Università degli Studi di Ferrara, Italy
Université Paul Sabatier, Toulouse, France

# Theory and applications of advanced techniques in quantum chemistry and their integration in a common infrastructure

Coordinators/Directeurs
Prof. Renzo CIMIRAGLIA
Prof. Stefano EVANGELISTI

Candidate/Candidat
Dr. Stefano BORINI

Year 2006

# Contents

# Acknowledgements

After three years of work spent in the development of this Ph.D. thesis, many steps toward the integration between the localization technique and the n-electron valence state perturbation theory have been made, paving the way for a powerful tool in applied quantum chemistry. Many other steps lie ahead, much more integration and coordination is needed, but in the long term every bit of knowledge will be integrated in a common, reliable and powerful meta-code, an important platform for theoretical and experimental investigations.

Chemistry is everywhere. New drugs and insights in biological processes, new materials for improving storage, computation, communication in a fast changing world. In developing new theories and computational approaches, we must keep in mind that the very high potential of these new tools still has to be completely unleashed. Once this happens, we will be able to perform a new kind of research.

I wish to thank my supervisors Prof. Renzo Cimiraglia, Prof. Stefano Evangelisti and also Dr. Daniel Maynau and Dr. Celestino Angeli for providing the passionate environment needed to develop a satisfying research.

I wish to thank Dr. Elda Rossi, Dr. Antonio Monari and Prof. Gianluigi Bendazzoli for the useful discussions to produce Q5Cost and F90xml libraries.

I wish to thank Dr. Mariachiara Pastore for the strong moral support and help during the Ph.D.

I wish to thank Prof. Maurizio Dal Colle for the help and hospitality.

I wish to thank Dr. Alex Cavallini, Dr. Manuel Sparta, Dr. Mirko Cestari, Dr. Lara Ferrighi for the fantastic research experience.

I wish to thank Prof. Gilli, for allowing me to discover new insights in my career.

Finally, I wish to thank my Family, Paola Marani and Walter Padovani, for the help in achieving all the professional steps needed to obtain this Ph.D.

# Preface

This Ph.D. thesis does not target a single technique: two quantum chemistry techniques are studied separately and then analyzed in order to combine them into a common infrastructure.

The main problem in applicability of *ab initio* methods to medium and large molecular systems is the very steep increase of the computational cost in terms of time, memory and disk space, with respect to the size of the molecule. Many different techniques have been developed to deal with this problem, ranging from improving algorithms to adding levels of approximation. In many cases changing the theoretical approach has proved successful, like in Density Functional Theory methods, but leading to other new interesting problems and limitations.

Since long time, the best approach to deal with this complexity has been the well known *Divide et Impera*: splitting a very large problem in smaller, more manageable problems. Along with this strategy goes the "right tool for the job" philosophy, leading to a number of techniques created to provide appropriate tools at different levels of theory, each one with its own strenghts and weaknesses, collaborating together to deal with the problem in synergic way.

The first technique presented in this work is the Freeze-and-Cut localized optimization, developed and implemented at Paul Sabatier University in Toulouse. The main advantage of this technique is to provide a purely *ab initio* evaluation on medium and large molecular systems, achieving at the same time saving of computational cost and partitioning of a large system in subsystems.

The second technique here presented is the n-Electron Valence state Perturbation Theory (NEVPT), a perturbative treatment developed at Ferrara University in partnership with Paul Sabatier University in Toulouse. The main advantage of this technique is the very accurate evaluation of spectroscopical properties of the studied chemical systems. Unfortunately NEVPT can be applied only to small molecular systems, being limited by the feasibility of Complete Active Space (CAS) evaluations.

Combining these two theories allows the computational chemist to perform a first low-level analysis with Freeze-and-Cut on the less critical parts of the molecule under study, and to perform a CASSCF plus NEVPT evaluation on the chemically interesting part, where a more detailed and high-level treatment is needed. This objective presents both theoretical and implementative hurdles. Refinements in the NEVPT theory have been introduced, driven by the need to deal with the localized orbitals approach provided by Freeze-and-Cut. New and improved standards for file formats and information exchange have been developed to satisfy the future needs of a common infrastructure, where the potentiality of

the methods can be further combined, used, and improved by other research laboratories.

This Ph.D. thesis is organized as follows:

1. The First chapter addresses the basic knowledge needed for the subsequent chapters

2. The Second chapter is devoted to localization techniques, focusing on theory and software implementation. A deep insight in the Freeze-and-Cut technique will follows, along with obtained results

3. The Third chapter presents NEVPT theory and applications on various carbonyl-based molecules. Also, various implementations of this theory will be discussed, such as Quasi-Degenerate and Non-Canonical approaches, needed in order to work with a localized set of orbitals

4. The Fourth chapter presents aspects and problems related to integration in a common grid infrastructure, a more manageable file format and a glance ahead toward the future of quantum chemistry computing. Particular focus will be granted to F77/F90xml, a C/Fortran binding library to parse the XML file format with Fortran, and Q5Cost, a purely Fortran 95 library to read and write a common file format with an intention-revealing interface on the *ab initio* quantum chemistry domain.

# Resumé chapitre 1
# Introduction

Ce premier chapitre concerne les notions de base pour les développements suivants.

La première section parle du théorème variational. Ce théorème dit que si $\epsilon$

$$\epsilon = \frac{\left\langle \tilde{\Psi} \left| \hat{\mathcal{H}} \right| \tilde{\Psi} \right\rangle}{\left\langle \tilde{\Psi} \middle| \tilde{\Psi} \right\rangle}$$

est l'énergie associée à une fonction d'onde approchée $\tilde{\Psi}$, et $E_0$ est l'énergie valeur propre de l'état fondamental $\Psi_0$, $E_0$ est une limite inférieure pour $\epsilon$. Pour un état fondamental non dégénérée, chaque fonction d'onde différente de la fonction d'onde $\Psi_0$ donne une énergie plus haute que l'énergie $E_0$.

La deuxième section parle du développement perturbatif de Rayleigh-Schrødinger dans le cas général. Ce développement obtient l'énergie avec la correction d'une fonction d'onde bien connue, la fonction d'onde al'ordre zéro en perturbation. Cette fonction est obtenue comme solution d'un Hamiltonien modèle $\hat{\mathcal{H}}_0 = \hat{\mathcal{H}} - \hat{V}$

$$\hat{\mathcal{H}}_0 \Psi_n^{(0)} = E_n^{(0)} \Psi_n^{(0)} \quad n = 0, 1, 2, \ldots$$

pour donner la correction à l'énergie au deuxième ordre

$$E_n^{(2)} = \sum_k \frac{\left| \left\langle \Psi_n^{(0)} \left| \hat{V} \right| \Psi_k^{(0)} \right\rangle \right|^2}{E_n^{(0)} - E_k^{(0)}}$$

La théorie perturbative Møller-Plesset au deuxième ordre (MP2) est le développement plus connu de cette approche.

En suite, on parlerà de l'importante nécessité pour la fonction d'onde d'être antisymétrique pour l'echange d'un couple d'électrons. Par exemple, dans le cas d'une fonction d'onde à deux electrons, on aura

$$\Psi(2,1) = -\Psi(1,2)$$

Cette loi naturelle est absolument fondamentale pour la validité physique d'une fonction d'onde.

Une nouvelle entité, le déterminant de Slater $\Phi$, a été rationalisée pour travailler avec une fonction d'onde antisymétrique par construction. Les déterminants de Slater représentent l'occupation des orbitales monoélectroniques, et ils peuvent être utilisés pour décrire la fonction d'onde vrai avec une combinaison

linéaire des tous les déterminants de Slater qui peuvent être créés avec un certain nombre d'orbitales monoelectroniques et d'électrons. Si la combinaison linéaire comprend tous les déterminants de Slater, une fonction d'onde Full-CI a été créée

$$\Psi = \sum_I c_I \Phi_I$$

qui décrit le mieux résultat quel'on peut obtenir avec une base d'orbitales monoelectroniques. Le coût computationnel nécessaire pour travailler avec une fonction d'onde Full-CI est très élevé, et donc des approximations doivent être appliquées.

Une première approximation est le choix d'un seul déterminant de Slater, $\Psi = \Phi_0$, et on utilise une procédure d'optimisation qui va améliorer les orbitales monoelectroniques de $\Phi_0$. Cette approche, bien connue sous le nom de technique Hartree-Fock (HF) ou Self Consistent Field (SCF), marche bien seulement dans les cas des molécules à la géométrie d'équilibre avec une couche fermé d'électrons.

Si on veut travailler avec des états excités, ou des distorsions de géométrie, l'approximation d'Hartree-Fock est insuffisante, et plusieurs déterminants de Slater doivent être considérés dans l'expansion lineaire. Dans ce cas, une fonction multiréférence est développée. La limite est l'expansion Full-CI, mais des méthodes alternatives de sélection des déterminants de Slater existent pour réduire le coût computationnel et avoir particulier propriété sur la fonction finale.

