, depending on the *ab initio* method into consideration. The values of the scaling factors determined with the experimental information are 2% lower comparing to the values obtained with the five-zeta level of quality. Furthermore, the distances between the computed *ab initio* points and these dissociative reference geometries are relatively large and so, their contribution to the interpolation procedure is small, mainly correcting the energies of the point close to the dissociation channels. We must emphasize that regarding the two types of *ab initio* methods used, the coupled-cluster approach gives a larger fraction of the correlation energy comparing with the multireference perturbation theory. To account for these differences and to warrant continuity of the overall PES, at the selected pivot references, instead of computing the difference between the CASPT2 and the CAS energies performed with a five-zeta basis set, we estimate the dynamical correlation energy from the CCSD(T) and CAS energies according to,

$$E^{\mathrm{corr,d}}_{\mathrm{X}=5,i} = E^{\mathrm{CCSD(T)}}_{\mathrm{X}=5,i} - E^{\mathrm{CAS}}_{\mathrm{X}=5,i} \tag{4.14}$$

for the five-zeta energies, or

$$E^{\mathrm{corr,d}}_{\infty,i} = E^{\mathrm{tot}}_{\mathrm{exp},i} - E^{\mathrm{CAS}}_{\infty,i} \tag{4.15}$$

for the dissociation fragments. In a way similar to the described for the scaling treatment of the CCSD(T) calculations, the scaling factors, γ_i, are computed at the reference

Table 4.3: Values of γ for the various references geometries.

	$\gamma_{\mathrm{CCSD(T)}}$	$\gamma_{\mathrm{CCSD(T)}}$	γ_{CASPT2}	γ_{CASPT2}
Reference	Supermolecule	O Fragments	Supermolecule	O Fragments
absolute minimum	1.0720	1.0769	1.1358	1.0990
local minimum	1.0732	1.0782	1.1565	1.1005
isomerization	1.0705	1.0784	1.1424	1.1006
dihedral_0	1.0719	1.0770	1.1538	1.0989
dihedral_180	1.0721	1.0769	1.1356	1.0989
ts $OH + OH$	1.0446	1.0818	1.1430	1.1050
ts $O_2 + H_2$	1.0589	1.0787	1.1059	1.1011
ts $H + HO_2$	1.0639	1.0783	1.1190	1.0999
H_approach	1.0505	1.0799	1.1591	1.1027
H_2O + O H_abstraction int.	1.1007	1.0801	1.1489	1.1023
H_2O + O H_abstraction asy.	1.0934	1.0810	1.1415	1.1032
H_2O + O O_insertion int.	1.0757	1.0798	1.1354	1.1019
H_2O + O O_insertion asy.	1.0899	1.0899	1.1367	1.1032
HO_2 + H	1.0795	1.0774	1.1206	1.1002
OH + OH	1.0796	1.0821	1.1241	1.1046
O_2 + H_2	1.0908	1.0802	1.1079	1.1024

geometries by applying the expression,

$$\gamma_i = \frac{E^{\mathrm{corr,d}}_{\mathrm{X}=5,i}}{E^{\mathrm{corr,d}}_{\mathrm{X}=3,i}}. \tag{4.16}$$

Table 4.3 summarizes the γ values computed for the all set of reference geometries. The first two columns represent the values of γ for the supermolecule and oxygen fragments obtained at the CCSD(T) level of theory, and the last two columns give the values recovered with the CASPT2 method. The values of the O fragments are given by the mean value between the results of γ determined for the two oxygen atoms in the molecule. Note that the values concerning the supermolecule and the O fragments at the CASPT2 level of theory are slightly higher than those calculated with the CCSD(T) method since, as already mentioned, in the first ones the dynamical part of the correlation energy was estimated in order to match the correlation energy given by the CCSD(T) at the five-zeta level. The values of γ_i for each point were interpolated using Equation 4.6 in a way similar to the one applied for the α values.

This approach was tested at some points in the transition regions between the CCSD(T) and the CASPT2 calculations. The difference in the extrapolated energies so obtain was less than $1.0 \times 10^{-3}\,E_{\mathrm{h}}$, which is lower than the precision of this work.

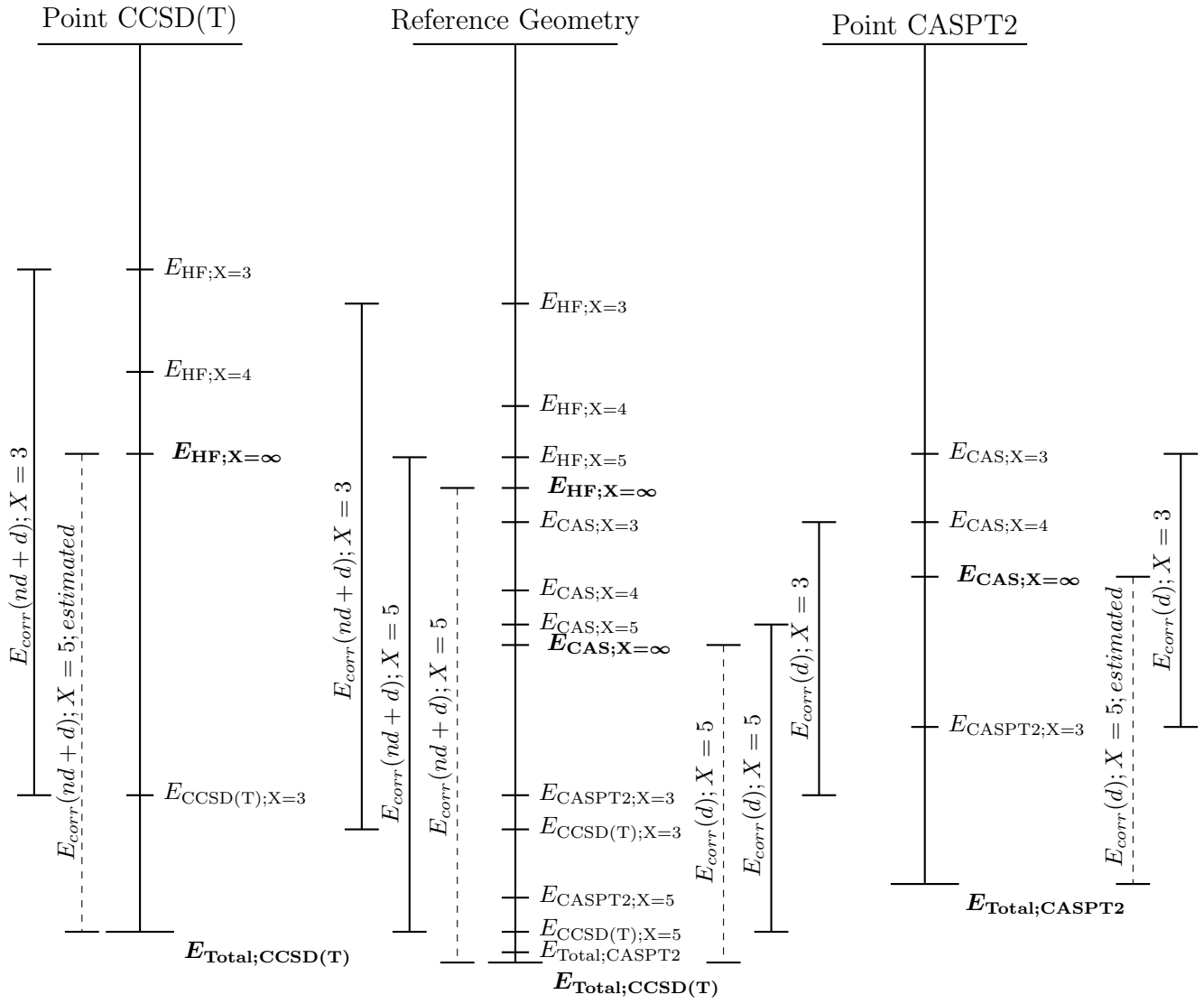


Figure 4.3: Extrapolation/scaling scheme. See text for explanation.

In short, our treatment of the correlation energy aims to extrapolate the triple-zeta energies to the "exact" energies, being the "exact" energies the best that we could obtain with the two calculation methods employed to account for the electronic correlation.

Figure 4.3 attempts to give a better understanding of the extrapolation/scaling scheme applied to accomplish an accurate PES for the H_2O_2 system.

The vertical bar at the centre represents the calculations performed at the reference geometries. For each of these structures, we have computed HF and CASSCF energies at the $X = 3$, 4 and 5 basis set levels. This allowed us to determine the values of α for both energies, quoted in Table 4.2, page 54, for the reference and also to extrapolate the HF and the CASSCF energies toward the CBS limit. The CBS limit energies are

represented in bold by ($E_{\mathrm{HF};\,X=\infty}$) and ($E_{\mathrm{CAS};\,X=\infty}$). The left side of the central bar represents the correlation energies obtained with the CCSD(T) method at the triple-zeta level, ($E_{\mathrm{corr}}(nd+d); X=3$), and at the five-zeta level, ($E_{\mathrm{corr}}(nd+d);\ X=5$). The ratio between the five-zeta and the triple-zeta results gives the corresponding value of $\gamma_{\mathrm{CCSD(T)}}$ for the reference, quoted in columns 2 and 3 of Table 4.3. The dashed line represents the addition of the five-zeta correlation to the CBS HF energy giving the total CCSD(T) estimated energy, in bold at the bottom of the central bar. This is the total energy of our reference points. Similarly, the right side of the central bar corresponds to the correlation energies computed with CASPT2 method in the reference point. As explained before, the dynamical correlation energy at the five-zeta level was estimated by making the difference between the CCSD(T) and the CASSCF energies at the same five-zeta level, ($E_{\mathrm{corr}}(d);\ X=5$). The dashed line on the right represents the addition of this energy to the CBS CASSCF energy, giving the energy, $E_{\mathrm{Total;\,CASPT2}}$, close to that obtained in the CCSD(T) treatment. The γ_{CASPT2} value for the reference is then given by the ratio between ($E_{\mathrm{corr}}(d);\ X=5$) and ($E_{\mathrm{corr}}(d);\ X=3$), quoted in columns 4 and 5 in Table 4.3. Although represented in the central line, the CASPT2 five-zeta energy of the reference is not used in this extrapolation procedure.

The vertical bar at the left side represents the extrapolation procedure for CCSD(T) energy points for which we have computed HF and CCSD(T) energies with the aug-cc-pVTZ basis set as well as HF energies with the aug-cc-pVQZ basis set. From the values of ($E_{\mathrm{HF};\,X=3}$), ($E_{\mathrm{HF};\,X=4}$) and α obtained by interpolation between the various reference geometries, it was possible to compute the value $E_{\mathrm{HF};\,X=\infty}$ for each point. The correlation energies computed at the triple-zeta basis set level, ($E_{\mathrm{corr}}(nd+d);\ X=3$), which include the nondynamical (nd) and dynamical energies (d), are multiplied by the scaling factor, γ, also interpolated between all the reference geometries. The energy value for the correlation so obtained is then added to the ($E_{\mathrm{HF};\,X=\infty}$) to estimate the total energy to be adjusted, ($E_{\mathrm{Total;\,CCSD(T)}}$). The same explanation can be extended to the points of the PES computed with the CASPT2 method represented by the vertical bar at the right side of Figure 4.3. The main difference is that, for these point energies, the nondynamical part of the correlation energy is treated on the extrapolation procedure toward the CBS limit.

Chapter 5

Functional form and Fit

After performing the *ab initio* calculations and the extrapolation/scaling procedure, the next step is to develop an analytical potential function invariant to the permutation of identical atoms within the DMBE framework. In the next sections we will explain how we carried out this task, in particular, the different many-body terms used, the symmetric coordinates applied in the invariant polynomials, the definition of the reference geometries and the range factors that should envelope the different polynomials. At the end, we will detail the fitting procedure adopted.

5.1 Functional form of the PES

Due to spin conservation, the singlet ground state of the hydrogen peroxide molecule accommodates the following dissociation scheme,

$$\mathrm{H_2O_2(X,^1\,A)} \rightarrow \begin{cases} \mathrm{OH}(X\,^2\Pi) + \mathrm{OH}(X\,^2\Pi) & -0.34023E_\mathrm{h} \\ \mathrm{H_2}(X\,^1\Sigma_g^+) + \mathrm{O_2}(a\,^1\Delta_g) & -0.32997E_\mathrm{h} \\ \mathrm{O}(^1D) + \mathrm{H_2O}(X\,^1A_1) & -0.29845E_\mathrm{h} \\ \mathrm{H}(^2S) + \mathrm{HO_2}(X\,^2A'') & -0.27818E_\mathrm{h} \\ 2\mathrm{H}(^2S) + \mathrm{O_2}(^3\Sigma_g^-) & -0.19157E_\mathrm{h} \\ \mathrm{H_2}(X\,^1\Sigma_g^+) + 2\mathrm{O}(^3P) & -0.17448E_\mathrm{h} \\ \mathrm{H}(^2S) + \mathrm{O}(^3P) + \mathrm{OH}(X\,^2\Pi) & -0.17014E_\mathrm{h} \\ 2\mathrm{H}(^2S) + 2\mathrm{O}(^3P) & 0.00000E_\mathrm{h} \end{cases}, \tag{5.1}$$

where the quoted energies correspond to the energy of the asymptotic potentials used to construct the PES. Since we aim at an accurate global PES, the functional form adopted for their representation should be able to describe all the previous equations. To accomplish it, we proposed a three-valued representation of the H_2O_2 potential energy

surface as the lowest eigenvalue of a symmetric 3×3 diabatic potential matrix of the form,

$$\begin{bmatrix} V_{11}(\mathbf{R}) & V_{12}(\mathbf{R}) & V_{13}(\mathbf{R}) \\ V_{21}(\mathbf{R}) & V_{22}(\mathbf{R}) & V_{23}(\mathbf{R}) \\ V_{31}(\mathbf{R}) & V_{32}(\mathbf{R}) & V_{33}(\mathbf{R}) \end{bmatrix}. \tag{5.2}$$

The diagonal terms $V_{11}(\mathbf{R})$, $V_{22}(\mathbf{R})$, and $V_{33}(\mathbf{R})$ represent effective diabatic potentials associated to the dissociation channels illustrated in Equation 5.1 within the double many-body expansion framework. The reason for the 3×3 matrix option may be understood based on the different dissociation states of the O atom and the O_2 molecule. The diabatic function $V_{11}(\mathbf{R})$ dissociates to O(^1D) while the $V_{22}(\mathbf{R})$ dissociates to O(^3P), whereas $V_{33}(\mathbf{R})$ is necessary to describe the O_2 ($a\,^1\Delta_g$) dissociation channel which also correlates to O(^3P) atoms.

The nondiagonal terms $V_{12}(\mathbf{R})$, $V_{13}(\mathbf{R})$, and $V_{23}(\mathbf{R})$ reproduce the coupling between the effective diabatic terms. These effective diabatic terms are functional forms and do not correspond to any diabatic calculation result. In Equation 5.2, $\mathbf{R}$ represents the vector of the six internuclear distances defined in Figure 4.1.