Par exemple, la sélection Complete Active Space (CAS) partage les orbitales dans trois groupes: core, actif et virtuelles. Les orbitales core sont toujour occupées en toutes les déterminants de Slater sélectionnés, les orbitales virtuelles sont toujour vides, et les orbitales actifs ont toutes les occupations possibles pour un certain nombre d'électrons actifs dans les orbitales actifs.

Une algèbre particuliére a été développée pour travailler avec les déterminants de Slater, appellée deuxième quantification. Les règles principales sont présentées dans cette thèse: deux operateurs particuliérs ont été décrits: l'operateur de création $a_i^+$ crée un électron dans l'orbitale $i$ du déterminant de Slater, et l'operateur de destruction $a_i$ détruit un électron dans l'orbitale.

Enfin, des approches d'optimisation de la fonction d'onde sont présentées: l'approche Newton-Rhapson (NR) et l'approche Super-CI. L'approche Newton-Rhapson se base sur le développement de Taylor de l'énergie.

L'approche Super-CI, utilisée dans cette thèse pour optimiser une fonction d'onde localisée, se base sur le théorème de Brillouin pour les cases multiréférence: l'interaction entre une fonction d'onde optimisée et ses excitations monoélectroniques via l'hamiltonien sont nuls. L'advantage de l'approche Super-CI est qui converge toujour à un minimum. Au contraire, l'approche Newton-Rhapson peut converger vers un point de selle.

# Chapter 1

# Introduction

## 1.1 Variational principle

Let $\hat{\mathcal{H}}$ be the true Hamiltonian, and $\tilde{\Psi}$ an arbitrary approximated wavefunction of the true wavefunction $\Psi_0$ for the ground state of a non-degenerate case. The energy $\epsilon$ associated to this function is given by the equation

$$\epsilon = \frac{\left\langle \tilde{\Psi} \middle| \hat{\mathcal{H}} \middle| \tilde{\Psi} \right\rangle}{\left\langle \tilde{\Psi} \middle| \tilde{\Psi} \right\rangle} \tag{1.1.1}$$

Under the variational theorem, the energy $E_0$ (eigenvalue of the true ground state $\Psi_0$) is a lower bound for the energy $\epsilon$ associated to the approximated wavefunction $\tilde{\Psi}$.

$$\epsilon \geq E_0 \quad \forall \tilde{\Psi} \quad , \quad \epsilon = E_0 \Leftrightarrow \tilde{\Psi} = \Psi_0 \tag{1.1.2}$$

Proof of this theorem can be found in quantum chemistry books (for example Ref. 1, 2). The theorem guarantees that given a generic function $\tilde{\Psi}$, the energy evaluated by means of Eqn. 1.1.1 will be greater, or at best equal than the true ground state energy, with the equality holding if the function $\tilde{\Psi}$ is the exact function for the electronic state. It is possible to parametrize $\tilde{\Psi}$ through a set of parameters, and to optimize them by minimizing the difference between $\epsilon$ and the true energy $E_0$.

Given the linear expansion of $\tilde{\Psi}$ as a linear combination of a set of functions $\Phi_i$

$$\tilde{\Psi} = \sum_i c_i \Phi_i \tag{1.1.3}$$

the minimization condition to satisfy is

$$\frac{\partial \epsilon}{\partial c_i} = 0 \quad \forall i \tag{1.1.4}$$

supposing real $c_i$ coefficients being part of a $\mathbf{c}$ array we obtain

$$\begin{aligned} \epsilon &= \frac{\sum_{i,j} c_i c_j \left\langle \Phi_i \left| \hat{\mathcal{H}} \right| \Phi_j \right\rangle}{\sum_{i,j} c_i c_j \langle \Phi_i | \Phi_j \rangle} \\ &= \frac{\mathbf{c}^+ \mathbf{H} \mathbf{c}}{\mathbf{c}^+ \mathbf{S} \mathbf{c}} \end{aligned} \tag{1.1.5}$$

Performing a differentiation of the energy with respect to the $\mathbf{c}$ vector we obtain the final minimization condition

$$\mathbf{H}\mathbf{c} = \epsilon \mathbf{S}\mathbf{c} \tag{1.1.6}$$

The $\mathbf{c}$ vector represents the linear combination of the $\Phi_i$ set producing the lowest possible value for the associated energy. An extension of the variational theorem also asserts that each solution $\epsilon_i$ obtained from Eqn. 1.1.6 satisfies the same rule for the energy $E_i$ (MacDonald's theorem).

## 1.2 Perturbation theory

Let $\hat{\mathcal{H}}$ be the true Hamiltonian and let us suppose it can be expressed in the form

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \lambda \hat{V} \tag{1.2.1}$$

where $\hat{\mathcal{H}}_0$ is a model Hamiltonian, $\lambda$ is a parametric value expressing the intensity of the perturbative effect and $\hat{V}$ is a perturbative operator.

Supposing a non-degenerate case and that all eigenvalues and eigenvectors of the model Hamiltonian $\hat{\mathcal{H}}_0$ are known (therefore knowing its spectral decomposition) we have, at zero order of perturbation

$$\hat{\mathcal{H}}_0 \Psi_n^{(0)} = E_n^{(0)} \Psi_n^{(0)} \quad n = 0, 1, 2, \ldots \tag{1.2.2}$$

the $\Psi_n^{(0)}$ eigenfunctions ($|n\rangle$ in shorter notation) define a complete basis set of wavefunctions. If no perturbative effects are present, the descriptive wavefunction for the state $n$ will be $\Psi_n^{(0)}$ with $E_n^{(0)}$ as associated energy.

Imposing a perturbation, the descriptive wavefunction and the associated energy will change as a function of the perturbative parameter $\lambda$. We can perform a Taylor's expansion to obtain

$$\Psi_n = |n\rangle + \lambda \Psi_n^{(1)} + \lambda^2 \Psi_n^{(2)} + \ldots \tag{1.2.3}$$

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \ldots \tag{1.2.4}$$

We suppose that the eigenfunctions of $\hat{\mathcal{H}}_0$ are normalized ($\langle n|n\rangle = 1$) and we choose the normalization of $\Psi_n$ such that $\langle n|\Psi_n\rangle = 1$ (*intermediate normalization*). This normalization can always be performed when $|n\rangle$ and $\Psi_n$ are not

orthogonal, simplifying the treatment and leading to

$$\left\langle n \middle| \Psi_n^{(k)} \right\rangle = 0 \quad k = 1, 2, 3, \ldots \tag{1.2.5}$$

Given the final relationship to solve

$$\hat{\mathcal{H}}\Psi_n = E_n\Psi_n \tag{1.2.6}$$

a substitution can be performed to obtain

$$\begin{aligned} & \lambda^0 \left( \hat{\mathcal{H}}_0 |n\rangle - E_n^{(0)} |n\rangle \right) \\ + \quad & \lambda^1 \left( \hat{\mathcal{H}}_0 \Psi_n^{(1)} + \hat{V} |n\rangle - E_n^{(0)} \Psi_n^{(1)} - E_n^{(1)} |n\rangle \right) \\ + \quad & \lambda^2 \left( \hat{\mathcal{H}}_0 \Psi_n^{(2)} + \hat{V} \Psi_n^{(1)} - E_n^{(0)} \Psi_n^{(2)} - E_n^{(1)} \Psi_n^{(1)} - E_n^{(2)} |n\rangle \right) \\ + \quad & \ldots = 0 \end{aligned} \tag{1.2.7}$$

In order to satisfy 1.2.7, terms inside parentheses must be zero

$$\hat{\mathcal{H}}_0 |n\rangle = E_n^{(0)} |n\rangle \tag{1.2.8}$$

$$\left( \hat{\mathcal{H}}_0 - E_n^{(0)} \right) \Psi_n^{(1)} = \left( E_n^{(1)} - \hat{V} \right) |n\rangle \tag{1.2.9}$$

$$\left( \hat{\mathcal{H}}_0 - E_n^{(0)} \right) \Psi_n^{(2)} = E_n^{(2)} |n\rangle + \left( E_n^{(1)} - \hat{V} \right) \Psi_n^{(1)} \tag{1.2.10}$$

$$\vdots$$

which also lead to

$$E_n^{(0)} = \left\langle n \middle| \hat{\mathcal{H}}_0 \middle| n \right\rangle \tag{1.2.11}$$

$$E_n^{(1)} = \left\langle n \middle| \hat{V} \middle| n \right\rangle \tag{1.2.12}$$

$$E_n^{(2)} = \left\langle n \middle| \hat{V} \middle| \Psi_n^{(1)} \right\rangle \tag{1.2.13}$$

We can find the first-order correction to the wavefunction by expressing the $\Psi_n^{(1)}$ on the complete basis set defined by the solutions of the zero-order Hamiltonian

$$\Psi_n^{(1)} = \sum_k c_k^{(1)} |k\rangle \tag{1.2.14}$$

It is clear that, since the eigenfunctions of $\hat{\mathcal{H}}_0$ are orthonormal, we can write

$$\left\langle k \middle| \Psi_n^{(1)} \right\rangle = c_k^{(1)} \tag{1.2.15}$$

and then

$$\Psi_n^{(1)} = \sum_k c_k^{(1)} |k\rangle \tag{1.2.16}$$

$$\Psi_n^{(1)} = \sum_k |k\rangle \left\langle k \middle| \Psi_n^{(1)} \right\rangle \tag{1.2.17}$$

We can now multiply 1.2.9 by $\langle k|$

$$\left(E_n^{(0)} - E_k^{(0)}\right) \left\langle k \middle| \Psi_n^{(1)} \right\rangle = \left\langle k \middle| \hat{V} \middle| n \right\rangle \tag{1.2.18}$$

and combining 1.2.17 with 1.2.13

$$E_n^{(2)} = \left\langle n \middle| \hat{V} \middle| \Psi_n^{(1)} \right\rangle = \sum_k \left\langle n \middle| \hat{V} \middle| k \right\rangle \left\langle k \middle| \Psi_n^{(1)} \right\rangle \tag{1.2.19}$$

combining with 1.2.18 leads to the second-order correction to the energy

$$E_n^{(2)} = \sum_k \frac{\left| \left\langle n \middle| \hat{V} \middle| k \right\rangle \right|^2}{E_n^{(0)} - E_k^{(0)}} \tag{1.2.20}$$

## 1.3 Antisymmetry

A natural law imposes that fermionic particles like electrons must be described by a wavefunction which is antisymmetric with respect to the exchange of two particles. For a simple two-electron wavefunction the relationship

$$\Psi(2,1) = -\Psi(1,2) \tag{1.3.1}$$

must be satisfied.

It is always possible to express a polyelectronic wavefunction as a combination of products of one-electron spinorbitals. If $\Psi(1,2,\ldots,N)$ is a N-electrons wavefunction, the spatial position of particles $(2,\ldots,N)$ can be fixed. The resulting function can be expanded on a single-particle basis set $\{\psi_i\}$

$$\Psi(1,\underbrace{2,\ldots,N}_{\text{fixed}}) = \sum_i c_i(2,\ldots,N)\psi_i(1) \tag{1.3.2}$$

where the $c_i$ coefficients hold the dependency against the fixed particles, therefore being functions themselves. The procedure can be repeated for $c_i(2,\ldots,N)$, fixing $(3,\ldots,N)$ and expanding the function on the same basis set

$$c_i(2,\underbrace{3,\ldots,N}_{\text{fixed}}) = \sum_j c'_{ij}(3,\ldots,N)\psi_j(2) \tag{1.3.3}$$

Repeating for the coordinates of $N$ particles, we obtain

$$\Psi(1,2,\ldots,N) = \sum_{i,j,k,\ldots,l} c_{ijk\ldots l}\psi_i(1)\psi_j(2)\psi_k(3)\ldots\psi_l(N) \tag{1.3.4}$$

where $c_{ijk\ldots l}$ is a purely numeric factor. Applying Eqn. 1.3.4 to a two-particles case, we obtain

$$\begin{aligned} \Psi(1,2) &= \sum_{i,j} c_{ij}\psi_i(1)\psi_j(2) \\ &= c_{11}\psi_1(1)\psi_1(2) + c_{12}\psi_1(1)\psi_2(2) + c_{21}\psi_2(1)\psi_1(2) \\ &\quad + c_{22}\psi_2(1)\psi_2(2) + \ldots \end{aligned} \tag{1.3.5}$$

Describing a system with the electrons swapped

$$\begin{aligned}\Psi(2,1) &= c_{11}\psi_1(2)\psi_1(1) + c_{12}\psi_1(2)\psi_2(1) + c_{21}\psi_2(2)\psi_1(1) \\ &\quad + c_{22}\psi_2(2)\psi_2(1) + \dots \end{aligned} \tag{1.3.6}$$

in order to satisfy the antisymmetry rule the following conditions must hold

$$\begin{aligned} c_{ii} &= -c_{ii} \rightarrow c_{ii} = 0 \\ c_{ij} &= -c_{ji} \end{aligned}$$

Therefore, the sum can be reduced to

$$\begin{aligned}\Psi(1,2) &= c_{12}\left(\psi_1(1)\psi_2(2) - \psi_2(1)\psi_1(2)\right) \\ &\quad + c_{13}\left(\psi_1(1)\psi_3(2) - \psi_3(1)\psi_1(2)\right) + \dots \\ &= \sum_{i<j} c_{ij}\left(\psi_i(1)\psi_j(2) - \psi_j(1)\psi_i(2)\right) \\ &\rightarrow \sum_{i<j} c_{ij}\,\|\psi_i\psi_j\| \end{aligned} \tag{1.3.7}$$

where

$$\|\psi_i\psi_j\| = 2^{-\frac{1}{2}} \begin{vmatrix} \psi_i(1) & \psi_j(1) \\ \psi_i(2) & \psi_j(2) \end{vmatrix} \tag{1.3.8}$$

is the normalized *Slater determinant*. The presented scheme can be generalized for a $N$ electron case

$$\Psi(1,2,\dots,N) = \sum_{i_1<i_2<\dots<i_N} c_{i_1,i_2,\dots,i_N}\,\|\psi_{i_1}\psi_{i_2}\dots\psi_{i_N}\| \tag{1.3.9}$$

$$= \sum_I c_I \Phi_I \tag{1.3.10}$$

where

$$\|\psi_{i_1}\psi_{i_2}\dots\psi_{i_N}\| = (N!)^{-\frac{1}{2}} \begin{vmatrix} \psi_{i_1}(1) & \psi_{i_2}(1) & \dots & \psi_{i_N}(1) \\ \psi_{i_1}(2) & \psi_{i_2}(2) & \dots & \psi_{i_N}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{i_1}(N) & \psi_{i_2}(N) & \dots & \psi_{i_N}(N) \end{vmatrix} \tag{1.3.11}$$

The wavefunction can be expressed as a linear combination of Slater determinants, each one describing a possible electronic distribution. Slater determinants satisfy by construction the antisymmetry requisite: exchanging two particles is expressed as a column swap, which changes the sign of the resulting determinant. Another interesting property, Pauli's rule, is a direct consequence of the antisymmetry requisite: if two particles are forced to occupy the same one-electron

function, the representative determinant is

$$\|\psi_{i_1}\psi_{i_1}\dots\psi_{i_N}\| = (N!)^{-\frac{1}{2}} \begin{vmatrix} \psi_{i_1}(1) & \psi_{i_1}(1) & \dots & \psi_{i_N}(1) \\ \psi_{i_1}(2) & \psi_{i_1}(2) & \dots & \psi_{i_N}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{i_1}(N) & \psi_{i_1}(N) & \dots & \psi_{i_N}(N) \end{vmatrix} \tag{1.3.12}$$

which evaluates as zero (two columns are equal).

## 1.4 Second quantization

### Creation and annihilation operators

A frequently used representation for Slater determinants is by spinorbital occupation. An occupation vector $|\mathbf{n}\rangle$ can be defined with elements $n_i$ which evaluate 1 if a given spinorbital is occupied by an electron, 0 if not occupied. Of course, the sum $\sum_i n_i$ evaluates the total number $N$ of electrons in the molecular system.

The occupation vector represents a particular distribution of $N$ electrons in $m$ spinorbitals and belongs to a so-called Fock space $F(m)$, an abstract linear vector space defined only by the occupation number of the spinorbitals.

A particular vector of the Fock space is the vacuum state

$$|\text{vac}\rangle = |0_1, 0_2, 0_3, \dots, 0_m\rangle \tag{1.4.1}$$

and it represents the case where no spinorbital is occupied by electrons. The vacuum vector defines a subspace $F(0,m)$ of the Fock space, spanned by the single vector given above. In general the subspace $F(N,m)$ of the Fock space has a higher dimensionality, being defined by all those vectors obtained by distributing $N$ electrons in $m$ spinorbitals.

For a basis of orthogonal spinorbitals, the product between two occupation vectors $|n\rangle$ and $|k\rangle$ is defined as

$$\langle n|k\rangle = \prod_{i=1}^{m} \delta_{n_i,k_i} \tag{1.4.2}$$

This product represents the overlap between Slater determinants, evaluating to one when the electronic distribution is the same, otherwise zero.