5.1.1 One-, two- and three-body terms

As we are only interested in the lower PES, we further simplify our work considering an effective four-body term for all the diabatic potentials and couplings between them. This way, the diagonal terms $V_{11}(\mathbf{R})$, $V_{22}(\mathbf{R})$, and $V_{33}(\mathbf{R})$ are described only by one-, two- and three-body terms. These terms, noted by a parenthesis in the superscript, are taken from the H_2O ($^1A_1$), H_2O ($^3A'$) and HO_2 ($^2A''$) potential energy surfaces [85–87],

$$\begin{aligned} V_{11}(\mathbf{R}) \;=\; & V^{(1)}_{\mathrm{O(^1D)}} + V^{(1)}_{\mathrm{O(^1D)}} + V^{(2)}_{\mathrm{O_2(^3\Sigma_g^-)}}(R_1) + V^{(2)}_{\mathrm{H_2(^1\Sigma_g^+)}}(R_2) + V^{(2)}_{\mathrm{OH(^2\Sigma)}}(R_3) \\ & + V^{(2)}_{\mathrm{OH(^2\Sigma)}}(R_4) + V^{(2)}_{\mathrm{OH(^2\Sigma)}}(R_5) + V^{(2)}_{\mathrm{OH(^2\Sigma)}}(R_6) \\ & + V^{(3)}_{\mathrm{HO_2\,(^2A'')}}(R_1, R_3, R_5) + V^{(3)}_{\mathrm{HO_2\,(^2A'')}}(R_1, R_4, R_6) \\ & + V^{(3)^1}_{\mathrm{H_2O\,(^1A_1)}}(R_2, R_3, R_6) + V^{(3)^1}_{\mathrm{H_2O\,(^1A_1)}}(R_2, R_4, R_5), \end{aligned} \tag{5.3}$$

$$\begin{aligned} V_{22}(\mathbf{R}) \;=\; & V^{(2)}_{\mathrm{O_2(^3\Sigma_g^-)}}(R_1) + V^{(2)}_{\mathrm{H_2(^3\Sigma_u^+)}}(R_2) + V^{(2)}_{\mathrm{OH(^2\Pi)}}(R_3) + V^{(2)}_{\mathrm{OH(^2\Pi)}}(R_4) \\ & + V^{(2)}_{\mathrm{OH(^2\Pi)}}(R_5) + V^{(2)}_{\mathrm{OH(^2\Pi)}}(R_6) + V^{(3)}_{\mathrm{HO_2\,(^2A'')}}(R_1, R_3, R_5) \\ & + V^{(3)}_{\mathrm{HO_2\,(^2A'')}}(R_1, R_4, R_6) + V^{(3)^2}_{\mathrm{H_2O\,(^1A_1)}}(R_2, R_3, R_6) \\ & + V^{(3)^2}_{\mathrm{H_2O\,(^1A_1)}}(R_2, R_4, R_5), \end{aligned} \tag{5.4}$$

Table 5.1: Values of the potential parameters used to fit the extended-Rydberg potential function for the $O_2(a\,^1\Delta_g)$.

	D_e/cm^{-1}	$r_e/Å$	$a_1/Å^{-1}$	$a_2/Å^{-2}$	$a_3/Å^{-3}$
$O_2(a\,^1\Delta_g)$	34127	1.21563	6.000	10.075	8.714

$$\begin{aligned} V_{33}(\mathbf{R}) &= V^{(2)}_{O_2(^1\Delta_g)}(R_1) + V^{(2)}_{H_2(^1\Sigma_g^+)}(R_2) + V^{(2)}_{OH(^2\Pi)}(R_3) + V^{(2)}_{OH(^2\Pi)}(R_4) \\ &\quad + V^{(2)}_{OH(^2\Pi)}(R_5) + V^{(2)}_{OH(^2\Pi)}(R_6) + V^{(3)}_{HO_2\,(^2A'')}(R_1, R_3, R_5) \\ &\quad + V^{(3)}_{HO_2\,(^2A'')}(R_1, R_4, R_6) + V^{(3)}_{H_2O\,(^3A')}(R_2, R_3, R_6) \\ &\quad + V^{(3)}_{H_2O\,(^3A')}(R_2, R_4, R_5). \end{aligned} \tag{5.5}$$

The analytical potential for the $^1\Delta_g$ state of the molecular oxygen, which was not considered in the potential surfaces above referred to, was based on a three parameter functional form, the extended-Rydberg function, proposed by Murrel and Sorbie [88] for the description of diatomic potentials,

$$V(\rho) = -D_e(1 + a_1\rho + a_2\rho^2 + a_3\rho^3)\exp(-a_1\rho), \tag{5.6}$$

where $\rho = R_1 - R_e$, R_e is the equilibrium internuclear distance and D_e is the diatomic dissociation energy. The values of the parameters a_1, a_2, and a_3, listed in Table 5.1, were taken from a previous paper by Nieh and Valenti [89]. As the $H_2O\,(^1A_1)$ PES [85] is described by two diabatic surfaces and a crossing term, the particular function used is identified by an additional superscript.

The $V_{12}(\mathbf{R}) = V_{21}(\mathbf{R})$ terms, which describe the diabatic crossings within the $H_2O(^1A_1)$ molecule, are given by,

$$V_{12}(\mathbf{R}) = V^{(3)^{12}}_{H_2O\,(^1A_1)}(R_2, R_3, R_6) + V^{(3)^{12}}_{H_2O\,(^1A_1)}(R_2, R_4, R_5). \tag{5.7}$$

For simplicity, we further consider the other nondiagonal terms (V_{13}, V_{23}, V_{31} and V_{32}) of the matrix to be zero. This is a result of the above mentioned approach of an effective four-body term, which should also take into account the coupling between these three-body terms. By assuming V_{13} and V_{23} to be zero the matrix form represented in Equation 5.2 becomes a block-diagonal matrix. Since the eigenvalues of a block-diagonal matrix are the eigenvalues of the diagonal blocks, the lowest eigenvalue of the 3×3 matrix is simply

given by,

$$V = Min(V_{33}, V_{interm}), \tag{5.8}$$

in which V_{interm} is,

$$V_{interm} = \frac{(V_{11} + V_{22}) - \sqrt{(V_{11} + V_{22})^2 - 4(V_{12})^2}}{2}. \tag{5.9}$$

Using the previous equations for the one-, two- and three-body terms, the lower eigenvalues of Equation 5.2, henceforth referred to as $V^{(3)}_{H_2O_2}$, accurately reproduce all the dissociation channels presented in Equation 5.1. The overall behaviour of $V^{(3)}_{H_2O_2}$ term is very close to the computed energies, being the average absolute difference of $0.02\,E_h$.

5.1.2 Four-body dipole-dipole interaction

Considering the four-body terms and due to the different functional forms, it is useful to separate the long-range interactions from the short and intermediate energies. With regard to the long-range interactions between the different fragments, one of the most important contributions to the dynamics of this system is the electrostatic interaction between two OH molecules. This is the only long-range term we explicitly treat in this PES.

This electrostatic interaction may arise from two different OH pairs, $O_1H_1 \cdots O_2H_2$ and $O_1H_2 \cdots O_2H_1$, where the subscripts label the atoms. The total electrostatic interaction is computed as a sum of both terms. Since we are considering the interaction between two neutral diatomic molecules, the dipole-dipole interaction component, resulting from the multipole expansion of the long-range electrostatic interaction, is the leading term that retains most of the information of the electronic energy [90] and is given by

$$V_{\mu\mu}(r, \theta_1, \theta_2, \phi) = \frac{-\mu\mu}{r^3}(2\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2\cos\phi), \tag{5.10}$$

where θ_1 and θ_2 are the angles formed by each dipole with respect to the line connecting the centre of mass of the two linear charge distributions with distance r and ϕ denoting the rotation of the two distributions about the line joining them.

The *ab initio* dipole moment for the OH molecule and its dependence on the interatomic distance has already been studied by Brandão and Rio [91]. Figure 5.1 shows the theoretical calculations performed at the MCSCF level with an augmented triple-zeta quality basis set for different OH internuclear distances. To our purpose, we constrained the dipole moment of the OH molecule to these theoretical values. Regarding the dipole moment of the two OH dipoles, in order to avoid the use of angles and their derivatives, we

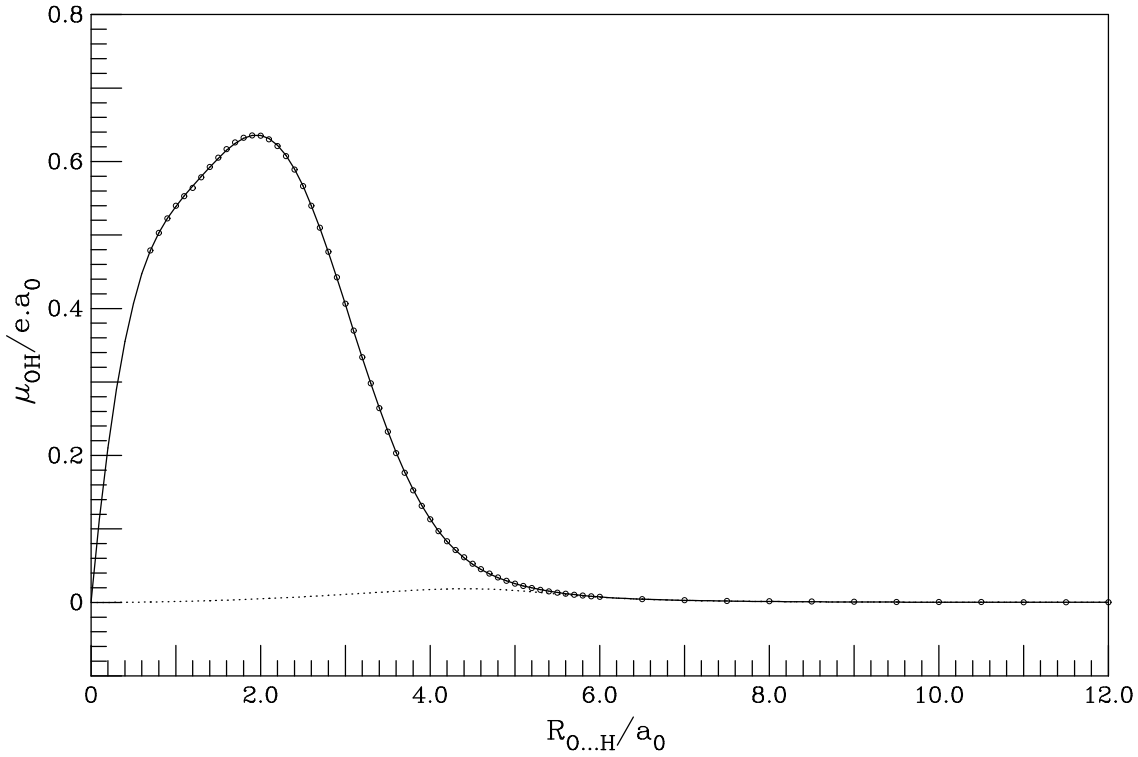


Figure 5.1: Dipole moment of the diatomic OH($^2\Pi$); the thicked line corresponds to the fit of the computed *ab initio* points ∘, the dashed line corresponds to the r^{-3} term. See in Reference [91] the functional form and the corresponding fitted coefficients.

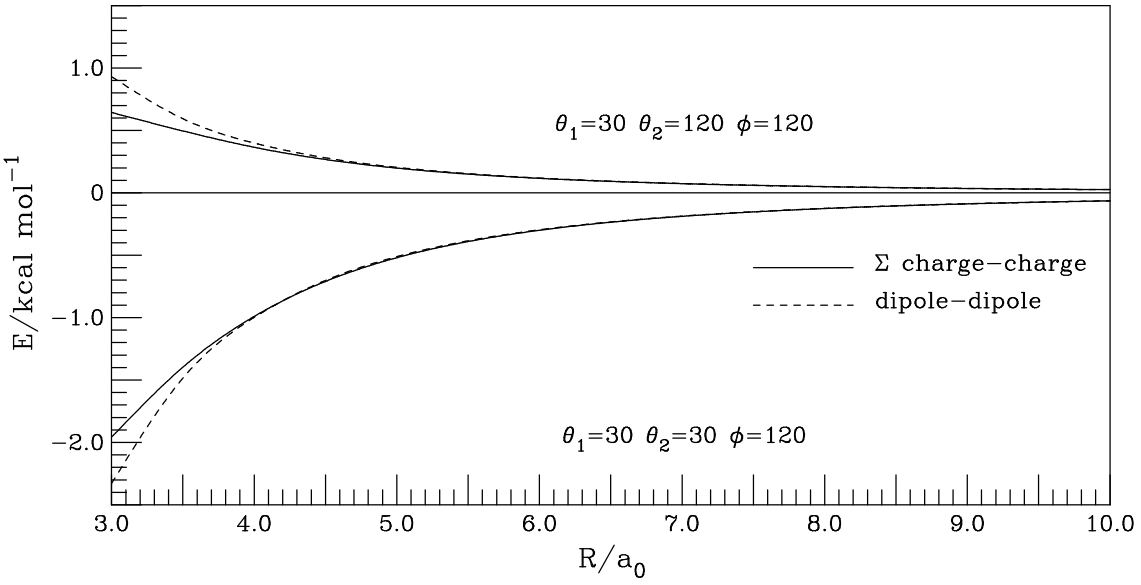


Figure 5.2: Comparision between $V_{\mu\mu}(r, \theta_1, \theta_2, \phi)$ and $V^{(4)}_{\text{elect}}(\mathbf{R})$ as a function of the distance between the centre of the OH bonds.

split the dipole moment of each OH diatomic in two symmetric atomic charges located at each nucleus, $Q_{\text{H}_\text{i}} = -Q_{\text{O}_\text{i}} = \mu_i/R_i$, and used the different charge-charge interactions to compute the interaction energy instead of applying the usual multipole expansion equation to represent the electrostatic energy. Thus, this interaction energy is computed as,

$$V^{(4)}_{\text{elect}}(\mathbf{R}) = \frac{Q_{\text{O}_1}Q_{\text{H}_1}}{R_3}\chi_1(R_3) + \frac{Q_{\text{O}_2}Q_{\text{H}_2}}{R_4}\chi_1(R_4) + \frac{Q_{\text{O}_1}Q_{\text{O}_2}}{R_1}\chi_1(R_1) + \frac{Q_{\text{H}_1}Q_{\text{H}_2}}{R_2}\chi_1(R_2), \quad (5.11)$$

where $\chi_1(R)$ are the usual damping functions for the multipole expansion [92, 93] to account for the breakdown of the charge separation due to orbital overlap. The validity of this function, $V^{(4)}_{\text{elect}}$, was tested by comparison with the dipole moment expansion term, Equation 5.10. Figure 5.2 shows the results as a function of the O-H distance for two different orientations of the dipoles. As we can observe, the behaviour of both functions are in close agreement with each other being the differences negligible at large distances, for $R > 4.5\, a_0$, where the multipole expansion becomes valid. The dependence of the dipole-dipole interaction on the orientation of the two molecules is quite noticeable being attractive (negative region of the plot) or repulsive (positive area) as one dipole is rotated with respect to the other.

5.1.3 Short-range four-body energy

The short-range four-body term, $V^{(4)}_{sr}$, is obtained by fitting the difference between the energies of the *ab initio* calculations and the sum of the $V^{(3)}_{\text{H}_2\text{O}_2}$ and $V^{(4)}_{elect}$ terms of the

potential energy surface. The functional form of $V_{sr}^{(4)}$ is given by a sum of polynomial functions of the fourth degree each multiplied by a range factor term, further represented by $\Theta(\mathbf{R})$, both centred at specific reference geometries. The coordinates used to construct the polynomial functions, the determination of the reference geometries and the range factor terms are discussed at length in the following section.

5.1.3.1 Symmetric coordinates

Due to the invariance of the electronic Hamiltonian, the Potential Energy Surface (PES) for a molecular system should be invariant to the permutation of identical atoms. Despite the development of the theory of invariant functions by Molien in 1897 [94], its application to physical science was not immediate. However, the construction of invariant functions of finite transformation groups has long been applied to crystal point groups [95], since the thermodynamic potentials of any symmetric system need to be invariant to the group of its symmetry. One of the first applications of Molien's theorem to build a PES that included the correct permutation symmetry of identical atoms was proposed by Schmelzer and Murrell [96]. In this work the analytic form of an invariant potential function obtained was applied to tetra-atomic homonuclear molecules, a X_4 system.