Two operators, named creation and annihilation operators, can now be defined. The creation operator $a_i^+$ is defined as

$$a_i^+ |n_1, n_2, \dots, 0_i, \dots, n_m\rangle = \Gamma(n)_i \, |n_1, n_2, \dots, 1_i, \dots, n_m\rangle \tag{1.4.3}$$

$$a_i^+ |n_1, n_2, \dots, 1_i, \dots, n_m\rangle = 0 \tag{1.4.4}$$

This operator imposes the occupation of a previously unoccupied spinorbital $i$ by changing its occupation number from 0 to 1. If the selected spinorbital $i$ is already occupied, the result is zero. This can be explained knowing that a Slater determinant with the same spinorbital included twice is zero, due to the Pauli exclusion principle. The $\Gamma(n)$ value is called *phase factor* and is defined as

$$\Gamma(n)_i = (-1)^{\sum_{j=1}^{i-1} n_j} \tag{1.4.5}$$

The effect of the phase factor is to change the sign of the resulting determinant if the number of occupied spinorbitals before $i$ is odd, or leaving it unchanged if even.

The annihilation operator $a_i$ removes an electron from an occupied spinorbital, or produces zero if the spinorbital is already unoccupied

$$a_i |n_1, n_2, \ldots, 1_i, \ldots, n_m\rangle = \Gamma(n)_i |n_1, n_2, \ldots, 0_i, \ldots, n_m\rangle \tag{1.4.6}$$

$$a_i |n_1, n_2, \ldots, 0_i, \ldots, n_m\rangle = 0 \tag{1.4.7}$$

The creation operator can be proved to be the adjoint of the annihilation operator, and viceversa.

Important considerations can be derived from these definitions: every vector of the Fock space can be generated from the vacuum state by application of an ordered product of creation operators

$$|\mathbf{n}\rangle = \prod_{i=1}^{N} \left(a_i^+\right)^{n_i} |\text{vac}\rangle \tag{1.4.8}$$

Another important fact is represented by anticommutation relationships between the operators. It can be demonstrated that

$$\left[a_i^+, a_j^+\right]_+ = a_i^+ a_j^+ + a_j^+ a_i^+ = 0 \tag{1.4.9}$$

$$[a_i, a_j]_+ = a_i a_j + a_j a_i = 0 \tag{1.4.10}$$

$$\left[a_i^+, a_j\right]_+ = a_i^+ a_j + a_j a_i^+ = \delta_{ij} \tag{1.4.11}$$

These relationships allow the definition of an efficient algebra for handling determinants, operators and interactions between them.

## Operators in second quantization

First quantization operators such as the kinetic operator or the Hamiltonian operator can be described using the second quantization formalism. The main difference between first and second quantization expression resides in the different structure of the second quantization approach, which incorporates the effect of

the spinorbital basis into the operator itself, and not into the Slater determinant, which holds only information about the occupation.

Operators in second quantization are expressed as a combination of products of elementary creation and annihilation operators. For example, the kinetic operator in first quantization is written as

$$\hat{f}_{1q} = \sum_{i=1}^{N} \hat{f}(x_i) \tag{1.4.12}$$

where $x_i$ represents the spatial and spin coordinates of the electron $i$, and the sum runs over the number of electrons. To express the same operator in second quantization, we should keep in mind that no change in the number of electrons is performed and only one electron at a time is involved. Therefore, the kinetic operator in second quantization is represented as a linear combination of excitation operators

$$\hat{f}_{2q} = \sum_{rs} f_{rs} a_r^+ a_s \tag{1.4.13}$$

where the sum runs over all spinorbitals. The operator is invariant with respect to the number of electrons, since this information is contained into the Fock space vector. The $a_r^+ a_s$ operator effectively moves one electron from spinorbital $s$ to spinorbital $r$.

Following the rules for orthogonality and application of the operators given above, the application of the second quantization $\left\langle n \left| \hat{f} \right| m \right\rangle$ operator produces three cases

1. the $|n\rangle$ and $|m\rangle$ vectors represent the same occupation

$$\left\langle n \left| \hat{f} \right| m \right\rangle = \sum_{r} n_r f_{rr} \tag{1.4.14}$$

2. the $|n\rangle$ and $|m\rangle$ differ in a single occupation pair

$$|n\rangle = |n_1, \ldots, 1_i, \ldots, 0_j, \ldots, n_m\rangle \tag{1.4.15}$$

$$|m\rangle = |n_1, \ldots, 0_i, \ldots, 1_j, \ldots, n_m\rangle \tag{1.4.16}$$

$$\left\langle n \left| \hat{f} \right| m \right\rangle = \Gamma(n_1)_j \Gamma(n_1)_i f_{ij} \tag{1.4.17}$$

3. the $|n\rangle$ and $|m\rangle$ differ in more than one occupation pairs

$$\left\langle n \left| \hat{f} \right| m \right\rangle = 0 \tag{1.4.18}$$

which reproduce the Slater rules between Slater determinants if we choose

$$f_{rs} = \int \psi_r^*(x) \hat{f}(x) \psi_s(x) dx \tag{1.4.19}$$

Therefore, the second quantization representation of a one-electron operator is a linear combination of single excitation operators, where the coefficients are the values of the integrals given above.

The same strategy can be arranged for two-electron operators. A two-electron operator gives non-zero matrix elements between two Slater determinants which differ at most for two pairs of occupation values. Considering again that no change in the number of electrons is performed by the operator, it could be expanded as a linear combination

$$\hat{g} = \frac{1}{2}\sum_{ijkl} g_{ijkl} a_i^+ a_j^+ a_l a_k \tag{1.4.20}$$

where, as usual, the $g_{ijkl}$ elements can be calculated evaluating the matrix element

$$g_{ijkl} = \int \psi_i^*(x)\psi_k^*(x')\hat{g}(x,x') = \langle ij\,|\hat{g}|\,kl\rangle\,\psi_j(x)\psi_l(x') \tag{1.4.21}$$

From these results, a representation of the Hamiltonian operator in second quantization can be devised

$$\hat{\mathcal{H}} = \sum_{ij} h_{ij} a_i^+ a_j + \frac{1}{2}\sum_{ijkl} g_{ijkl} a_i^+ a_j^+ a_l a_k \tag{1.4.22}$$

where $h_{ij}$ is the integral of kinetic energy plus the electron-nuclear operators, and $g_{ijkl}$ is the electron-electron repulsion term.

Summarizing, the differences between first and second quantization operators are:

- In first quantization the Slater determinants depend on the orbital basis, and the operators are independent of this information. In second quantization the operators contain this information, and the Slater determinants are represented only as occupation vectors in the Fock space, without references to the basis set
- In first quantization, the operators depend on the number of electrons. In second quantization, they are independent of this parameter.
- First quantization operators are exact operators, while second quantization operators are projected operators.

### Density matrix

As we know, the expectation value of the energy for a normalized wavefunction can be obtained by

$$E = \left\langle \Psi \left| \hat{\mathcal{H}} \right| \Psi \right\rangle \tag{1.4.23}$$

As we saw in Sec. 1.1, we can define the wavefunction as a linear combination of functions, in our case Slater determinants

$$\Psi = \sum_i c_i \Phi_i \tag{1.4.24}$$

If we substitute we obtain

$$E = \left\langle \sum_i c_i \Phi_i \left| \hat{\mathcal{H}} \right| \sum_j c_j \Phi_j \right\rangle = \sum_{ij} c_i c_j H_{ij} \tag{1.4.25}$$

Where $H_{ij}$ is the matrix element of the Hamiltonian operator between two Slater determinants. We can expand further this expression keeping into account Eqn. 1.4.22

$$E = \sum_{ij} c_i c_j \left( \sum_{ab} \gamma_{ab}^{ij} h_{ab} + \frac{1}{2} \sum_{abcd} \Gamma_{abcd}^{ij} g_{abcd} \right) \tag{1.4.26}$$

where $\gamma_{ab}^{ij} = \langle \Phi_i \left| a_a^+ a_b \right| \Phi_j \rangle$ and $\Gamma_{abcd}^{ij} = \left\langle \Phi_i \left| a_a^+ a_b^+ a_d a_c \right| \Phi_j \right\rangle$. We can remove the dependency on indexes $i$ and $j$ by including the coefficients and defining two entities: the one-particle density matrix

$$\gamma_{ab} = \sum_{ij} c_i c_j \gamma_{ab}^{ij} \tag{1.4.27}$$

and the two-particles density matrix

$$\gamma_{abcd} = \sum_{ij} c_i c_j \Gamma_{abcd}^{ij} \tag{1.4.28}$$

The energy can now be written as

$$E = \sum_{ab} \gamma_{ab} h_{ab} + \frac{1}{2} \sum_{abcd} \gamma_{abcd} g_{abcd} \tag{1.4.29}$$

These matrixes are central in modern quantum chemistry. An interesting property of the one-particle density matrix can be found in the sum of the diagonal elements (trace), which adds to the total number of electrons

$$\mathrm{Tr}(\gamma_{ij}) = N \tag{1.4.30}$$

The eigenvectors of the density matrix are called *natural orbitals*. The corresponding eigenvalues are called *occupation numbers*, and express the electronic occupation of the corresponding natural orbital.

## 1.5 Hartree-Fock approximation and SCF method

A wavefunction can be defined in a space of Slater determinants, each one describing a possible electronic distribution, by creating a linear expansion over this basis

$$\Psi(1,2,\ldots,N) = \sum_{i_1<i_2<\ldots<i_N} c_{i_1,i_2,\ldots,i_N} \|\psi_{i_1}\psi_{i_2}\ldots\psi_{i_N}\| \quad (1.5.1)$$

$$= \sum_I c_I \Phi_I \quad (1.5.2)$$

This wavefunction satisfies the requisite of antisymmetry, and spans the full space of determinants of the Fock space defined by $N$ electrons and an infinite number of spinorbitals. For practical reasons in daily computations, the number of spinorbitals in the expansion is finite, and so is the expansion.

In Sec. 1.1 we reported the possibility to obtain the best description for the ground state of a molecular system, making use of the variational theorem and optimizing a given set of parameters of the wavefunction. Two degrees of freedom are available: the coefficients $c_I$ of the linear expansion (a linear optimization) and the spinorbitals $\psi_i$ (a non-linear optimization).

One possible strategy for the optimization is to work with the complete space of determinants, providing the *Full Configuration Interaction* (Full-CI) description. This approach is often unfeasible due to the very high number of coefficients to optimize. Although special algorithms have been developed to reduce the computational cost, Full-CI scales factorially with respect to the size of the system, and its applicability is limited to small molecules.

On the opposite side, a tough but effective approximation is to consider a single determinant to describe the system. The multideterminant expansion

$$\Psi(1,2,\ldots,N) = \sum_{i_1<i_2<\ldots<i_N} c_{i_1,i_2,\ldots,i_N} \|\psi_{i_1}\psi_{i_2}\ldots\psi_{i_N}\| \quad (1.5.3)$$

$$= \sum_I c_I \Phi_I \quad (1.5.4)$$

is forced to

$$c_I = 0 \qquad \forall I \neq 0$$
$$c_0 = 1$$

leading to

$$\Psi(1,2,\ldots,N) = \Phi_0 \quad (1.5.5)$$

This approach drops any degree of freedom provided by the linear expansion, and focuses on the optimization of the one-electron spinorbitals.

It is possible to show that the best spinorbitals for the single determinant description of a molecular system are the eigenfunctions of a one-electron operator, named *Fock operator*. The procedure goes under the name of *Self Consistent Field*, since the Fock operator itself depends on the spinorbitals, and the solution is built iteratively until autoconsistency is reached. The *Hartree-Fock* equation

$$\hat{\mathcal{F}}\psi_i = \epsilon_i \psi_i \tag{1.5.6}$$

gives a set of one-electron spinorbitals which describe at best (in terms of energy minimization) the electronic situation represented by the single determinant approximation. These spinorbitals are named *canonical*.

The corresponding eigenvalues $\epsilon_i$ are named *orbital energies*, and possess an intrinsic physical meaning, being an approximation of the energy needed for the ionization of the electron occupying the spinorbital $\psi_i$ (Koopmans theorem[3]). Spinorbitals are normally ordered by their respective orbital energies, from the lowest one to the highest one.

On an infinite space of spinorbitals, the Hartree-Fock equation has infinite solutions. In a real case, the number of solutions is limited by the size of the atomic basis set used to describe the spinorbitals, but in any case the lowest total energy is obtained when the first $N$ spinorbitals (where $N$ is the number of electrons) are occupied by electrons. These spinorbitals are therefore named *occupied*, and the remaining empty spinorbitals are named *virtual*.

The Hartree-Fock approximation provides good results when the molecule is evaluated at the equilibrium geometry and electronic ground state, for closed-shell systems. Hartree-Fock theory keeps into account only averaged interactions between electrons. The one-electron nature of the Fock operator aims at optimizing the interactions of each electron with an averaged charge cloud described by the remaining electrons, completely discarding instantaneous interactions.

## 1.6 Brillouin theorem

The Hartree-Fock equation gives a set of optimized spinorbitals which can be occupied by electrons to build the single determinant wavefunction $\Psi = \Phi_0$ with the lowest energy. As a vector of the Fock space, this wavefunction is not the only electronic distribution which can be built with the given set of spinorbitals. Each different electronic distribution can be described as an excitation of a given number of electrons from the occupied orbitals to the virtual ones. We can now reformulate Eqn. 1.5.1 with a different grouping strategy, honoring the excitation level (referred to the $\Phi_0$) of each determinant

$$\Psi = c_0 \left|\Phi_0\right\rangle + \sum_{ir} c_i^r \Phi_i^r + \sum_{ijrs} c_{ij}^{rs} \Phi_{ij}^{rs} + \dots \tag{1.6.1}$$

where $r$ and $i$ denote the excitation of an electron from the $\psi_i$ orbital to $\psi_r$.

To improve the single determinant description, one would be tempted to enrich the description including the singly excited contributions, and to perform a diagonalization of the Hamiltonian inside this subspace of determinants $\{\Phi_0, \Phi_i^r\}$ defined by the Hartree-Fock determinant and its single excitations. This produces

$$\begin{pmatrix} \left\langle \Phi_0 \left| \hat{\mathcal{H}} \right| \Phi_0 \right\rangle & \left\langle \Phi_0 \left| \hat{\mathcal{H}} \right| \Phi_i^r \right\rangle \\ \left\langle \Phi_i^r \left| \hat{\mathcal{H}} \right| \Phi_0 \right\rangle & \left\langle \Phi_i^r \left| \hat{\mathcal{H}} \right| \Phi_{i'}^{r'} \right\rangle \end{pmatrix} \begin{pmatrix} c_0 \\ c_a^r \end{pmatrix} = E_0 \begin{pmatrix} c_0 \\ c_a^r \end{pmatrix} \tag{1.6.2}$$

Brillouin's theorem[4,5] guarantees that when spinorbitals satisfying the Hartree-Fock equation are used to build the determinants, all the interactions $\left\langle \Phi_0 \left| \hat{\mathcal{H}} \right| \Phi_i^r \right\rangle$ are zero. Thanks to the Slater rules, this relationship can be proved to be equivalent to a spinorbital one

$$\left\langle \psi_a \left| \hat{\mathcal{F}} \right| \psi_i \right\rangle = 0 \tag{1.6.3}$$

Each $\hat{\mathcal{F}}\psi_i$ therefore belongs to a space which is totally orthogonal to the virtual space.

## 1.7 Multi Configuration SCF

The Hartree-Fock method is efficient when applied to closed-shell systems at the equilibrium geometry. When these conditions are not met, the limited flexibility of the wavefunction is not sufficient to describe the system. The result is an incorrect representation of those electronic situations which intrinsically need more than one determinant, and a wrong description of the electron-electron repulsion effects. The evaluation of the resulting error, named *correlation energy*, is central in any post-SCF evaluation.

As already stated, the best solution is to consider the Full-CI expansion, but practical computational factors limit this strategy to small molecular systems. An intermediate solution is to pick a limited selection of determinants, chosen under a given criterium. This approach is generally called Multi Configuration or Multi Reference Configuration Interaction (MRCI). A multireference approach provides additional flexibility to the wavefunction, which can be used to describe those cases where the single-determinant Hartree-Fock approximation does not suffice, namely geometrical distortion, bond breaking, or excited states.

### A simple case: the hydrogen molecule

The hydrogen molecule is the most frequently reported case to explain problems related to a single-configuration approximation. The molecular orbitals can be

written as a linear combination of $1s$ atomic orbitals (Linear Combination of Atomic Orbitals, LCAO), normalized with a constant factor $N_1$:

$$\varphi_1 = N_1 \left(1s_a + 1s_b\right) \tag{1.7.1}$$

The resulting spatial orbital is doubly occupied, and the Slater determinant is

$$\Phi_1 = \varphi_1(1)\varphi_1(2)\frac{\alpha\beta - \beta\alpha}{\sqrt{2}} \tag{1.7.2}$$

where the spatial part $\varphi$ and the spin part have been separated for the singlet closed-shell system. Using this wavefunction to describe the equilibrium geometry, the obtained results match qualitatively the experimental values.