Developing an analytical potential function able to express the symmetry-invariance property of a molecule containing two or more indistinguishable nuclei is important from the practical point of view. This restriction of the PES, i.e., the fact that the potential function must be totally symmetric concerning the permutation of atoms of the same chemical species, greatly reduces the number of *ab initio* calculations needed to cover the regions of chemical interest of the PES since a part of the configuration space becomes redundant.

This topic is a whole research field in Algebra [97, 98] and so, to our purpose, we limit ourselves to describe how to apply the theory of invariant functions without going into further details. It should be noted that the following description for the construction of a symmetric polynomial for the H_2O_2 system can be also applied to any other system that includes in its structure two pairs of identical atoms which, in a generalized way, may be represented by a A_2B_2 system.

5.1.3.2 The construction of invariant polynomials

Let us consider the hydrogen peroxide molecule illustrated in Figure 4.1, page 46, in which the atoms were labelled as O_1, O_2, H_1, H_2. To describe the molecule, we need to define a set of internal coordinates. For a four-atom system ($N = 4$) we have a total of $3N - 6$ internal coordinates, i.e., a total of six internal coordinates, each one corresponding to a

distance between two atoms. These 6 internal coordinates are labelled as R_1, R_2, R_3, R_4, R_5 and R_6, where R_1 stands for the distance between the two O atoms, R_2 between the two H atoms and R_3, R_4, R_5 and R_6 represent the four possible distances between the O and H atoms.

By using the previous labels we can define the complete nuclear permutation group (CNP) [99] of the H_2O_2 molecule which represents all possible permutations of identical atoms in the molecule. Since we have two sets of identical atoms, the CNP of the molecule can be obtained from two distinct groups, being the first one the group of all oxygen nuclei permutations and the second the group of all hydrogen nuclei permutations, which we can represent, respectively, as $S_2^{(O)}$= {E; O_1O_2} and $S_2^{(H)}$= {E; H_1H_2}. So, each group will exhibit an identity operation (E) and a nuclear permutation operation relative to the atom under consideration (O_1O_2 or H_1H_2). Each of these symmetry operations acts on the molecule changing, or not, the order of the internal coordinates. Exemplifying, the permutation operation of two oxygen atoms leads to the following result,

$$(\mathrm{O}_1\,\mathrm{O}_2) = \begin{pmatrix} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 \\ R_1 & R_2 & R_5 & R_6 & R_3 & R_4 \end{pmatrix}.$$

It turns out that by applying this operation the internal coordinates R_1 and R_2 remain the same but all the others change their value, R_3 assumes the value of R_5 and R_4 assumes the value of R_6 and vice versa. The group of all permutations of identical atoms in the molecule is given by the combination of all elements of the $S_2^{(O)}$ and all the elements of the $S_2^{(H)}$ group, which corresponds to $2! \times 2! = 4$ elements in the total [99]. In other words, we may say that the CNP group of the molecule is the direct product of the groups $S_2^{(O)}$ and $S_2^{(H)}$ and it can be written as $G^{\mathrm{CNP}} = S_2^{(O)} \otimes S_2^{(H)}$, which is isomorphic to the D_2 point group. It consists of a total of four permutation operations,

$$G^{\mathrm{CNP}} = \{\mathrm{E}; (\mathrm{O}_1\mathrm{O}_2); (\mathrm{H}_1\mathrm{H}_2); (\mathrm{O}_1\mathrm{O}_2)(\mathrm{H}_1\mathrm{H}_2)\}, \tag{5.12}$$

where E is the identity operation (in which the molecule remains unchanged); (O_1O_2) is the permutation operation of the two identical O atoms; (H_1H_2) is the permutation operation of the two identical H atoms and (O_1O_2)(H_1H_2) is the concomitant permutation operation of the two identical O atoms and the two identical H atoms. From now on, these four operations will be referred to as g.

To obtain an invariant polynomial concerning the permutations of identical atoms, we must first understand the meaning of it. We may define a six dimension vector $\vec{r}$ with

components equal to the internal coordinates,

$$\vec{r} = (R_1; R_2; R_3; R_4; R_5; R_6). \tag{5.13}$$

A polynomial f is symmetric if,

$$f(\vec{r}) = f(g(\vec{r})), \forall g \in G^{\mathrm{CNP}}. \tag{5.14}$$

This can be accomplished by constructing a set of invariant functions, $\varphi(\vec{r})$, so that

$$f(\vec{r}) = F(\varphi_1(\vec{r}); \varphi_2(\vec{r}); \ldots; \varphi_n(\vec{r})). \tag{5.15}$$

Once defined the concept of an invariant polynomial three questions must be answered in order to construct it: how many terms of each degree we should use; what is the smallest number of terms, called integrity basis, that can be used to build the polynomial and how we can define this integrity basis. Let us focus on the answer to the first question.

As we have already exemplified previously, the permutation operations g can act directly on the H_2O_2 molecule, hence on the $\vec{r}$. Thus, g can be described in a matricial 6-dimensional representation which is defined as the $\Re$ representation of this group for this coordinate basis,

$$\mathrm{E} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (\mathrm{A_1A_2})(\mathrm{B_1B_2}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(\mathrm{A_1A_2}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (\mathrm{B_1B_2}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{5.16}$$

Molien [94] has shown that by using a set of independent coordinates, ξ_i, any invariant

polynomial of a finite group can be obtained by the following generating function

$$\phi_G(\lambda) = \frac{1}{|G|} \sum_{g \in G} \frac{1}{det[\mathbf{I} - \lambda \mathbf{M}(g)]}, \tag{5.17}$$

which is the product of the inverse of the order of the group, $|G|$, by the sum, for all the g permutation operations of the G group, of the inverse of the determinant of the difference between the identity matrix $\mathbf{I}$ and the product of a dummy variable λ by the matrices of the representation in the basis ξ_i represented as $\mathbf{M}(g)$. Applying this expression to our $\Re$ representation we have

$$\phi_G(\lambda) = \frac{1}{4} \left[\frac{1}{(1-\lambda)^6} + \frac{3}{(1-\lambda)^2(1-\lambda^2)^2} \right], \tag{5.18}$$

which, as showed by Molien, can be rearranged in a more condensed way as

$$\phi_G(\lambda) = \frac{1+\lambda^3}{(1-\lambda)^3(1-\lambda^2)^3}. \tag{5.19}$$

Any invariant polynomial function of a finite group can be developed in terms of a finite number of invariant polynomials, called the integrity basis. If we split all the terms in Equation 5.19 we can rewrite it as,

$$\phi_G(\lambda) = (1+\lambda^3) \times \frac{1}{(1-\lambda)} \times \frac{1}{(1-\lambda)} \times \frac{1}{(1-\lambda)} \times \frac{1}{(1-\lambda^2)} \times \frac{1}{(1-\lambda^2)} \times \frac{1}{(1-\lambda^2)}. \tag{5.20}$$

Since the power series expansion of the geometric series $\frac{1}{1-x}$ gives rise to the following sum of powers of x,

$$\frac{1}{(1-x)} = \sum_{i=0}^{\infty} x^i = 1 + x + x^2 + x^3 + x^4 + \cdots, \tag{5.21}$$

we can, in the same way, expand each of the fractions of the generating function. The expansion leads to a product of seven polynomials

$$\begin{aligned} \phi_G(\lambda) =& (1+\lambda^3) \times (1+\lambda+\lambda^2+\lambda^3+\cdots) \times (1+\lambda+\lambda^2+\lambda^3+\cdots) \times \\ & (1+\lambda+\lambda^2+\lambda^3+\cdots) \times (1+\lambda^2+\lambda^4+\lambda^6+\cdots) \times \\ & (1+\lambda^2+\lambda^4+\lambda^6+\cdots) \times (1+\lambda^2+\lambda^4+\lambda^6+\cdots). \end{aligned}$$

The first polynomial is the numerator, the three following polynomials result from the complete expansions of the first degree terms, and the three last ones result from the

complete expansion of the second degree terms. Therefore, we can say that the integrity basis consists of three terms totally symmetric of the first degree $(1-\lambda)$, three terms totally symmetric of the second degree $(1-\lambda^2)$ and one term totally symmetric of the third degree $(1-\lambda^3)$, which can only be used once in each monomial. The multiplication of the seven polynomials yields

$$\phi_G(\lambda) = 1 + 3\lambda + 9\lambda^2 + 20\lambda^3 + 42\lambda^4 + 78\lambda^5 + \cdots . \tag{5.22}$$

Equation 5.22 gives us the information relative to the number of terms of each degree that we will need in order to build an invariant polynomial. So, for our tetratomic system we will have a term of degree 0; 3 terms of degree 1; 9 of degree 2; and so on.

So far we have become aware of the dimension of the integrity basis and the number of terms of each degree that we can use to construct the invariant polynomial; we are now going to focus our attention in the way used to define the integrity basis.

Knowing the character, $\chi^{\Re}$, of the matrices that represent the permutation operations g and the character table of the D_2 point group, to which our permutation group is isomorphic to, we can decompose the reducible representation, $\Re$, of the $G^{\mathrm{CNP}} = S_2^{(O)} \otimes S_2^{(H)}$ group into a direct sum of irreducible representations. The number of times, m_i, that each irreducible representation is contained in the reducible representation is given by,

$$m_i = \frac{1}{|G|} \sum_{g \in G} \chi^{\Re}_{(g)} \chi^{i}_{(g)}, \tag{5.23}$$

where $|G|$ represents the order of the group, $\chi^{\Re}_{(g)}$ is the character of the symmetric operation g for the reducible representation $\Re$, and $\chi^{i}_{(g)}$ is the character of the symmetric operation g for the irreducible representation i. Considering the irreducible representation A we have,

$$m_A = \frac{1}{4}[(6 \times 1) + (2 \times 1) + (2 \times 1) + (2 \times 1)] = 3. \tag{5.24}$$

Applying this expression to the remaining irreducible representations of the D_2 group, we find that the complete reduced structure of the reducible representation, $\Re$, is,

$$\Re = 3A \oplus 1B_1 \oplus 1B_2 \oplus 1B_3. \tag{5.25}$$

Thus, with the six internal coordinates R_1, R_2, R_3, R_4, R_5 and R_6, we can construct new ones in which three will behave according to the irreducible representation A; one will behave according to the irreducible representation B_1; one will behave according to the irreducible representation B_2; and one according to the irreducible representation B_3. This new coordinates, called symmetric coordinates, can be defined by the application of

the projection operator, $\mathcal{P}^i$. The equation defining it is,

$$\mathcal{P}^i = \frac{n_i}{|G|} \sum_g \chi_g^{(i)} \Theta_g, \tag{5.26}$$

where n_i is the dimension of the irreducible representation i, $|G|$ is the group order, $\chi_g^{(i)}$ is the character of the irreducible representation i and Θ_g is the operator that performs the symmetry operation g. There is one projection operator for each irreducible representation of the D_2 group, for a total of four operators. For example, considering the B_1 representation, the projection operator is,

$$\mathcal{P}^{B_1} = \frac{1}{4}[E + C_2(z) - C_2(y) - C_2(x)]. \tag{5.27}$$

Since the $G^{\text{CNP}} = S_2^{(A)} \otimes S_2^{(B)}$ is isomorphic to the D_2 group, this is equivalent to say that for a specific symmetry operation of the point group D_2, there exists a symmetric permutation operation of $G^{\text{CNP}} = S_2^{(A)} \otimes S_2^{(B)}$ that behaves in the same way as the one of the D_2 group, i.e., there is a one-to-one correspondence between the two groups. This implies that the previous operator for the B_1 representation of the D_2 point group can be expressed for the $G^{\text{CNP}} = S_2^{(A)} \otimes S_2^{(B)}$ as,

$$\mathcal{P}^{B_1} = \frac{1}{4}[E + (A_1A_2)(B_1B_2) - (A_1A_2) - (B_1B_2)]. \tag{5.28}$$

The application of this operator, for instance, to the internal coordinate R_3 yields the following result,

$$\mathcal{P}^{B_1}(R_3) = \frac{1}{4}(R_3 + R_4 - R_5 - R_6). \tag{5.29}$$

By using the same procedure for the other three projection operators, we obtain combinations of the six internal coordinates that behave according to the irreducible representation of the point group D_2,

$$\begin{aligned}
\rho_1 &= R_1 && \in A \\
\rho_2 &= R_2 && \in A \\
\rho_3 &= \frac{1}{4}(R_3 + R_4 + R_5 + R_6) && \in A \\
\rho_4 &= \frac{1}{4}(R_3 + R_4 - R_5 - R_6) && \in B_1 \\
\rho_5 &= \frac{1}{4}(R_3 - R_4 + R_5 - R_6) && \in B_2 \\
\rho_6 &= \frac{1}{4}(R_3 - R_4 - R_5 + R_6) && \in B_3.
\end{aligned} \tag{5.30}$$

Table 5.2: Algebraically independent and auxiliary invariants for the H_2O_2 system.

Degree			
1	$\varphi_1(\vec{r}) = \rho_1 = R_1$	$\varphi_2(\vec{r}) = \rho_2 = R_2$	$\varphi_3(\vec{r}) = \rho_3$
2	$\varphi_4(\vec{r}) = \rho_4^2$	$\varphi_5(\vec{r}) = \rho_5^2$	$\varphi_6(\vec{r}) = \rho_6^2$
3	$\varphi_7(\vec{r}) = \rho_4\rho_5\rho_6$		

Table 5.3: Invariant polynomial terms constructed from the seven basis terms.

Degree	Number	Terms
0	1	Constant
1	3	$\varphi_1(\vec{r});\ \varphi_2(\vec{r});\ \varphi_3(\vec{r})$
2	9	$\varphi_1(\vec{r})^2;\ \varphi_1(\vec{r})\,\varphi_2(\vec{r});\ \varphi_1(\vec{r})\,\varphi_3(\vec{r});$ $\varphi_2(\vec{r})^2;\ \varphi_2(\vec{r})\,\varphi_3(\vec{r});\ \varphi_3(\vec{r})^2;$ $\varphi_4(\vec{r});\ \varphi_5(\vec{r});\ \varphi_6(\vec{r})$
3	20	$\varphi_1(\vec{r})^3;\ \varphi_1(\vec{r})^2\,\varphi_2(\vec{r});\ \varphi_1(\vec{r})^2\,\varphi_3(\vec{r});$ $\varphi_1(\vec{r})\,\varphi_2(\vec{r})^2;\ \varphi_1(\vec{r})\,\varphi_2(\vec{r})\,\varphi_3(\vec{r});\ \varphi_1(\vec{r})\,\varphi_3(\vec{r})^2$ $\varphi_1(\vec{r})\,\varphi_4(\vec{r});\ \varphi_1(\vec{r})\,\varphi_5(\vec{r});\ \varphi_1(\vec{r})\,\varphi_6(\vec{r});$ $\varphi_2(\vec{r})^3;\ \varphi_2(\vec{r})^2\,\varphi_3(\vec{r});\ \varphi_2(\vec{r})\,\varphi_3(\vec{r})^2;\ \varphi_2(\vec{r})\,\varphi_4(\vec{r});$ $\varphi_2(\vec{r})\,\varphi_5(\vec{r});\ \varphi_2(\vec{r})\,\varphi_6(\vec{r});\ \varphi_3(\vec{r})^3;\ \varphi_3(\vec{r})\,\varphi_4(\vec{r});$ $\varphi_3(\vec{r})\,\varphi_5(\vec{r});\ \varphi_3(\vec{r})\,\varphi_6(\vec{r});\ \varphi_7(\vec{r})$
4	42	...

As would be expected, the two diatomic distances R_1 and R_2 belong to the totally symmetric representation A since they do not exchange when any of the four permutation operations is applied to the molecule. The sum of R_3, R_4, R_5 and R_6, denoted as ρ_3, also belongs to the totally symmetric representation A, but the other three combinations do not. To build totally symmetric coordinate terms from the ρ_4, ρ_5 and ρ_6 we can use the direct product multiplication table of the D_2 group. It can be easily shown that ρ_4^2, ρ_5^2 and ρ_6^2 constitute the three invariant terms of the second degree and that $\rho_4 \times \rho_5 \times \rho_6$ is the third degree invariant term.

Summing up, the invariant polynomial for an H_2O_2 molecule will be constructed as combinations of the seven terms described in Table 5.2, which constitute the integrity basis.

Table 5.3 shows the first polynomial terms to use up to the third degree.

The coordinates ρ_1 and ρ_2 have a well defined physical meaning. However, it is important to remark the relation between the remaining coordinates and the conformation of the molecule. Coordinates ρ_3 up to ρ_6 expressed in Equation 5.30 are related to the asymptotic channels of the H_2O_2 molecule.

High values for ρ_1, ρ_2 and ρ_3 mean that all the atoms of the molecule are far apart, that is, the molecule is dissociated into its atomic fragments. On the other hand, small values of ρ_1 together with large values for ρ_2 and ρ_3 imply the presence of an O_2 molecule. Yet, when the ρ_2 has a small value an H_2 is formed. If we consider only ρ_3 to be large and ρ_1 and ρ_2 both small, then we will be in the presence of the asymptotic channel that leads to the formation of O_2 plus H_2.

The coordinate ρ_4 gives us information relative to the dissociation channel OH plus OH. As for the bond length representation in Figure 4.1, as the value of ρ_4 increases two pairs of OH molecules are being formed. Negative values for ρ_4 imply the formation of the pairs O_1H_1 and O_2H_2 while a positive value represents the presence of the O_1H_2, O_2H_1 pairs. As by symmetry the energy is the same whatever possibility we consider, ρ_4 must be squared in order to be used as an invariant coordinate which we define as $\varphi_4(\vec{r})$.

The coordinate ρ_5 is related to the distance of each hydrogen atom to the oxygen atoms. It assumes positive values if H_1 is farther from the O atoms than the H_2 atom, and it will have negative values if we consider the opposite case. So, large values of ρ_5 signify that an H atom is approaching an HO_2 molecule and the sign identifies the hydrogen atom that integrates the HO_2 molecule. Again, since we have two possibilities which are energetically indistinguishable, we must take $\varphi_5(\vec{r})$ (square of ρ_5) as an invariant coordinate.

The same reasoning can be applied to the ρ_6, which has a higher weight in the dissociation channel that leads to the formation of O plus H_2O. For high and positive values of ρ_6 we will have O_1 plus H_2O. High and negative values of ρ_6 yield O_2 plus H_2O. The square of this coordinate is referred to as $\varphi_6(\vec{r})$.

The domain of the φ coordinates expressed in Table 5.2 ranges from 0 to ∞ with the exception of the seventh invariant term of the third degree, which can take either a positive or a negative value depending on the signs of ρ_4, ρ_5 and ρ_6. Since $\varphi_7(\vec{r})$ results from the product of the three previous variables, there are four possibilities to yield a positive value and another four possibilities to yield a negative value for this symmetric coordinate. All of these eight possibilities correspond to two distinct groups of symmetric molecular conformations of the H_2O_2 molecule, one with a positive and another with a negative sign for the $\varphi_7(\vec{r})$.

It should be noted that a change of sign in $\varphi_7(\vec{r})$ is not associated with the occurrence of any of the permutation symmetry operations of the hydrogen peroxide molecule. To

Table 5.4: Displacement integrity basis.

Degree			
1	$\varphi_1(\vec{r}) = \rho_1 - \rho_{1,ref}$	$\varphi_2(\vec{r}) = \rho_2 - \rho_{2,ref}$	$\varphi_3(\vec{r}) = \rho_3 - \rho_{3,ref}$
2	$\varphi_4(\vec{r}) = \rho_4^2 - \rho_{4,ref}^2$	$\varphi_5(\vec{r}) = \rho_5^2 - \rho_{5,ref}^2$	$\varphi_6(\vec{r}) = \rho_6^2 - \rho_{6,ref}^2$
3	$\varphi_7(\vec{r}) = \rho_4\rho_5\rho_6 - \rho_{4,ref}\rho_{5,ref}\rho_{6,ref}$		

corroborate this observation let us look at what happens to $\varphi_7(\vec{r})$, whatever the permutation operation. The permutation of two oxygen atoms changes the sign of ρ_4 and ρ_6; the permutation of two hydrogen atoms reverses the sign of ρ_4 and ρ_5; and the concomitant switch of two oxygen and two hydrogen atoms alters the sign of ρ_5 and ρ_6. We realize that any permutation operation gives rise to a matched number of swaps in the signs of the coordinates and consequently the sign of $\varphi_7(\vec{r})$ remains unchanged. Following this line of reasoning we can state that, due to a lack of connection between the shift in the sign of $\varphi_7(\vec{r})$ and the permutation operations of the CNP group of the H_2O_2 molecule, two geometric structures with opposite sign of $\varphi_7(\vec{r})$ are not equivalent and the energies of these structures can be different.

5.1.3.3 Reference geometries

Totally symmetric polynomials are usually computed using displacement coordinates from some reference geometry, which is usually chosen as the equilibrium structure of the molecule or a saddle point configuration. The major limitation to this approach lies in the fact that these structures can seldom be used to define the displacement coordinates since they do not meet the symmetry requirement [100]. In the present work, instead of using displacement coordinates we resort to the application of displacement integrity basis. By doing so, each term of the integrity basis should be subtracted by the equivalent term computed at the reference geometry, being the polynomial function obtained from this displacement integrity basis. Since the values of the terms obtained at the reference geometry are constants, their subtraction do not change the invariance of the polynomial. The displacement integrity basis for the H_2O_2 system is represented in Table 5.4. Within this approach, all the integrity basis terms will vanish at the reference geometry, remaining only the constant term. There are no restrictions concerning the choice of the reference geometry and several polynomial functions can be expanded from different reference geometries giving rise to a multifunctional fit.

Since the regions of greatest importance in the dynamics of the chemical reactions that

occur in the potential surface include minima and saddle points, the structures described in Figure 4.2 were the first choices for the reference geometries. After a preliminary fitting, we observe that some regions of the potential were not well described, which drives us to the introduction of three new polynomial centres. To warrant a good description of the insertion and abstraction entrance channels for the $O(^1D)+ H_2O(X\ ^1A_1)$, we account for two new polynomial centres, denoted as O_insertion and H_abstraction, at planar geometries. The description of the approach of a hydrogen atom to the hydroperoxyl radical was further improved by the inclusion of a third centre positioned along the dissociation channel $H\cdots HO_2$, which was designated as H_approach. These eleven configurations constitute a basis of fixed references.

Due to the large spread of the computed *ab initio* points, it was necessary to consider further references to screen the configurational space that falls out of the regions covered by the eleven fixed structure references.

A fundamental problem that arises is how to find homogeneous groups of *ab initio* points in our data set and how to properly assign a reference geometry to each one of these collections of points. To do so, we may take advantage of cluster techniques. The partition of a data set, known as clustering, is a widely used procedure to recognize patterns, or clusters, in data and finds a broad number of applications in different research domains.

Our great challenge is to define each of these clusters (regions in the configuration space of the molecule in which the density of points is locally higher in comparison with other space regions) by stipulating a "similarity" criterion or measure so that the *ab initio* points within a cluster can be as similar to each other as possible, while different from the others belonging to other clusters.

There is an enormous variety of clustering algorithms among which one of the most popular is the k-means [101]. In this work we have chosen this algorithm to construct non-overlapping clusters from our set of *ab initio* points mainly because of its simplicity, easiness in implementation and capacity to yield, in general, fast results.

The basic idea behind k-means algorithm is to partition the data into k sets or centres in an iterative way, being each centre assigned to a centroid, also known as prototype or seed, which represents the general features of that specific centre.

In our case, the data to be clustered is a set of N *ab initio* points, $X = \{x_1; x_2; \cdots ; x_N\}$, in which each point corresponds to a six-dimensional (six interatomic distances) variable x. The goal is to form a number of K disjoint subsets C= $\{c_1; c_2; \cdots ; c_k\}$, each containing n points. The data points are distributed throughout the subsets based on the minimization of the squared Euclidean distances between each point x_n and the centroid m_k of the

subset c_k. The objective function to be optimized is then defined as,

$$KM(X,C) = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n - m_k\|^2, \tag{5.31}$$

where r_{nk} is a boolean variable. If the point x_n is assigned to cluster k then $\mathrm{r}_{nk} = 1$, otherwise it will be zero,

$$r_{nk} = \begin{cases} 1 & \text{if } k = \mathrm{argmin}_j\|x_n - m_j\|^2 \\ 0 & \text{otherwise} \end{cases}, \tag{5.32}$$

where $\mathrm{argmin}_j\|x_n - m_j\|^2$ is the value of j that minimizes $\|x_n - m_j\|^2$. After the assignment of the data to the closest centre, the following step is to proceed to the optimization of m_k by setting it equal to the mean of all the x_n points attributed to cluster c_k.

To briefly summarize, the methodology behind the k-means algorithm proceeds as follows [102]:

- Select randomly a number of k cluster centres $c_1, c_2, \cdots, c_k$;
- Assign each data point to its nearest cluster centroid based on Equation 5.1.3.3 (classification step);
- Recalculate each cluster centroid by setting it as the centre of all the points of the cluster to which it belongs (minimization step);
- Repeat the second and third steps until no data point is reassigned.

It must be pointed out that k-means has some important disadvantages [103]. First the user needs to identify in advance the number of clusters to give as an input, second the performance of the algorithm depends on the initial definition of the positions of the cluster centres, and third the algorithm may lead to the so called "dead unit" problem, which means that if a centre is not initialized properly, it may never win a data point in the classification step.

In our specific case the first and second problems are attenuated, thanks to the previous knowledge of the most important regions of the PES, mainly minimum and transition states. This enables us to use these structures as clusters centroids. Since these configurations are crucial to assure a good description of the topology of the PES they are kept fixed during the minimization step. Only the other cluster prototypes, used to cover the *ab initio* points that fall out the fixed centroids, are allowed to be optimized in the algorithm procedure. The initial reference geometries have been randomly generated. In this

Table 5.5: Geometry references used in this work.

Reference	Property	R_1	R_2	R_3	R_4	R_5	R_6
Ref_1	absolute minimum	2.759	4.518	3.560	3.560	1.826	1.826
Ref_2	local minimum	2.913	2.937	3.707	1.832	1.832	3.706
Ref_3	ts H+HO_2	2.527	3.698	5.110	1.846	3.295	3.486
Ref_4	ts OH+OH	5.503	6.050	5.730	5.444	1.842	1.842
Ref_5	isomerization	3.074	3.029	3.789	1.962	1.833	2.624
Ref_6	ts O_2+H_2	2.460	1.990	5.195	2.083	3.643	3.751
Ref_7	dihedral_0	2.780	3.672	3.680	3.680	1.827	1.827
Ref_8	dihedral_180	2.779	4.896	3.538	3.538	1.824	1.824
Ref_9	O_insertion int.	4.070	2.862	5.372	1.810	1.810	5.372
Ref_{10}	H_abstraction int.	5.060	2.862	4.203	1.810	1.810	4.203
Ref_{11}	H_approach	2.681	5.681	6.309	2.057	4.648	3.581
Ref_{12}	generated	4.818	3.967	4.837	3.998	1.968	1.997
Ref_{13}	generated	2.486	5.169	2.306	1.898	3.985	3.797
Ref_{14}	generated	2.966	3.713	3.658	2.855	1.854	1.989
Ref_{15}	generated	3.520	3.439	4.760	2.031	2.300	4.273
Ref_{16}	generated	3.616	4.979	4.432	4.068	2.016	1.966
Ref_{17}	generated	2.897	5.390	3.680	2.031	4.897	3.699
Ref_{18}	generated	2.606	6.000	6.000	2.074	6.000	3.592
Ref_{19}	generated	2.723	3.310	4.162	1.910	2.407	3.619
Ref_{20}	generated	4.553	5.528	5.103	4.717	1.917	1.932
Ref_{21}	generated	2.642	2.069	4.102	2.121	2.328	3.866

procedure all the six internal distances of the molecular system are uniformly distributed between 1.0 and 5.5 Å. To overcome the "dead unit" issue and the formation of small sets of points, only references to which are assigned more than three hundred points may be accepted as a reference point. If the previous requirement is not fulfilled, a new randomly generated reference will replace the older one and all the procedure restarted.

The twenty one reference geometries used in this work are quoted in Table 5.5. The fixed reference geometries have been obtained in a preliminary study using a triple-zeta basis set and do not constitute stationary points of the final PES.

Another important aspect to take into consideration is the fact that due to the symmetry invariance condition of the system each molecule configuration has four equivalent geometries. To ascribe the points to the closest reference geometry, we define the metric between the point and the reference as the lowest Cartesian distance, in interatomic coordinates, between their different equivalent geometries.

5.1.3.4 Range factors

To represent the four-body short-range interaction term of the H_2O_2 molecular system, each of the twenty one polynomials were multiplied by a range factor, which must also be invariant to any symmetry operation of the system and should also be computed using the displacement integrity basis.

A functional form that is still in common use for the representation of the range factor consists of hyperbolic tangent switching functions [104]. Since the objective is to proceed to a multifunctional fit of the four-body term $V_{sr}^{(4)}$, our approach was to use "modified" Gaussian functions rather than hyperbolic tangents as range factors, so that the polynomials would have the least possible interference with each other.

We must, however, emphasize that we do not have a normal distribution of *ab initio* points. The resource to "modified" Gaussian functions intends mainly to provide us with information concerning the dispersion of the points around the reference geometry to which they are assigned. This way, the functional form of the "modified" Gaussian functions used in the present work include a first order term to account for the spatially varying behaviour of the local points.

Our range factors, $\Theta(\mathbf{R})$, are then defined as a product of exponential functions, centred at the reference geometries of the polynomials, each composed by a quadratic and a linear term,

$$\Theta(\mathbf{R}) = \prod_{i=1,7} \exp \frac{\eta_i v_i - v_i^2}{8\lambda_i}, \tag{5.33}$$

where v_i represents displacement coordinates from the reference geometry. The negative sign on the quadratic term assures that $\Theta(\mathbf{R})$ approaches zero for geometries far from the reference. The parameters η and λ in each exponential function are estimated based on the local distribution of the points. This is important since the functional form of our fitting model is not linear with respect to this parameters and so, by setting its values according to the shape of the subset of neighbouring points, we are avoiding fitting convergence problems of the fit. These two parameters were specified by defining, in a first step, the variance-covariance matrix for each subset of points using the displacement symmetric coordinates, φ, followed by its diagonalization. The eigenvectors, $\vec{u}_i$, correspond to the directions in which the covariances are zero. The eigenvalues, λ_i, give the magnitude of the variance in each direction (along each eigenvector). The displacement coordinates v_i, used in Equation 5.33, represent the projection of the displacement symmetric coordinates, φ_j defined in Table 5.4, page 75, onto each eigenvector,

$$v_i = (\vec{\varphi}.\vec{u_i}) = \sum_{j=1,7} \varphi_j \times u_{ij}. \tag{5.34}$$

In a second step, as the point distribution around the reference may not be symmetric, the asymmetric factor for each eigenvector direction, η_i, is computed in such a way that $\Theta(\mathbf{R})$ for two extreme points should be the same. These extreme points are defined as the average of the 100 farthest points v_i for each positive (v_{imax}) and negative (v_{imin}) sides,

$$\exp \frac{\eta_i v_{imax} - v_{imax}^2}{8\lambda_i} = \exp \frac{\eta_i v_{imin} - v_{imin}^2}{8\lambda_i}. \tag{5.35}$$

Once these average extreme points are defined, the η_i values are calculated as their sum,

$$\eta_i = v_{imax} + v_{imin}. \tag{5.36}$$

To avoid the exclusion (small weight) of the farthest neighbour points, we found convenient to use a scaling factor 8 instead of the value 2 used in the Gaussian distribution to multiply the variance. This creates a superposition between neighbour regions, which favours the potential continuity.