Developing the product for $\Phi_1$

$$\begin{aligned}\Phi_1 &= N_1^2 \left(1s_a(1) + 1s_b(1)\right)\left(1s_a(2) + 1s_b(2)\right)\mathcal{S}_0(s_1, s_2) \\ &= N_1^2 \left[\left(1s_a(1)1s_a(2) + 1s_b(1)1s_b(2)\right)\right. \\ &\quad \left. + \left(1s_a(1)1s_b(2) + 1s_b(1)1s_a(2)\right)\right]\mathcal{S}_0(s_1, s_2)\end{aligned} \tag{1.7.3}$$

where $\mathcal{S}_0(s_1, s_2) = \frac{\alpha\beta - \beta\alpha}{\sqrt{2}}$. It is possible to recognize a ionic component (two electrons belong to the same atom) and a covalent one (two equally shared electrons). The single determinant expansion gives the same weight to these representations, regardless of the interatomic distance. Rather obviously, this interpretation does not hold when the molecule is completely dissociated, and the ionic component with both electrons on the same atom is supposed to be zero due to the high distance. As a resulting consequence, the resulting SCF energy behaves incorrectly (Fig. 1.7.1)

If we build a two determinant MRCI

$$\Psi_{\text{MRCI}} = c_1\Phi_1 + c_2\Phi_2 \tag{1.7.4}$$

where $\Phi_2$ is the antibonding configuration

$$\Phi_2 = \varphi_2(1)\varphi_2(2)\frac{\alpha\beta - \beta\alpha}{\sqrt{2}} \tag{1.7.5}$$

made with the difference

$$\varphi_2 = N_2 \left(1s_a - 1s_b\right) \tag{1.7.6}$$

the resulting wavefunction can be optimized with additional degrees of freedom, which provide a way for balancing the ionic and covalent components with respect to the interatomic distance. This function shows the correct dissociation curve.

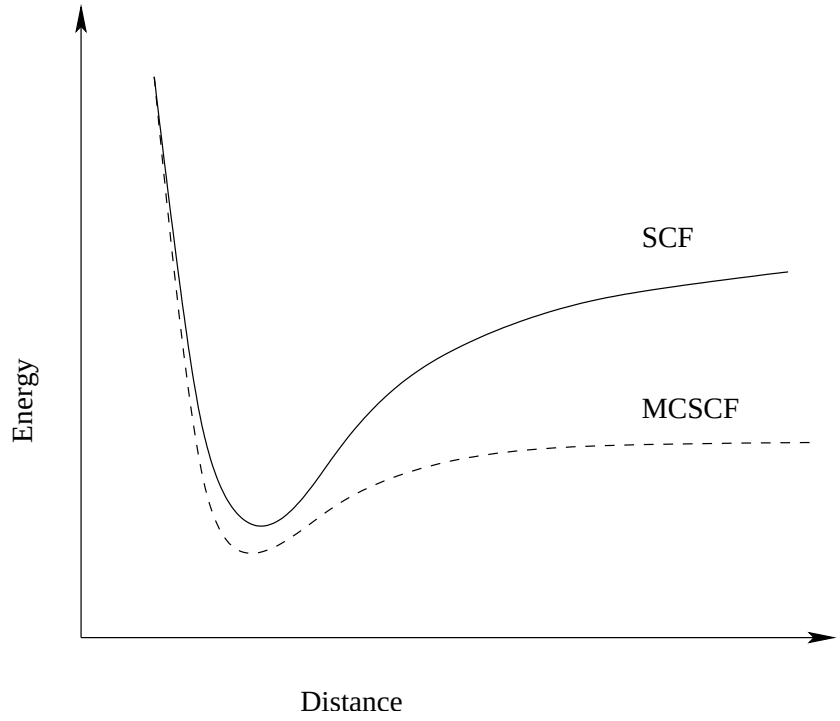


**Figure 1.7.1:** A qualitative behavior for the dissociation of the hydrogen molecule.

### Complete Active Space

The Complete Active Space (CAS) multiconfiguration selection is based on the classification of the orbitals in three distinct groups: *virtual orbitals*, always empty in each determinant of the CI expansion, *active orbitals* occupied in all the possible configurations by a given number of *active electrons*, and *core orbitals*, always occupied by electrons in each determinant of the CI expansion.

This approach resembles a Full-CI limited to the set of active orbitals, and reduces the computational cost without renouncing too much on wavefunction flexibility. The current limit for this technique is 14 active orbitals distributed in 14 electrons. Higher selections are difficult to evaluate without dedicated algorithms, due to the high computational cost involved in the operation.

### Newton-Raphson algorithm

To obtain an optimized multireference wavefunction, an optimization strategy is needed. The Newton-Raphson method is one of these strategies. In this algorithm, the energy $E$ is considered as a function of a set of parameters, arranged as a vector $\mathbf{p}$. The function is then expressed as a Taylor's expansion around a point $\mathbf{p}_0$

$$E(\mathbf{p}) = E(\mathbf{p}_0) + \sum_i \left(\frac{\partial E}{\partial p_i}\right)_{\mathbf{p}_0} p_i + \frac{1}{2}\sum_{i,j} p_i \left(\frac{\partial^2 E}{\partial p_i \partial p_j}\right)_{\mathbf{p}_0} p_j + \ldots \qquad (1.7.7)$$

In matrix notation

$$E(\mathbf{p}) = E(\mathbf{p}_0) + \mathbf{g}^+\mathbf{p} + \frac{1}{2}\mathbf{p}^+\mathbf{H}\mathbf{p} + \dots \tag{1.7.8}$$

where $\mathbf{g}$ is the *energy gradient vector* and $\mathbf{H}$ is the *energy Hessian matrix*. When evaluated in an energy minimum, the gradient vector is zero. An iterative procedure exists, which involves the resolution of the linear equation

$$\mathbf{g} + \mathbf{H}\mathbf{p} = \mathbf{0} \tag{1.7.9}$$

then redefining a new zero point from the obtained solution, evaluating again $\mathbf{g}$ and $\mathbf{H}$ and stepping toward the solution. The procedure is not guaranteed to converge to a minimum: in some cases it can converge to saddle points or even diverge if the guess vector is too far from the solution.

### Super-CI method

An alternative method is the Super-CI approach, which is based on the extended version of the Brillouin theorem. This theorem states that the interaction between an optimized multireference wavefunction and its single contracted excitations is zero.[6] The optimization procedure aims at nullifying iteratively these interactions to produce the optimized solution.

A Super-CI wavefunction is built as

$$|\mathrm{SCI}\rangle = |\Psi_0\rangle + \sum_{p>q} t_{pq} E^-_{pq} |\Psi_0\rangle \tag{1.7.10}$$

where $E^-_{pq} = E_{pq} - E_{qp}$ and $|\Psi_0\rangle$ is a multireference wavefunction. The $E_{pq}$ operator is called *shift operator* and is useful to work with spatial orbitals, regardless of spin. It is expressed as

$$E_{pq} = a^+_{p\alpha} a_{q\alpha} + a^+_{p\beta} a_{q\beta} \tag{1.7.11}$$

The convergence is reached when

$$\left\langle \Psi_0 \left| \hat{\mathcal{H}} \right| E^-_{pq} \Psi_0 \right\rangle = 0 \tag{1.7.12}$$

We must note that single excitations are not orthonormal

$$\begin{aligned} S_{pq,rs} &= \left\langle E^-_{pq}\Psi_0 \middle| E^-_{rs}\Psi_0 \right\rangle && (1.7.13)\\ &= \left\langle \Psi_0 \left| E^-_{qp} E^-_{rs} \right| \Psi_0 \right\rangle && (1.7.14)\\ &\neq \delta_{pq,rs} && (1.7.15) \end{aligned}$$

The Hamiltonian matrix elements are

$$\begin{aligned} \left\langle \Psi_0 \left| \hat{\mathcal{H}} \right| pq \right\rangle &= \left\langle \Psi_0 \left| \hat{\mathcal{H}} E^-_{pq} \right| \Psi_0 \right\rangle = w_{pq} && (1.7.16)\\ \left\langle pq \left| \hat{\mathcal{H}} \right| rs \right\rangle &= \left\langle \Psi_0 \left| E^-_{qp} \hat{\mathcal{H}} E^-_{rs} \right| \Psi_0 \right\rangle = d_{pq,rs} && (1.7.17) \end{aligned}$$

and the secular equation to solve is

$$\begin{pmatrix} E_0 & \mathbf{w}^+ \\ \mathbf{w} & \mathbf{d} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix} = E_{\mathrm{SCI}} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix} \tag{1.7.18}$$

which can be rewritten as

$$\begin{pmatrix} 0 & \mathbf{w}^+ \\ \mathbf{w} & \mathbf{d} - E_0\mathbf{S} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix} = (E_{\mathrm{SCI}} - E_0) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix} \tag{1.7.19}$$

The Super-CI method is guaranteed to converge to a minimum, in contrast to the Newton-Raphson method.

# Resumé chapitre 2
# Localisation

En général, la délocalisation des orbitales moléculaires n'est pas un requisite physique. La description délocalisée des orbitales est importante pour la connexion directe avec l'énergie de ionisation definie par le théorème de Koopmans. Elle donne aussi une bonne explication des systèmes à haute délocalisation, par exemple les systèmes conjugués ou le molécules aromatiques.

Si le système se comporte comme un ensemble de liaisons localisées, une description délocalisée est plus complexe. Par exemple, une approche localisée explique naturellement les similarités entre les liaisons C-H du méthane et de l'éthane. Une description délocalisée est plus complexe.