Figure 5.3 illustrates a contour/perspective plot cut of the decay function for the points centred at the equilibrium minimum (first reference geometry of Table 5.5) in the plane defined by the two eigenvectors, $\vec{u}_i$, of higher eigenvalues (higher variances).

The axis displayed in the graphic has different scales. The eigenvector of higher variance, i.e., the eigenvector with the higher eigenvalue, λ_i, is represented in the X-axis and its principal component, for this reference, turns to be the symmetric coordinate, $\varphi_4(\vec{r})$, see Table 5.2, page 73. In turn, the Y-axis corresponds to the eigenvector with the second higher variance that mainly depends on the $\varphi_7(\vec{r})$ symmetric coordinate.

The contour plot represents cuts at different heights of the decay function $\Theta(\mathbf{R})$ according to those directions. The contours start at 0.05, expressed by the letter B and go up, with increments of 0.05, as far as letter U, which assumes the value of 1. The grey area delimits the region where $\Theta(\mathbf{R}) > 0.5$. If no asymmetric factor was introduced, the centre of the contours would match the (0,0) point of the referential. By considering the asymmetry of the point distribution around the reference geometry according to the two eigenvectors directions, the centre will be dislocated by $-\eta_i/2$, which is the maximum value of Equation 5.33. From the analysis of the figure, we may say that the asymmetric factor is more pronounced according to the second eigenvector of higher variance since a considerable distortion of the exponential function is evident in the negative direction of this eigenvector. This distortion ensures that the area of higher concentration of points situated at the bottom of the plot are better adjusted by the polynomial function of this centre. Due to the almost symmetric distribution of the points along the direction of the eigenvector with higher eigenvalue, u_1, the effect of the asymmetric factor on the X-axis

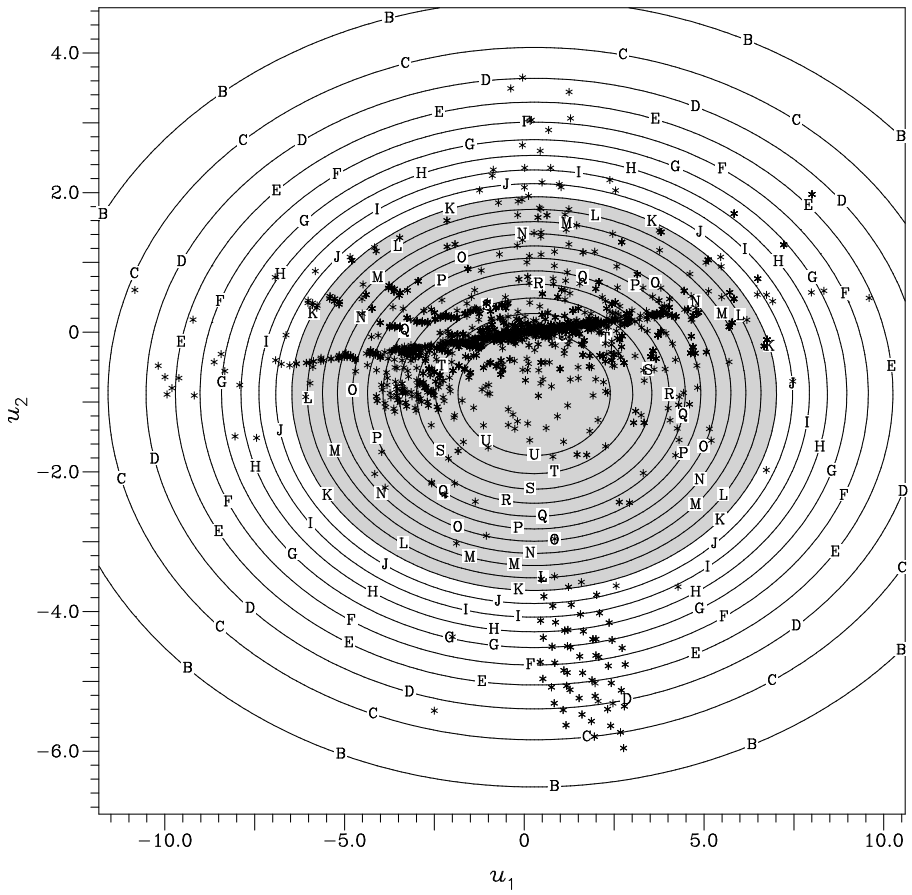


Figure 5.3: Contour plot of the decay function, $\Theta(\mathbf{R})$, in the vicinity of the equilibrium minimum according to the two principal eigenvectors with the highest eigenvalues. The contour increment is 0.05 starting at 0.05 (letter B) up to 1.

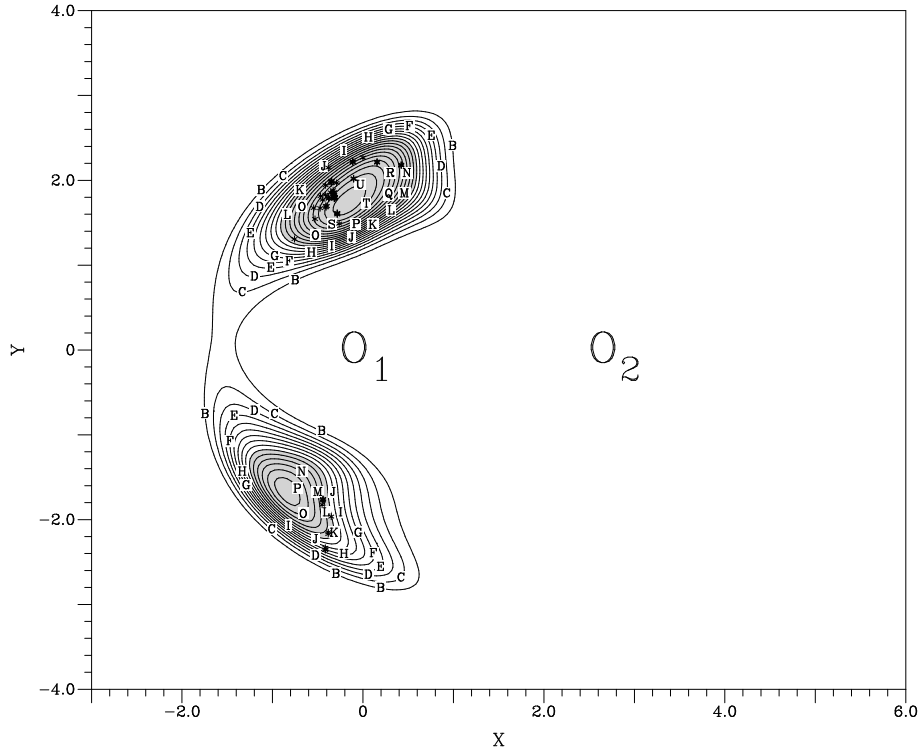


Figure 5.4: Contour plot of the exponential function, $\Theta(\mathbf{R})$, centred at the global minimum. See text for a description.

is not clear in the figure.

Figure 5.4 shows a contour plot of the exponential function $\Theta(\mathbf{R})$ centred at the H_2O_2 equilibrium minimum. To encompass this reference geometry, we fixed one hydrogen atom, H_2, and the two oxygen atoms, O_1 and O_2, at their positions at the equilibrium geometry of the hydrogen peroxide. The other hydrogen atom, H_1, moves in the plane of the figure. The dihedral angle H_1OOH_2 is 113.70° when this atom is above the oxygen atoms ($Y > 0.0$) and 66.40° otherwise. As we can see in the figure, this polynomial centre covers mostly the points located in the region of the equilibrium structure but it also extends to the region close to the *cis*-conformer.

5.2 Fitting procedure

The potential parameters were optimized by means of a least-squares fit. Twenty one reference geometries were selected and in each one of them was centred a polynomial function of the fourth degree, resulting in the adjustment of a total of $21 \times 75 = 1575$ parameters. Due to the fact that some energetic configurations of the molecule fall into intermediate regions covered by more than one polynomial function in conjunction to

the large size of the matrix for the least-squares polynomial fit, the fitting procedure has been addressed using the singular value decomposition method. We chose the subroutine DGLESS from the Linear Algebra Package, LAPACK [48], which has proved to be robust and fast.

Once we are mainly interested in the description of the regions accessible to the reactants involved in chemical reactions that take place in the ground state PES, see Equation 5.1, page 61, high lying configurations with energies above -0.12000 E_h were weighted with 0.1, being only used to guide the polynomial functions. For the chemically interesting regions, special weights were introduced as follows. To accomplish a good characterization of the configuration space in the vicinity of the equilibrium structure and the correspondent *cis* and *trans* configurations it was assigned the weights of 100. In addition, to fit the bottom well accurately, a weight of 10 000 has been attributed to energies lower than -0.425E_h and 100 000 to energies lower than -0.4255E_h.

To prevent an irregular behaviour of the four-body term in configurations far from the interaction region, we include 80 000 additional points, 20 000 for each one of the four dissociation channels. The energy for each configuration has been assigned to the sum of the energies of the one-, two- and three-body terms with the four-body dipole-dipole interaction. This ensures that the short four-body term will approach zero in the fitting procedure. Their geometries have been randomly generated, weighting the configuration of the fragments according to their probability at 10 000 K. A weigh of 1 has been ascribed to these points and to all the remaining *ab initio* points, except for the ones that fall far from the regions covered by the polynomial centres, to which were assigned a weight of 0.01.

Using these conditions, the final PES has a root mean square deviation of less than 0.50 kcal mol^{-1} to the fitted energies.

Chapter 6

Brief description of the H_2O_2 ($X\ ^1A$) PES

In this chapter the most important features of the potential energy surface are scrutinized through the study of the transition state and minimum geometries and energies, prediction of barrier heights, representation and analysis of contour plots for different regions of the potential surface, and one-dimensional cuts along selected coordinates. The quality of the results is evaluated having as a standard reliable *ab initio* calculations from literature and, whenever possible, from experiment.

6.1 Dissociation channels

Due to the many-body formalism, the potential energy surface must reproduce the dissociation energies for all possible combinations of atom-diatom, atom-triatom and diatom-diatom products. Table 6.1 summarizes the energies for the different dissociation limits. The distances of the main coordinates, those that move away as the reaction evolves towards the formation of products, are kept at a fixed distance of 10.00 Bohr. As for the remaining internuclear distances, the bond lengths are restricted to the values of the equilibrium structures of the correspondent fragments. The results obtained from the fitted potential function show that the values of the energies of the different fragments at those distances of 10.00 Bohr are a few tenth of mE_h lower than the asymptotic limit as it should be expected for these distances. This demonstrates that all the terms, other than the four-body interaction term, are accurately determined and our analytical function dissociates correctly to any of the possible reaction channels that can occur in the ground state surface of the system.

Table 6.1: Asymptotic potential energies for the dissociation products of the ground state potential of the H_2O_2 molecule at distances of 10.00 Bohr for the breaking bonds. Distances are given in Bohr and energies in Hartree.

Channels	R_1	R_2	R_3	R_4	R_5	R_6	This work	Asymptotic
OH +OH	10.00	10.00	1.834	1.834	10.00	10.00	-0.34061	-0.34023
H + HO_2	2.526	10.00	1.859	10.00	3.527	10.00	-0.27882	-0.27817
O + H_2O	10.00	2.861	1.810	10.00	10.00	1.810	-0.29876	-0.29845
O_2 + H_2	2.297	1.401	10.00	10.00	10.00	10.00	-0.33002	-0.32997

6.2 Bottom well

Due to the importance and peculiarities of the hydrogen peroxide system previously mentioned, one of the most important and well studied regions of its potential surface is the region of the global minimum and torsional potential mainly by spectroscopic [67] and tunneling splitting [70] studies. For this purpose, there is a considerable amount of literature focused on high quality *ab initio* calculations [66] performed for the local description of the equilibrium structure and surrounding area, which provide a good standard for assessing the quality of our work.

Table 6.2 represents geometry structures for the equilibrium minimum and for the *cis* and *trans* transition states estimated from our fitted analytical potential. Bond lengths are given in Bohr and the angles in degrees. The parameters so obtained are compared with the most accurate ones determined by Malyszek and Koput [66] at the CCSD(T)/aug-cc-pV7Z level of theory. The potential energy barriers for internal rotation correspond to the energy difference, in cm^{-1}, between the *cis/trans* isomers and the equilibrium structure.

Comparing the equilibrium geometry obtained in our work with the experimental data whose structure parameters are represented in Table 4.1, we may highlight that the theoretical results slightly overestimate the values of the experimental parameters for the O-O and O-H bond lengths in 0.7% and 0.1%, respectively. In respect to the $\angle$OOH and the dihedral angles both are underestimated approximately by 2.6% and 1.2%.

A different set of experimental data for the geometry of the global minimum was provided by Koput in an earlier paper [105]. For the O-O and O-H distances the values determined were 2.767 a_0 and 1.824 a_0, respectively. As for the angles, the estimates gave $\angle$OOH = 99.40° and $\angle$HOOH = 111.80°. Although our results are, in general, in more close agreement with the Koput experimental parameters, it must be enhanced that the experimental data may be questionable since the two sets diverge considerably one from each other. The floppiness of the hydrogen peroxide hinders the unambiguous determination of the internal parameters of the molecule. Thus, determining those parameters are

Table 6.2: Comparison between theoretically predicted geometries and energies for the global minimum and for the *cis* and *trans* transition states obtained in this work with those obtained by Malyszek and Koput [66]. Torsional barrier heights are also represented.

	This work	Malyszek and Koput [66]
Equilibrium Configuration (C_2)		
$R_1(a_0)$	2.770	2.741
$R_3(a_0)$	1.829	1.820
$\alpha(^\circ)$	99.73	100.10
$\omega(^\circ)$	112.30	112.75
Trans configuration		
$R_1(a_0)$	2.797	2.761
$R_3(a_0)$	1.822	1.818
$\alpha(^\circ)$	97.99	98.53
$\omega(^\circ)$	180.00	180.00
$\Delta E(cm^{-1})^a$	330.80	368.50
Cis configuration		
$R_1(a_0)$	2.826	2.761
$R_3(a_0)$	1.821	1.821
$\alpha(^\circ)$	103.42	104.35
$\omega(^\circ)$	0.00	0.00
$\Delta E(cm^{-1})^b$	2429.87	2506.30

[a]Energy difference between the *trans* conformation and the equilibrium structure.
[b]Energy difference between the *cis* conformation and the equilibrium structure.

only possible by assuming a fixed value for one coordinate, usually the OH distance, and deriving the other coordinates accordingly.

In the absence of more precise experimental information, we restricted the comparison to the results determined by Malyszek and Koput [66]. The differences between both theoretical estimates are relatively small, being of the order of 1.1% for the O-O and 0.4% for the O-H bond lengths and for the OOH and dihedral angles. This outcome shows that our fitted geometry is in consonance with the more accurate *ab initio* calculations in these internal coordinates.

The same limitations may be extended to the comparison of the fitted results determined for the *cis* and *trans* molecular parameters, as well as for the heights of the torsional barriers at those conformations with the experimental values. Since the experimental data [105, 106] available for both transition states is derived from one-dimensional cuts of the Hamiltonian and do not take into account the amplitude of the vibration of the other coordinates of the system, they are crude estimates to the correspondent potential energy barriers and they should not constitute a pattern to which our results may be directly compared.

At this point, it is important to emphasize that the constructed PES can be considered as having a five-zeta quality with errors less than $1\,\mathrm{kcal\,mol^{-1}}$. Since we are looking for a global potential function accurate enough to yield reliable results toward dynamic calculations for all possible dissociation channels, we only intend to achieve a good description of the minimum and vicinity areas of the configuration space and not the spectroscopic quality of the calculations we are comparing our results to.

The predicted values for the heights of the torsional barriers for the two conformers of the hydrogen peroxide are close to those calculated at the CCSD(T)/aug-cc-pV7Z level. The barrier $\Delta E(cis)$ was determined to be $2429.87\,\mathrm{cm^{-1}}$ comparing with $2506.30\,\mathrm{cm^{-1}}$ from Reference [66] corresponding to a energy difference of about $76.43\,\mathrm{cm^{-1}}$. As for the $\Delta E(trans)$ the energy diverges in about $37.70\,\mathrm{cm^{-1}}$. Also, the fitted structure parameters are considered to be reasonable, presenting deviations of about 2.4% for R_1 and, 0.9% for $\angle$OOH in the case of the *cis* configuration. As for the *trans* isomer the deviations are of 1.3% for R_1, 0.2% for R_3, and 0.6% for $\angle$OOH.

Table 6.2 shows that, as the torsion angle varies from 0.00° to 180.00°, the remaining parameters change considerably. The major difference between our results and the ones obtained in Reference [66] appears for the values for the O-O and O-H bond distances. Comparing these distances for the global minimum and the *trans* configuration, our bond distances are slightly higher although the behaviour is similar, since for both theoretical predictions the distances are higher for the *trans* isomer. As for the *cis* configuration, unlike the results predicted by Malyszek and Koput [66], our estimate shows a considerable

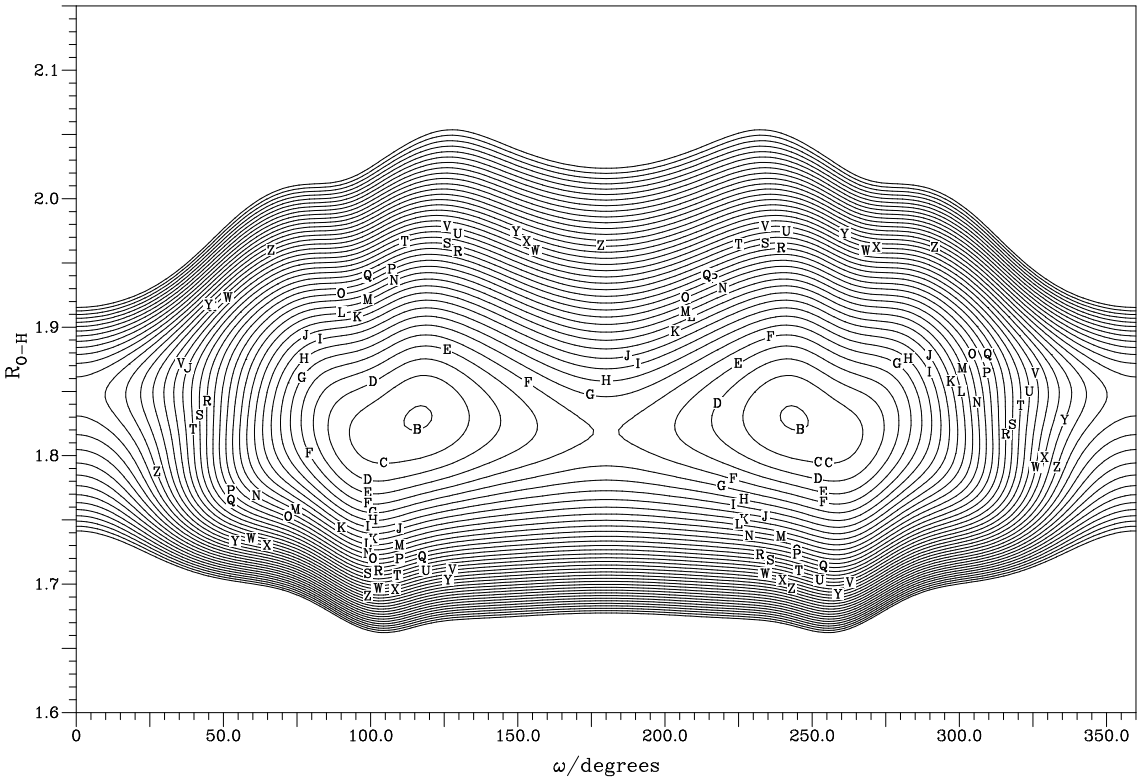


Figure 6.1: Contour plot for the dependence of the OH distances with the diahedral angle for $R_1 = 2.816\,a_0$ and $\angle$OOH = 100.00°. Countours start at -0.4258E_h and are equally spaced by $100\,\mathrm{cm}^{-1}$.

increase in the O-O bond length and a decrease in the O-H distance in comparison with the other two isomers. Our values for the OOH angles compare quite well with the most accurate ones. It's important to note that this internal parameter is the one that varies more with the dihedral angle, exhibiting a variation around 6% when it goes from the *trans* to the *cis* configurations. These observations relative to the distance and angle variations are expected due to conformation constrains when both hydrogen atoms are close to each other.

The dependence of the O-H bond on the dihedral angle is depicted in Figure 6.1, where, for clarity, we display all the 360.00° of rotation.

We fixed $R_{OO} = 2.816\,a_0$ and $\angle$OOH = 100.00° and as a consequence, the minimum and saddle points do not correspond exactly to those represented in Table 6.2.

A better overview of the torsional potential can be seen by a one-dimensional cut along the ω coordinate. This is depicted in Figure 6.2, where, for each dihedral angle, the O..O, O..H bond lengths and the OOH angles have been partially optimized near the equilibrium configuration. The analysis of the potential energy plot as a function of the dihedral angle demonstrates, once more, the good agreement of the predicted H_2O_2 torsion motions with

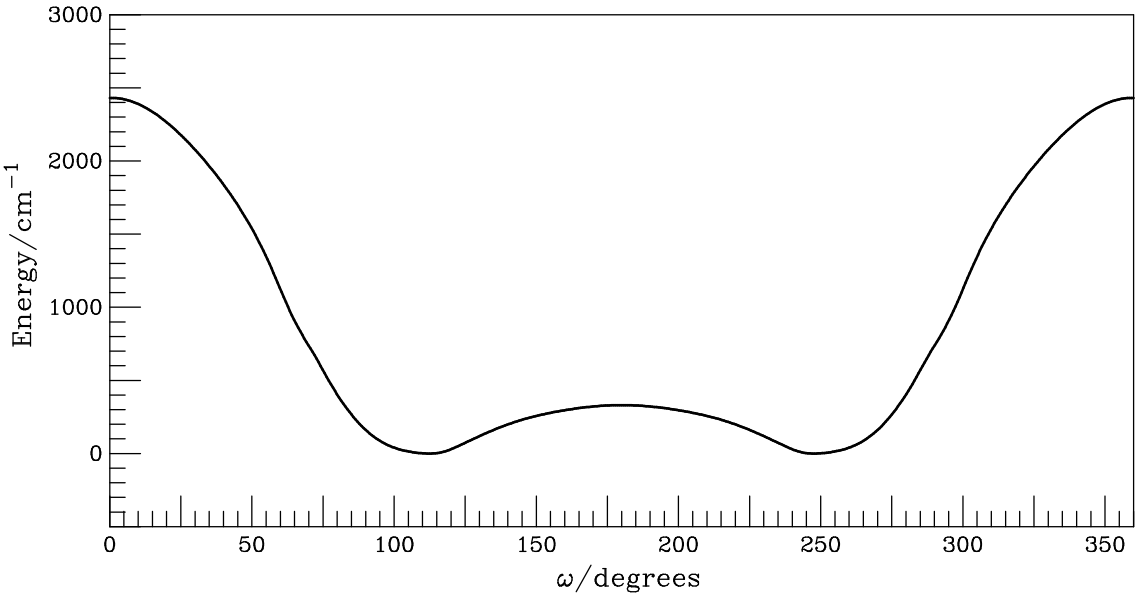


Figure 6.2: Torsional potencial shape, obtained from a one-dimensional cut along the torsional coordinate, and correspondent energy values for the *cis* and *trans* barriers in cm^{-1}.

the theoretical results from literature [66]. The fitted potential properly reproduces the features of the minimum region as it goes from 2429.87 cm^{-1} for the energy barrier of the *cis* conformer to approximately 330.80 cm^{-1} for the energy barrier of the *trans* conformer. The energies of the barriers are given as the difference between the energy obtained at 0.00 ° for the case of the *cis* structure and at 180.00 ° for in the case of *trans* structure, and the value of the potential at the equilibrium minimum, which, for simplicity, we defined in the figure as the zero energy.

6.3 Stationary points

The characterization of the most relevant stationary points of the potential surface is also an important tool to analyse its quality by comparing those structures with those predicted in literature. Table 6.3 gives an overall view of the geometry parameters of the stationary points, others than the equilibrium minimum and the *cis* and *trans* conformers, found for our potential function. These parameters are compared with those predicted by other authors [57, 78]. The geometric structures for these stationary points obtained by *ab initio* optimization at the CCSD(T) level of theory with an aug-cc-pVTZ basis set are represented in Figure 4.2, page 48.

One of the local minimum structures of the H_2O_2 potential energy hypersurface is referred to as oxywater (H_2OO). The existence of this structural isomer of hydrogen

Table 6.3: Geometry parameters of the stationary points found in the global ground state analytical potential of the H_2O_2 system. Distances are in Bohr and the angles are given in degrees.

Geometry parameters	local minimum	isomerization	ts $H + HO_2$	ts $O_2 + H_2$
R_1	2.892	3.375	2.423	2.396
R_2	-	-	-	2.399
R_3	1.849	2.390	1.713	1.869
R_6	3.729	1.888	3.084	-
$\angle O_2O_1H_1$	101.43	50.98	118.73	124.95
$\angle O_2O_1H_2$	101.43	96.74	111.17	-
$\angle O_1H_1H_2$	-	-	-	109.24
$\angle H_1O_1O_2H_2$	105.50	93.04	-	-
$\angle O_2O_1H_1H_2$	-	-	-	63.06
E/E_h	-0.35229	-0.34892	-0.28452	-0.28208
ΔE^a	46.08	48.20	88.60	90.13

[a]Energy in kcal mol^{-1} related to the global minimum of the hydrogen peroxide molecule (-0.42572 E_h).

peroxide is supported in literature. In 1983, Pople et al. [107] conducted several *ab initio* calculations to evaluate the existence of the singlet oxywater. The most accurate results performed at the MP4/6-31G** level, including single up to quadruple corrections, reveal that this minimum structure was unstable due to the absence of energy barrier associated to the reaction $H_2OO \rightarrow HOOH$. These results were, however, refuted in a set of papers that followed. Bach and co-workers [108], in 1990, predicted an energy barrier for the H_2O_2 [1,2]-hydrogen shift of 6.3 kcal mol^{-1} (3.1 kcal mol^{-1} considering the zero-point vibrational energy (ZPVE or ZPE) correction) applying a similar *ab initio* approach but adopting MP2 optimized structures. Two years later, in 1922, Schaefer et al. [109] corroborated the later results by estimating a barrier of 5.7 kcal mol^{-1} (3.2 kcal mol^{-1} with ZPE) at the CCSD(T)/TZ2P+f level of theory. They also brought solid support concerning the formation of oxywater through a [1,2]-hydrogen shift of the hydrogen peroxide, a mechanism that was previously studied by Bach et al. in 1991 [110].

More recent *ab initio* studies from González and co-workers [78] obtained at the CASPT2//CASSCF level reveal a much lower energy barrier of about 0.20 kcal mol^{-1} not only by comparison with the above mentioned values but also with the 2.83 kcal mol^{-1} obtained by the same authors [77] in an earlier paper at PUMP4//UMP2 level of theory.

Table 6.3 shows that our results for the geometric parameters of the oxywater (see Table 4.2, page 48, for the optimized structure and Table 6.3 for the results obtained

from the fitted function) compare quite nicely with the ones obtained by González at the CASSCF/cc-pVTZ level, where $R_{OO} = 2.978\,a_0$, $R_{OH} = 1.836\,a_0$, $\angle OOH = 104.86°$ and $\angle HOOH = 106.75°$, and with those obtained by Schaefer et al. [109], $R_{OO} = 2.982\,a_0$, $R_{OH} = 1.841\,a_0$, $\angle OOH = 105.80°$ and $\angle HOOH = 107.86°$ at the CCSD(T)/DZP level of theory. For the H_2OO-HOOH isomerization we obtained an energy of $2.11\,\mathrm{kcal\,mol^{-1}}$, not considering the ZPE effects.

The transition state structure for the reaction $H_2OO \rightarrow HOOH$ was also characterized. Similar to Schaefer et al. [109] results, the outcome values for the parameters support the occurrence of a [1,2]-hydrogen shift. The theoretical optimization as well as the values obtained from the fitted function, predicted an O-O distance of about $3.074\,a_0$ and $3.375\,a_0$, respectively. Both distances have a higher length compared with those from the corresponding hydrogen peroxide isomers. Also, as expected for a [1,2]-hydrogen shift, the O-H distance increases as the hydrogen atom migrates from one oxygen atom to the other.

We also optimized a transition state which, as stated by González et al. [78], represents the evolution of the system from the oxywater minimum up to the formation of a hydroperoxyl radical. The energies at the CASPT2/cc-pVTZ level of calculation placed this structure $1.10\,\mathrm{kcal\,mol^{-1}}$ (including ZPE) above the H + HO_2 dissociation channel. Our optimization based on the CCSD(T)/aug-cc-pVTZ level predicted an energy for this structure of $7.76\,\mathrm{kcal\,mol^{-1}}$, above the H + HO_2 dissociation energy, but after the extrapolation/scaling procedure the fitted potential gives a transition state $3.98\,\mathrm{kcal\,mol^{-1}}$ below dissociation, which indicates that this reaction is barrierless in this potential energy surface.

Similar results were achieved for the transition state that connects the hydroperoxyl radical to the formation of H_2 + O_2. Our most accurate results predicted an energy $2.45\,\mathrm{kcal\,mol^{-1}}$ below reactants, which advocates a barrier free reaction not considering zero-point vibrational correction to the energy values.

From what we explained above, one of the most important reactions of the ground potential surface of the hydrogen peroxide system is the $O(^1D)$ + H_2O reaction, which can occur via two distinct mechanisms. One of the possibilities is a nondirect insertion mechanism where the $O(^1D)$ atom approaches a water molecule from the oxygen atom side to produce a singlet oxywater [57, 78], followed by an isomerization process which involves the migration of a hydrogen atom from one oxygen atom to the other to yield the hydrogen peroxide molecule. Another way for the $O(^1D)$ + H_2O reaction to proceed is through the abstraction of an H atom of the water molecule by the $O(^1D)$ atom to provide the two OH radical products.

In a first attempt to characterize the ground state potential of the $O(^1D)$ + H_2O

system, González and co-workers [77] predicted, by employing Møller-Plesset methods, a transition state for the formation of oxywater from reactants. The same authors [78] used multireference methods to test the validity of the former report and concluded that indeed, the oxywater is formed directly from reactants without the formation of any intermediate species. This conclusion was later sustained by Head et al. [57] who examined the $O(^1D)$ + $H_2O \rightarrow HOOH$ reaction path at the CASSCF/aug-cc-pVTZ level.

Based on calculations performed at the CASSCF/cc-pVTZ [78] and CASPT2/aug-cc-pVTZ [57] levels of theory, the abstraction mechanism seems to proceed barrier free.

Further support to the absence of energy barrier, in at least one of those reaction channels, is given by the experimental information [111] since, for a temperature range between 200 and 350 K, it was not found any rate constant temperature dependence.