La délocalisation est souvent une consequence des procédures mathématiques utilisées pour l'optimisation des orbitales. Si on change l'approche mathématique, on peut obtenir des orbitales optimisées et localisées, qui donnent une representation plus facile pour le chimiste. Cette approche peut être aussi utilisée pour réduire le coût computationnel des méthodes *ab initio*.

l'évaluation de la corrélation statique et dynamique est centrale pour obtenir l'énergie d'un système moléculaire avec un haut niveau de précision. La corrélation électronique est un phénomène local, liée à la distribution spatiale des électrons. Une approche délocalisée oblige des évaluations très chères, parce que l'effet physique de la corrélation est distribué sur un nombre très élevé d'integrales.

En utilisant une description locale, les interactions entre orbitales eloignées est presque negligible, et peut être éliminée. Les effets corrélatifs sont concentrés sur un petit ensemble d'intégrals très importantes. La localité des orbitales moléculaires est reconnue importante pour le Linear Scaling, c'est à dire obtenir un algorithme d'évaluation avec coût computationnel linéaire avec la dimension du système moléculaire.

Des systèmes de Linear Scaling sont déjà implémentés pour les méthodes single référence, mais pour l'évaluation des états excitées, de rupture de liaison, de systèmes magnétiques ou de transfert de charge, l'approximation single référence n'est pas suffisante, et une approche multiréférence doit être utilisée. Le Linear Scaling pour les méthodes multireference n'existe pas encore, mais la localisation des orbitales est le premier pas pour l'obtention d'une telle méthode.

La localisation des orbitales peut aider aussi à contrôler la nature de l'espace actif sur des calculs Complete Active Space (CAS), et une reduction de la dimension de cet espace peut être obtenue.

Une méthode de localisation a été développée à Toulouse. Cette méthode est *a priori*, parce que la localisation est gardée à partir d'un set d'orbitales localisées qui sont optimisées à niveau Complete Active Space Self Consistent Field (CASSCF). Au contraire, les méthodes *a posteriori* travaillent avec des orbitales délocalisées qui sont localisées après optimisation.

Après localisation, les orbitales pas intéressées au phénomène chimique étiudé peuvent être gelées à un niveau de théorie plus bas, dans notre cas Self Consistent Field (SCF), pour réduire le poids computationnel, mais cette méthode change assez peu le travail nécessaire pour la transformation des integrales bielectroniques, qui a une complexité O( $N_{\mathrm{AO}}^4 N_{\mathrm{MO}}$ ) avec $N_{\mathrm{AO}}$ nombre d'orbitales atomiques (AO) et $N_{\mathrm{MO}}$ nombre d'orbitales moléculaires. Si on gèle des orbitales moléculaires, l'effet est limité au facteur $N_{\mathrm{MO}}$. Réduire le nombre d'orbitales atomiques et donc le facteur $N_{\mathrm{AO}}^4$ est central pour utiliser des méthodes *ab initio* sur grands systèmes moléculaires.

Une stratégie possible est l'effacement de la partie de molécule pas nécessaire pour la description des orbitales pas gelées. Ces orbitales sont projectées sur un sous-ensemble de la bas atomique original, et une optimisation CASSCF peut être réalisée sur un nombre d'atomes réduit avec un poids computational mineur, mais l'effet des orbitales gelées serà gardé pendant l'optimisation.

La technique, appellée Freeze-and-Cut, a été utilisée sur deux systèmes modèle. Le premièr test est relative au (7Z)-13 ammoniotridec-7-enoate, un aminoacide expressément construit pour évaluer la réponse de la méthode à l'interaction de charge. L'interconversion cis-trans autour d'un double liaison a été étudiée avec un espace actif minimal de deux électrons dans deux orbitales.

Le deuxième test a été appliqué à la molécule de tridequenal, pour évaluer la cohérence de la méthode sur l'estimation de l'énergie d'interconversion cisoid-transoid du groupe aldéhydique.

Une troisième évaluation a été effectué sur l'etat excité $n \rightarrow \pi^*$ de la molecule d'acetone avec 6 molecules d'eau. Dans ce cas, la dimensionalité plus grande de la bas atomique a démontré une sensibilité de la méthode Freeze-and-Cut à des bases diffuses. Une solution peut être l'utilise d'orbitales non-orthogonaux entre le groupe gelé et non-gelé. Cette solution donne des résultats plus corrects, mais la procedure courant doit être adaptée avec l'assumption de non orthogonalité, et donc autres développements theoretiques sont nécessaires.

# Chapter 2

# Localization

*This chapter, in particular for the computational evaluation, is based on the article at Ref. 7*

The need to evaluate the properties of large molecular systems using a high-level *ab initio* approach is an interesting and challenging task for the quantum chemist. Many useful results have been achieved in this direction, thanks to refinements of theoretical approaches, together with computational resources increasingly more efficient and cheap. However, the problem of the growth of the computational effort as a function of the system size still represents, in most cases, a serious bottleneck. This can lead to a huge computational cost even for relatively small systems.

Many techniques have been implemented to deal with large systems, like, for instance, metallic complexes with organic ligands, or biologically interesting molecules. For example, in the evaluation of potential energy surfaces of the ground state a Molecular Mechanics approach makes feasible calculations on large molecular systems. The price to pay, in this case, is to consider the system as purely classical.

When a quantum mechanical evaluation is needed (description of excited states, bond breaking, etc.), the mixing of molecular mechanics with quantum mechanics is a possible strategy, like Quantum Mechanics/Molecular Mechanics QM/MM[8] and the "Our own N-layered Integrated molecular Orbital and molecular Mechanics" ONIOM method.[9] For the calculation of excited states, Time Dependent Density Functional Theory (TDDFT) methods are also available, which behave well in some cases but with substantial errors in others, as pointed out for example in Ref. 10.

Performing purely *ab initio* evaluations is also feasible. A possible approach is

represented by the elimination of large molecular groups, replacing them by computationally cheaper mimicking entities, using Effective Group Potential (EGP) pseudo-potentials,[11,12] for example, or simply by hydrogen atoms. Other available methods, like the Symmetry Adapted Cluster-Configuration Interaction (SAC-CI)[13–16] have been applied successfully to large biological systems.[17]

Another possible strategy in the treatment of large molecular systems is given by Localized Molecular Orbitals (LMO). The use of localized orbitals in quantum chemistry has a long history.[18] Only in recent years, however, Localized Molecular Orbitals have become popular as a key step toward Linear Scaling *ab initio* approaches, where the computational cost of the evaluation grows linear with respect to the molecular system size. Using a localized philosophy, a reduction of the high computational cost imposed by delocalized methods can be obtained.

## 2.1 Advantages of a localized approach

In general, delocalization of molecular orbitals is not a physical requirement. A delocalized picture of the orbitals gained mainstream attention in connection with the ionization energy as expressed by the Koopmans theorem. Delocalization can also explain some phenomena in those systems that possess intrinsic physical delocalization, like for example polyene chains or aromatic molecules.

However, when a molecule behaves like a set of localized bonds, a delocalized description introduces a high degree of complexity. As an example, a localized point of view clearly explains the similar behavior of C-H bonds in methane, ethane and cyclohexane. A delocalized description of this physical phenomenon is more difficult.

Delocalization is also a frequent consequence of the mathematical procedures implemented for orbital optimization. Improving the mathematical asset in order to obtain spatially localized orbitals allows the quantum chemist to deal with a more practical representation of the charge distribution, and also provides possible new strategies in reducing the computational cost for *ab initio* evaluations.

A reliable evaluation of the static and dynamic correlation is the main limiting factor for a quantum chemical approach to medium and large molecular system. These contributions to the energy are critical, and their evaluation computationally demanding, but their nature is a local phenomenon, related to the spatial distribution of the electrons. A delocalized approach imposes very expensive evaluations: the physical effect of correlation is spread over a large number of integrals. Each of these integrals accounts for an almost unpredictable quantity to the energy, and their physical meaning is difficult to understand.

Making use of a localized description, interactions between electrons popu-

lating distant, non-overlapping orbitals can be considered negligible, and thus candidate for being eliminated from the evaluation. As a direct consequence, correlative effect are now concentrated on a reduced set of highly important integrals. Neglecting these long-range interactions due to locality of the molecular orbitals is an important strategy to reduce the computational cost.

Linear Scaling techniques are already available for SCF calculations,[19] Møller-Plesset perturbative treatments (MP) at different orders[20–26] Single and Double Configuration Interaction (SDCI),[27] Single and Double Coupled Cluster (CCSD)[27,28] and CCSD(T).[29,30] All the cited methods apply to a single reference wavefunction, and are therefore unbalanced when two or more determinants of nearly equal weight are critical in the description of the system. Although multireference Linear Scaling techniques are still to be devised, a localized orbitals approach is recognized as an important requisite in achieving this objective, allowing a "*divide et impera*" strategy on the molecule.