A contour plot of our fitted ground state potential energy surface for the approach of an oxygen atom to a water fragment fixed at its equilibrium structure is illustrated in Figure 6.3.

The two reaction possibilities are depicted in the plot. The first one, on the left side of the plot, corresponds to an insertion mechanism where an oxygen atom attacks the water molecule from the opposite side of the hydrogen atoms, leading to the formation of the local oxywater minimum. The other approach, on the right side of the plot, represents the abstraction of a hydrogen atom of the water molecule by the $O(^1D)$ atom, which leads directly to the abstraction products, OH + OH. The analysis of the above figure reveals that there is no barrier in the insertion mechanism but a small barrier of $3\,\mathrm{kcal\,mol^{-1}}$ in the abstraction one.

6.4 Other regions

The nature of the interaction potential between two OH radicals has been the subject of many studies for a long time. Previous works [62, 77, 78] supported the existence of a nonplanar transition state yielding the formation of two hydroxyl radicals from hydrogen peroxide. Due to the open shell character of these species, CASPT2 and MRCI levels of calculations were employed to optimize it. In both, we were unable to locate the saddle point due to convergence problems. This point was also not found in our fitted potential.

Figure 6.4 displays a contour plot for the approach of two OH fragments positioned with the two parallel HOO angles fixed at 90.00° and with a dihedral angle fixed at 113.00°. An "approximation" to the $O_2(a\ ^1\Delta_g)$ + $H_2(X\ ^1\Sigma_g^+)$ reaction channel is also shown in the figure y-axis. We say "approximation" because in this plot both O-O and H-H distances are kept fixed at an equal value. As inferred from the figure there is no

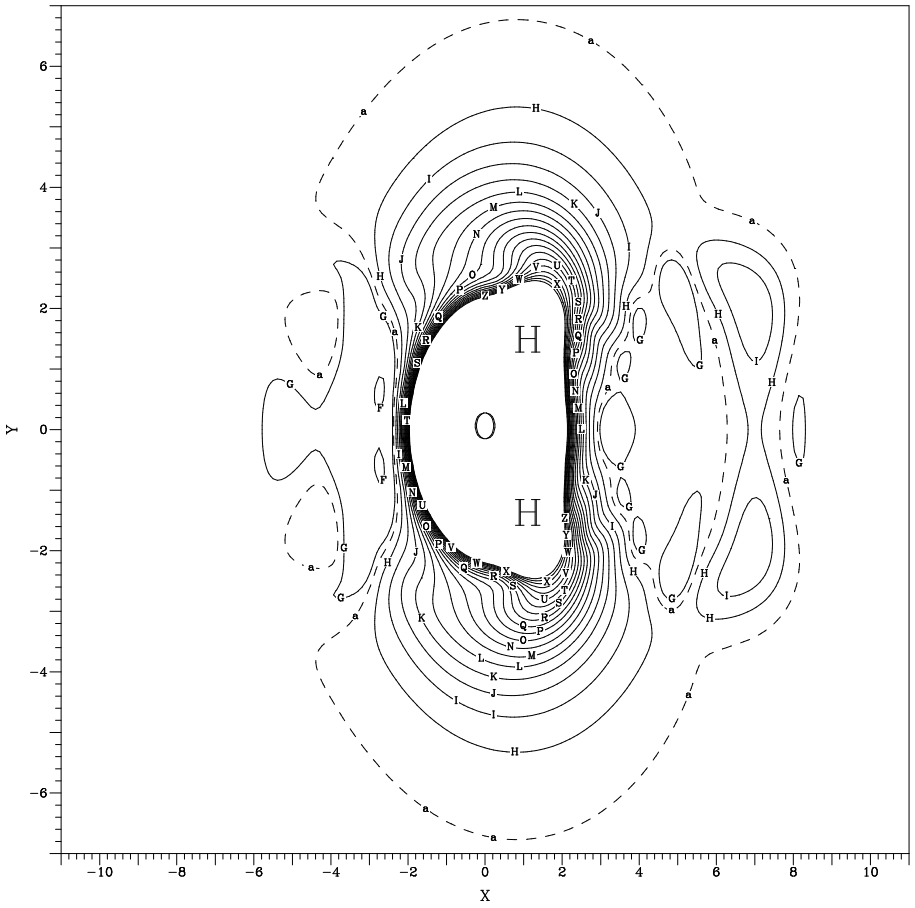


Figure 6.3: Contour plot for the $O(^1D)$ + H_2O planar interaction potential where an oxygen atom is moving around an H_2O molecule kept fixed at its equilibrium geometry. Contours start at -0.425E_h and are equally spaced by 0.02E_h. The dashed line represents the dissociation energy for the reaction (-0.29845E_h).

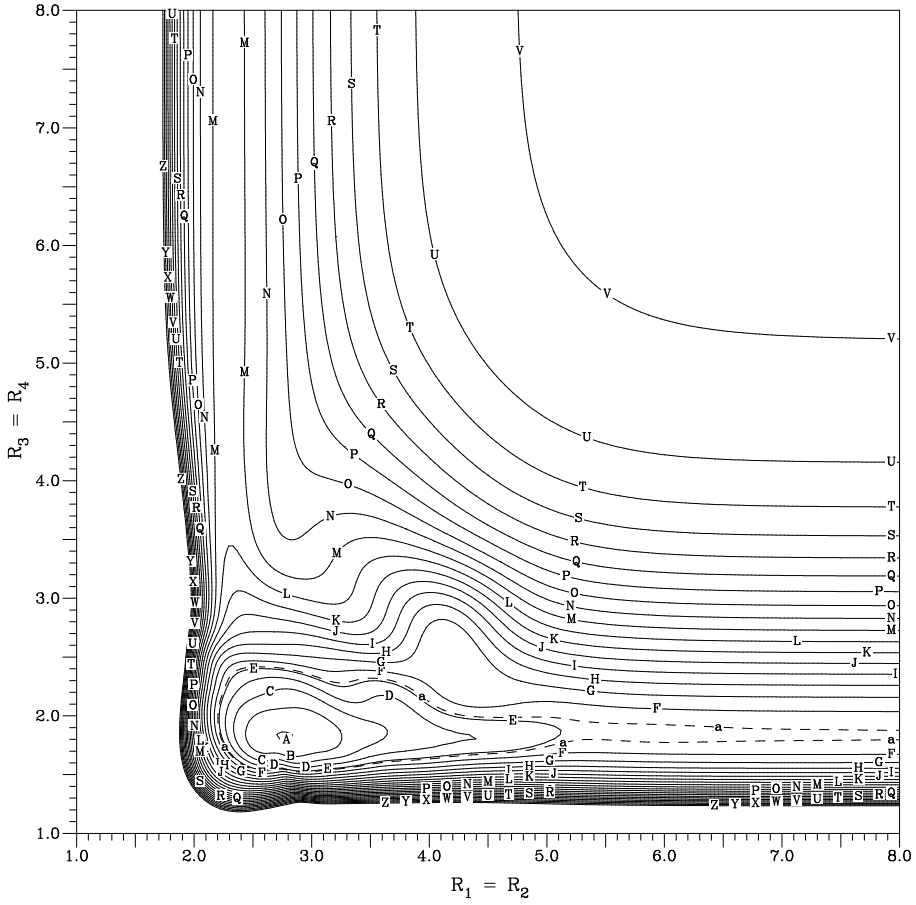


Figure 6.4: Contour plot for the approach of two OH radicals with parallel HOO angles of 90.00 and a dihedral angle of 113.00°. The "approximation" to the $O_2(a\ ^1\Delta_g)$ + $H_2(X\ ^1\Sigma_g^+)$ product channel is also represented in the y-axis for equal variations of the OH distances. The contours start at $-0.425E_h$, being equally spaced by $0.02E_h$. The dashed line indicates the OH + OH dissociation energy ($-0.34023E_h$).

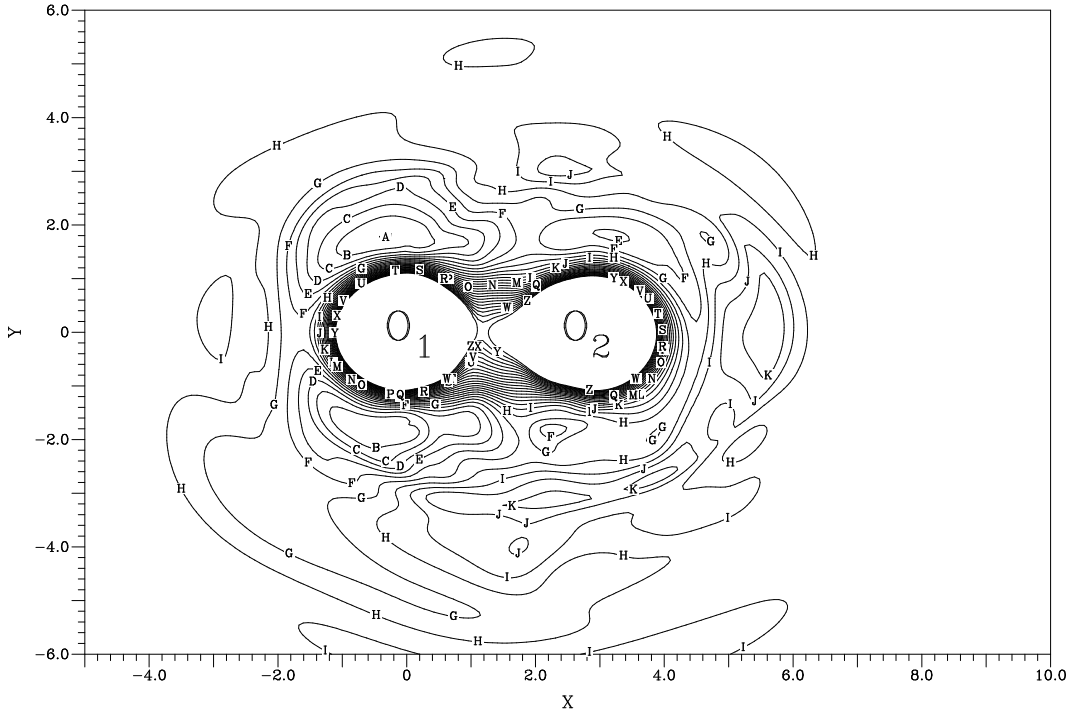


Figure 6.5: Contour plot of a hydrogen atom moving around HOO fragment at the experimental equilibrium geometry of the hydrogen peroxide. The contour plot increment is $0.02\,E_\mathrm{h}$, starting at $-0.425\,E_\mathrm{h}$. This panel shows the full DMBE potential energy surface.

energy barrier for the interaction of two OH radicals.

Another interesting perspective of the analytical potential is given in Figure 6.5, which shows an energy diagram of a hydrogen atom moving in the plan around an HOO molecule fixed at the experimental equilibrium geometry of the hydrogen peroxide (see Table 4.1). The hydrogen of the HOO fragment linked to the O_2 is fixed above the fourth quadrant of the plan of the two oxygen atoms forming a dihedral angle of 113.70° with the top positive plan of the plot. For a clear understanding of the contribution of the short-range four-body terms the corresponding contour plot for the $V^{(3)}_{\mathrm{H_2O_2}}$ and $V^{(4)}_{elect}$ body level of the potential energy surface is also displayed in Figure 6.6.

On comparing both graphics the short-range four-body term stands out for providing considerable refinements to the $V^{(3)}_{\mathrm{H_2O_2}}$ and $V^{(4)}_{elect}$ body level of the potential energy surface. The introduction of this term lowers the energy of the potential exacerbating the region of the global minimum of the H_2O_2, which is located at the left top side of the plan. The torsion potential effect is also prominent by the presence of another minimum in the lower left side of the plan. The minimum represented by letter F corresponds to the oxywater structure.

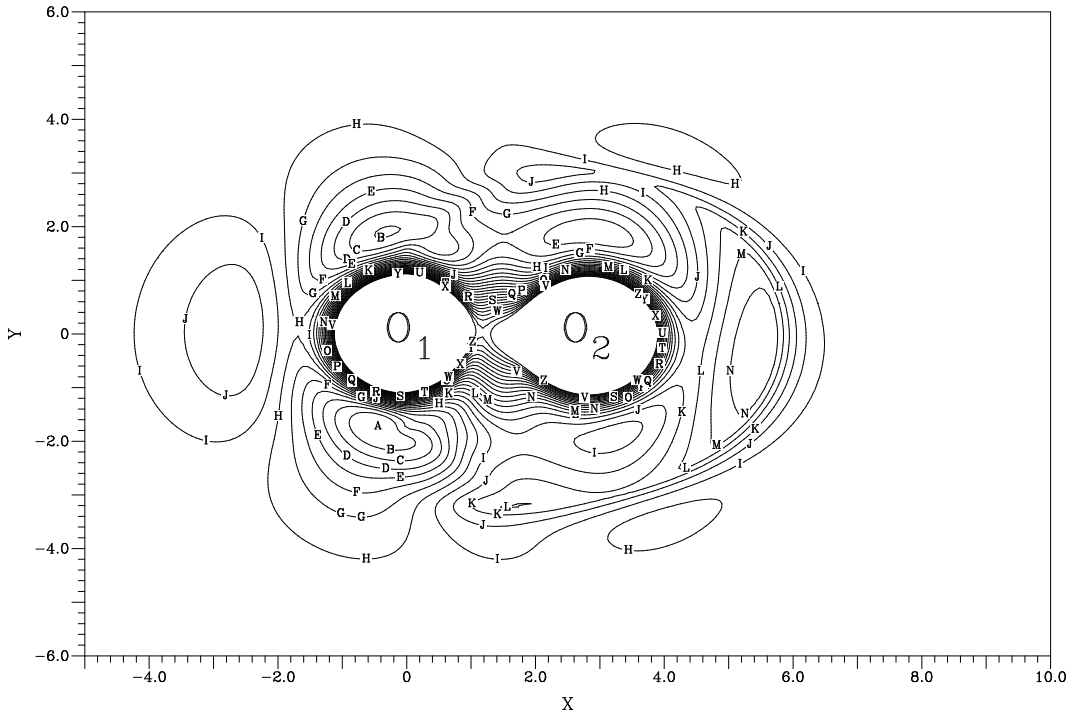


Figure 6.6: Contour plot of a hydrogen atom moving around HOO fragment with the geometry parameters fixed at the experimental equilibrium minimum of the hydrogen peroxide molecule (R_1=2.751 a_0, R_3=1.827 a_0, $\angle$OOH = 102.32° and $\angle$HOOH = 113.70°). The contour plot increment is 0.02 E_h, starting at $-0.425\,E_\mathrm{h}$. This panel shows the $V^{(3)}_{\mathrm{H_2O_2}}$ and $V^{(4)}_{elect}$ body level of the potential energy surface.

Chapter 7

Conclusions

A global ground state potential energy surface for the H_2O_2 system has been reported with chemical accuracy in order to provide reliable results for dynamical calculations of all the possible dissociation channels. In what follows, a brief description of the stages for the construction of the analytical potential surface is given together with a summary of the main conclusions for the present study.

To properly describe all the reaction channels that can take place in the singlet state PES of the hydrogen peroxide molecule, we used a symmetric 3×3 potential matrix wherein each diabatic term is represented by a many-body expansion. One-, two- and three-body terms were adopted from previous potential energy surfaces developed by others namely from the H_2O ($^1A_1$), H_2O ($^3A'$) and HO_2 ($^2A''$) potential energy surfaces [85–87].

The treatment of the four-body terms implied a split into long- and short-range interactions. For the characterization of the former, only the electrostatic interactions between two different pairs of OH molecules were taken into consideration.

As for the short-range four-body energies, we first started by performing a large set of *ab initio* calculations to cover all the most important regions of the configurational space with the resource to the CCSD(T) and CASPT2 levels of theory and employing Dunning-type basis sets. An extrapolation/scaling scheme has been proposed to achieve a high-accurate global PES for the ground state of the hydrogen peroxide from less expensive basis set calculations. For this purpose, calculations were conducted with triple-zeta and quadruple-zeta basis sets at all geometries plus quintuple-zeta only at a few reference geometries. The Hartree-Fock and CASSCF energies were then extrapolated toward the CBS limit through the application of a term obtained by interpolation from the different pivot geometries. Similarly, the same reference structures were used to obtain an interpolated factor to perform the correlation energies scaling up to the five-zeta quality.

Thus, the short-range four-body term is obtained by fitting the difference between the extrapolated energies and the sum of the four-body long-range interactions with the lowest eigenvalue of the 3×3 potential matrix.

After performing the *ab initio* calculations and the extrapolation/scaling procedure, the next step was to develop an analytical potential function invariant to the permutation of identical atoms. To do so, we first applied the Molien theorem to determine the dimension of the integrity basis and the number of terms of each degree that can be used to build the invariant polynomial function. The next step was to define the integrity basis through the application of projection operators, one for each irreducible representation of the D_2 point group, to which our permutation group is isomorphic. The invariant polynomial function were constructed as combinations of the seven integrity basis terms.

The fitting procedure involved a multifunctional fit in which several permutation invariant polynomial functions were centred at different reference geometries in a total of twenty one structures. Eleven of these reference geometries correspond to transition states of the surface as well as additional geometries in specific areas namely on the $O(^1D)$ + $H_2O(X\ ^1A_1)$ reaction channel. The *ab initio* points were then assigned to each of these reference geometries and ten additional ones based on a k-means algorithm. Due to the importance of the first eleven references to attain a good description of the surface, they are kept fixed during the algorithm process. As for the additional ten references, they were randomly generated and optimized, so that more than three hundred *ab initio* points were assigned to each one of them. To restrict each polynomial to the region of its centre, we multiply it by a range factor defined as a product of exponential functions centred at its own reference geometry.

For simplicity, fourth degree polynomial functions were used in all the regions of the PES. Since we defined 21 reference centres, a total of $21 \times 75 = 1575$ parameters were adjusted using the single value decomposition method implemented in the subroutine DGLESS, from the Linear Algebra Package, LAPACK.

By relying on the aforementioned conditions, it was possible to derive a six-dimensional potential energy hypersurface for the singlet hydrogen peroxide molecule with a root mean square deviation for the fitted energies of less than $1\,\mathrm{kcal\,mol^{-1}}$, which fulfils the requirement of chemical accuracy.

At an overall global level, the present surface of five-zeta quality found solid support in *ab initio* results previously reported by others. Nevertheless, there are several ways to further improve the surface. One of them consists in refining the global minimum area and surrounding configurational space to achieve spectroscopic precision in order to compute the vibration-rotation spectra for the electronic ground state of the H_2O_2 system. This may be accomplished by improving this region fit or by adopting a technique similar to

the one described in the work by Brandão and co-workers [87] for the HO_2 molecule.

The direct application of this global potential and also a way to assess its reliability is to perform dynamic studies for all possible dissociation channels. Preliminary studies on the dynamic of the $O(^1D)$ + $H_2O(X\ ^1A_1)$ reaction channel revealed, in consonance with the results obtained by González et al. [77], that contributions from excited potential energy surfaces must be considered in order to achieve a rate constant in good agreement to the experimental value.

Bibliography

[1] I. N. Levine, *Quantum Chemistry*, Prentice Hall, New Jersey, 4th edn. (1991).

[2] M. Born and J. R. Oppenheimer, Ann. Phys. **84**, 457 (1927).

[3] D. A. McQuarrie, *Quantum Chemistry*, University Science Books, California (1983).

[4] S. Bubin, M. Pavanello, W.-C. Tung, K. L. Sharkey and L. Adamowicz, Chem. Rev. **113**, 36 (2013).

[5] D. G. Truhlar, *The Encyclopedia of Physical Science and Technology*, vol. 13, Academic Press, New York (2001).

[6] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Chichester (1999).

[7] A. J. C. Varandas, Adv. Chem. Phys. **74**, 255 (1988).

[8] J. N. Murrell, Structure and Bonding **32**, 93 (1977).

[9] A. J. C. Varandas and P. Piecuch, Chem. Phys. Lett. **430**, 448 (2006).

[10] C. E. Dykstra, *Advances in Molecular Electronic Structure Theory - Calculation and Characterization of Molecular Potential Energy Surfaces*, vol. 1, JAI Press Ltd. (1990).

[11] H. B. Schlegel, WIREs Comput Mol Sci **1**, 790 (2011).

[12] E. G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of the Molecular and Quantum Mechanics*, Springer, 3rd edn. (2011).

[13] I. N. Levine, *Quantum Chemistry*, Prentice-Hall, Inc. (2000).

[14] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Macmillan, New York (1982).

[15] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[16] P. W. Atkins and R. S. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, Oxford (1997).

[17] L. Gagliardi and B. O. Roos, Chem. Soc. Rev. **36**, 893 (2007).

[18] T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic Structure Theory*, no. ISBN 0 471 96755 6, John Wiley & Sons Ltd (2000).

[19] H. J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).

[20] J. W. Hollet and P. M. W. Gill, J. Chem. Phys. **134**, 114111 (2011).

[21] R. J. Bartlett and M. Musial, Rev. Mod. Phys. **79**, 291 (2007).

[22] J. C. Slater, Phys. Rev. **36**, 57 (1930).

[23] S. F. Boys, Proc. R. Soc. (London) A **200**, 542 (1950).

[24] I. Mayer, *Simple Theorems, Proofs and Derivations in Quantum Chemistry*, Springer Science+Business Media, New York (2003).

[25] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).

[26] T. H. D. Jr., J. Chem. Phys. **53**, 2823 (1970).

[27] W. J. Hehre, R. F. Stewart and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).

[28] R. Ditchfield, W. J. Hehre and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).

[29] P. C. Hariharan and J. A. Pople, Theoretica Chimica Acta **28**, 213 (1973).

[30] M. J. Frisch, J. A. Pople and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

[31] J. T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[32] D. E. Woon and T. H. D. Jr., J. Chem. Phys. **103**, 4572 (1995).

[33] R. A. Kendall, T. H. D. Jr. and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[34] T. Helgaker and P. R.Taylor, *Modern Electronic Structure Theory - Part II*, vol. 2, World Scientific, Singapore (1995).

[35] S. F. Boys and F. Bernardi, Molec. Phys. **19**, 553 (1970).

[36] J. S. Wright and S. K. Gray, J. Chem. Phys. **69**, 67 (1978).

[37] P. J. Kuntz, in *Dynamics of Molecular Collisions*, edited by W. Miller, 53, Plenum, New York (1976).

[38] D. M. Hirst, *Potential Energy Surfaces*, Taylor & Francis, London (1985).

[39] J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley and A. J. C. Varandas, *Molecular Potential Energy Functions*, Wiley, Chichester (1984).

[40] K. S. Sorbie and J. N. Murrell, Mol. Phys. **31**, 905 (1976).

[41] A. J. C. Varandas, Chem. Phys. Lett. **194**, 333 (1992).

[42] A. J. C. Varandas, Int. Rev. Phys. Chem. **19**, 199 (2000).

[43] G. C. Schatz, *Advances in Molecular Electronic Structure Theory - The analytic representation of the potential energy surfaces for chemical reactions.*, vol. 1, JAI Press Ltd., London (1990).

[44] A. J. C. Varandas, Mol. Phys. **53**, 1303 (1984).

[45] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, Inc., New York, 3rd edn. (1998).

[46] D. Kalman, The College Mathematics Journal **27**, 1 (1996).

[47] W. H. Press, S. A. Teukolsky, W. T. VerterliNG and B. P. Flannery, *Numerical Recipes*, Cambrige Univrersity Press, Cambridge, 2nd edn. (1992).

[48] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen, *LAPACK User's Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 3rd edn. (1999).

[49] B. Lewis and G. von Elbe, *Combustion, Flames and Explosion of Gases*, Academic Press Inc., New York, 2nd edn. (1961).

[50] N. D. C. Allen, G. G. Abad, P. F. Bernath and C. D. Boone, Journal of Quantitative Spectroscopy & Radiative Transfer **115**, 66 (2013).

[51] D. D. Davis, Can. J. Chem. **52**, 1405 (1974).

[52] T. Encrenaz, T. K. Greathouse, F. Lefèvre and S. K. Atreya, Planetary and Space Science **68**, 3 (2012).

[53] P. Berhman, B. Parise, R. Liseau, B. Larsson, H. Olofsson, K. M. Menten and Güsten, Astronomy & Astrophysics **531**, L8 (2011).

[54] J. R. Wiesenfeld, Acc. Chem. Res. **15**, 110 (1982).

[55] J. Troe, Combust. Flame **158**, 594 (2011).

[56] A. R. Hendrix, C. A. Barth, A. I. F. Stewart, C. W. Hord and A. L. Lane, Lunar and Planetary Institute Conference Abstrats 2043 (1999).

[57] G. Yingbin, K. Olsen, R. I. Kaiser and J. D. Head, Astrochemistry, From Laboratory Studies to Astronomical Observations **855**, 253 (2006).

[58] W. Zheng, D. Jewitt and R. I. Kaiser, The Astrophysical Journal **648**, 753 (2006).

[59] X. Luo, P. R. Fleming and T. R. Rizzo, J. Chem. Phys. **96**, 5659 (1992).

[60] T. M. Ticich, H. R. Dübal and F. F. Crim, J. Chem. Phys. **84**, 1508 (1986).

[61] L. B. Harding, J. Phys. Chem. **93**, 8004 (1989).

[62] L. B. Harding, J. Chem. Phys. **95**, 8653 (1991).

[63] J. Koput, Chem. Phys. Lett. **236**, 516 (1995).

[64] J. Koput, S. Carter and N. C. Handy, J. Phys. Chem. **102**, 6325 (1998).

[65] J. Koput, S. Carter and N. C. Handy, J. Chem. Phys. **115**, 8345 (2001).

[66] P. Malyszek and J. Koput, J. Comput. Chem. **34**, 337 (2013).

[67] O. L. Polyansky, I. N. Kozin, R. I. Ovsyannikov, P. Malyszek, J. Koput, J. Tennyson and S. N. Yurchenko, J. Phys. Chem. A **177**, 7367 (2013).

[68] B. Kuhn, T. R. Rizzo, D. Luckhaus, M. Quack and M. A. Suhm, J. Chem. Phys. **111**, 2565 (1999).

[69] B. Fehrensen, D. Luckhaus and M. Quack, Chem. Phys. Lett. **300**, 312 (1999).

[70] B. Fehrensen, D. Luckhaus and M. Quack, Chem. Phys. **338**, 90 (2007).

[71] R. Chen, G. Ma and H. Guo, Chem. Phys. Lett. **320**, 567 (2000).

[72] R. Chen, G. Ma and H. Guo, J. Chem. Phys. **114**, 4763 (2001).

[73] S. Y. Lin and H. Guo, J. Chem. Phys. **119**, 5867 (2003).

[74] S. Carter, A. R. Sharma and J. M. Bowman, J. Chem. Phys. **135**, 014308 (2011).

[75] J. Troe and V. G. Ushakov, Phys. Chem. Chem. Phys. **10**, 3915 (2008).

[76] A. Bakasov, R. Berger, T.-K. Ha and M. Quack, Int. J. Quantum Chem. **99**, 393 (2004).

[77] R. Sayós, C. Oliva and M. Gonzálex, J. Chem. Phys. **113**, 6736 (2000).

[78] R. Sayós, C. Oliva and M. González, J. Chem. Phys. **115**, 8828 (2001).

[79] R. Sayós, C. Oliva and M. Gonzálex **75**, 279 (2000).

[80] G. Pelz, K. M. T. Yamada and G. Winnewisser, J. Mol. Spectrosc. **159**, 507 (1993).

[81] A. J. C. Varandas, J. Chem. Phys. **113**, 8880 (2000).

[82] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[83] A. J. C. Varandas, Chem. Phys. Lett. **443**, 398 (2007).

[84] A. J. C. Varandas, J. Chem. Phys. **127**, 114316 (2007).

[85] J. Brandão and C. M. A. Rio, J. Chem. Phys. **119**, 3148 (2003).

[86] J. Brandão, C. Mogo and B. C. Silva, J. Chem. Phys. **121**, 8861 (2004).

[87] J. Brandão, C. M. A. Rio and J. Tennyson, J. Chem. Phys. **130**, 134309 (2009).

[88] J. N. Murrell and K. S. Sorbie, J. Chem. Soc., Faraday Trans. 2 **70**, 1552 (1974).

[89] J.-C. Nieh and J. J. Valentini, J. Phys. Chem. **91**, 1370 (1987).

[90] G. C. Maitland, M. Rigby, E. B. Smith and W. A. Wakeham, *Intermolecular Forces*, Clarendon Press, Oxford (1981).

[91] J. Brandão and C. M. A. Rio, Chem. Phys. Lett. **372**, 866 (2003).

[92] A. J. C. Varandas and J. Brandão, Mol. Phys. **45**, 857 (1982).

[93] A. J. C. Varandas, J. Mol. Struct. (Theochem) **120**, 401 (1985).

[94] T. Molien, Sitz. Königl. Preuss. Akad. Wiss. 1152–1156 (1897).

[95] A. G. McLellan, J. Phys. C.: Solid St. Phys. **12**, 753 (1979).

[96] A. Schmelzer and J. N. Murrell, Int. J. Quantum Chem. **28**, 287 (1985).

[97] H. Weyl, *The Classical Groups*, Princeton Mathematical Series, New Jersey (1946).

[98] H. Derksen and G. Kemper, *Computational Invariant Theory*, vol. 130, Springer, New york (2002).

[99] P. R. Bunker and P. Jensen, *Molecular Symmetry and Spectroscopy*, NRC Research Press, Ottawa, Ontario, Canada, 2nd edn. (1998).

[100] S. Farantos, E. C. Leisegang, J. N. Murrell, K. Sorbie, J. J. C. Teixeira-Dias and A. J. C. Varandas, Molec. Phys. **34**, 947 (1977).

[101] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, New York (2006).

[102] K. B. Lipkowitz and D. B. Boyd, *Reviews in Computational Chemistry*, vol. 18, John Wiley & Sons, Inc., New York (2002).

[103] K. R. Žalik, Phys. Rev. Lett **29**, 1385 (2008).

[104] K. S. Sorbie and J. N. Murrell, Mol. Phys. **29**, 1387 (1975).

[105] J. Koput, J. Mol. Spectrosc. **115**, 438 (1986).

[106] J.-M. Flaud, C. Camy-Peyret, J. W. C. Johns and B. Carli, J. Chem. Phys. **91**, 1504 (1989).

[107] J. A. Pople, K. Raghavachari, M. J. Frisch, J. S. Binkley and P. V. R. Schleyer, J. Am. Chem. Soc. **105**, 6389 (1983).

[108] R. D. Bach, J. J. W. McDouall, A. L. Owensby and H. B. Schlegel, J. Am. Chem. Soc. **112**, 7065 (1990).

[109] C. Meredith, T. P. Hamilton and H. F. Schaefer, J. Phys. Chem. **96**, 9250 (1992).

[110] R. D. Bach, A. L. Owensby, C. Gonzales, H. B. Schlegel and J. J. W. McDouall, J. Am. Chem. Soc. **113**, 6001 (1991).

[111] R. Altkinson, D. L. Baulch, R. A. Cox, R. F. Hampson, Jr., J. A. Kerr and J. Troe, J. Phys. Chem. Ref. Data **21**, 1125 (1992).