It is important to note that many chemically interesting problems require the application of a multireference scheme in order to produce reliable results. Examples of these problems are the treatment of bond breakings, electronically excited states, magnetic systems, and charge transfer. Treatment of these systems is important to gather a deep insight in biological catalysis, photochemical reactions, high atmosphere degradation of polluting compounds, technological improvements for data storage, data transfer and computation.

Localization of molecular orbitals can also benefit multireference evaluations in terms of quality and size of the active space. Phenomena involving electronic excitation or bond breaking usually happen in a well localized region of the molecule.

Working with delocalized orbitals, the active space needed to describe these phenomena is normally defined by those MOs bringing the highest correlative effects, regardless of the spatial or physical nature. Moreover, the user has poor or no control of the active orbital selection. Using local orbitals, a well defined and physically clear active space can be chosen and maintained during the optimization.

Finally, a reduction of the reference space can also be obtained: a localized description clearly depicts which determinants are not important for the description of the wavefunction, because they express highly ionic electronic distributions where a large number of electrons are concentrated in a certain region of space.

## 2.2 Existent methods of localization

In the past, localized orbitals describing the molecule were obtained performing an appropriate transformation of a delocalized manifold, i.e. applying a unitary

transformation to the spatial orbitals row vector $\psi$

$$\psi_{\mathrm{loc}} = \psi_{\mathrm{deloc}}\mathbf{U} \tag{2.2.1}$$

Finding localized, equivalent orbitals is normally achieved choosing a $\mathbf{U}$ matrix which maximizes a "localization function". For example, the contribution from Edmiston and Ruedenberg[31] aims at minimizing the Coulomb repulsion between electron pairs occupying two different orbitals. If $\phi_i$ and $\phi_j$ are two different orbitals, then satisfying the condition

$$\sum_{i<j}^{\mathrm{occ}} \langle \phi_i\phi_j \,|g|\, \phi_i\phi_j \rangle = \text{minimum} \tag{2.2.2}$$

imposes locality of the charge distribution. Another approach from Boys[32] is to maximize the distance between the centroids of the transformed orbitals. Other methods, like for example those by Pipek,[33] Angeli et al.[34] work along the same principle, and can be classified as *intrinsic methods* of localizations. *Extrinsic methods* work equally well, and they are realized performing a projection of localized MO's onto a delocalized optimized manifold.[18]

All these methods resort on *a posteriori* localization techniques, therefore a prior evaluation of delocalized canonical MOs is needed. Obtaining a localized set of orbitals is also possible using *a priori* methods, where a controlled optimization procedure is applied on a localized but unoptimized guess of orbitals. The controlled procedure preserves the locality, rather than imposing it. Methods that address SCF equations already exist.[35–37]

The main advantage of an *a priori* method is the possibility to choose and preserve the nature of the active space during a CAS evaluations. This property is of particular importance in order to build a CAS space using the orbitals involved in the chemical or physical process under study.

In all these methods, the local orbitals are expressed as an orthogonal set. In general this approach is computationally more efficient compared to Generalized Valence Bond methods[38,39] where a set of non-orthogonal orbitals with atomic character is variationally optimized.

## 2.3 Toulouse method of localization

An *a priori* localization method has been recently proposed by the working group in Toulouse, with the goal to address multireference problems. The method has been applied to the description of the excited state of polyenes polyenals, polyendials and polyenones,[40–43] $C_{60}$ Fullerene + Sodium interaction,[44] nickel and copper chains[45] in a mixed localized/delocalized approach, Cu magnetic systems,[46] Hydrogen chains,[47] and Mo-Cl complexes.[48]

Briefly, this method preserves the local character of LMO, converging to a CASSCF wavefunction. A localized guess is obtained through the projection of Localized Molecular Orbitals (LMO) onto the delocalized SCF manifold. Then, the locality is preserved avoiding a complete diagonalization of the density matrix.

Three versions of the optimization algorithm are currently available: a variational uncontracted,[49] developed in Toulouse, a perturbative contracted[50] and the recently deployed variational contracted, both developed in Ferrara.

The subsequent work is based on the variational uncontracted approach. This approach does not converge to a rigorous CASSCF solution, because of the uncontracted nature of the implemented CI expansion. However, as already shown in Ref. 49, this fact does not seem to have severe practical effects.

### Creation of the guess

The method starts with the creation of a set of Orthogonal Atomic Orbitals (OAO), obtained through a Schmidt orthogonalization and a $S^{-\frac{1}{2}}$ normalization of the atomic basis set. ANO type[51] is preferred, since the orbitals are described by a minimal number of coefficients.

The orthonormalization is performed against a hierarchical classification of the orbitals: the core orbitals are orthogonalized among themselves, then the valence orbitals against the cores and among themselves, finally the remaining orbitals. The obtained OAOs are then used to build Localized Molecular Orbitals (LMOs) as linear combinations of OAOs, using the one-electron density matrix of a one-electron evaluation (Huckel or SCF):

$$R_{ij} = \left\langle \Phi_0 \left| a_i^+ a_j \right| \Phi_0 \right\rangle \tag{2.3.1}$$

It is possible to define a projector for an atom K (built using the projector onto the $\tilde{\chi}_l$ OAO functions) or for a molecular fragment F (from a sum of projectors onto the involved atoms)

$$P_K = \sum_{l \in K} |\tilde{\chi}_l\rangle \langle \tilde{\chi}_l| \tag{2.3.2}$$

$$P_F = \sum_{K \in F} P_K \tag{2.3.3}$$

The density matrix obtained by the projected wavefunction $P_F \Phi_0$ can be diagonalized. This step leads to a linear combination of OAO, representing MO's localized on the fragment. Some of these MO's exhibit occupation numbers which are close to two or zero, and they represent a localized bond between the atoms involved in the fragment. Other orbitals can exhibit occupation numbers close to one, and they are rejected since they are hybrid orbitals that are representative of an "half-bond" between non considered atoms.

As an example, the creation of LMO's in the carbonyl group of acetone is here presented.

```
Extracted matrix
----------------

O1 1s   1.6881  0.0000  0.0000  0.3780  0.2764  0.0000  0.0000  0.5185
O1 2px  0.0000  1.2406  0.0000  0.0000  0.0000  0.9340  0.0000  0.0000
O1 2py  0.0000  0.0000  1.8980  0.0000  0.0000  0.0000  0.2703  0.0000
O1 2pz  0.3780  0.0000  0.0000  1.4621 -0.4912  0.0000  0.0000 -0.6081
C1 1s   0.2764  0.0000  0.0000 -0.4912  0.9477  0.0000  0.0000 -0.0968
C1 2px  0.0000  0.9340  0.0000  0.0000  0.0000  0.8168  0.0000  0.0000
C1 2py  0.0000  0.0000  0.2703  0.0000  0.0000  0.0000  1.0402  0.0000
C1 2pz  0.5185  0.0000  0.0000 -0.6081 -0.0968  0.0000  0.0000  0.9164

Orbitals
--------

Occup.
Number  1.9975  1.9865  1.9777  1.9761  1.0201  0.9622  0.0709  0.0190

O1 1s  -0.5478  0.0000  0.7384  0.0000 -0.0855  0.0000  0.0000 -0.3838
O1 2px  0.0000  0.7814  0.0000  0.0000  0.0000  0.0000 -0.6240  0.0000
O1 2py  0.0000  0.0000  0.0000  0.9607  0.0000 -0.2775  0.0000  0.0000
O1 2pz  0.5405  0.0000  0.6639  0.0000 -0.0413  0.0000  0.0000  0.5151
C1 1s  -0.3477  0.0000 -0.1176  0.0000 -0.8134  0.0000  0.0000  0.4512
C1 2px  0.0000  0.6240  0.0000  0.0000  0.0000  0.0000  0.7814  0.0000
C1 2py  0.0000  0.0000  0.0000  0.2775  0.0000  0.9607  0.0000  0.0000
C1 2pz -0.5356  0.0000 -0.0089  0.0000  0.5739  0.0000  0.0000  0.6195

-------------------------------------------------------------------------
    Occupied              Eliminated             Virtuals
 1.9975... 1.9761 /  /  1.0201...  0.9622 /  /  0.0709 ... 0.0190
```

As can be seen, the extracted density matrix leads to a localized set of orbitals. The first four, with occupation numbers close to two, are the C-O $\sigma$ bond, the CO $\pi$ bond, and the lone pairs $n_z$ and $n_y$. Two orbitals express the half $\sigma$ bonds directed toward the remaining carbon atoms. These bonds will be kept into account in other selections, so they are discarded. Finally, two orbitals with occupation number close to zero express the C-O $\pi^*$ and $\sigma^*$. The resulting set of LMOs is not orthogonal, so a new hierarchical orthonormalization is performed.

The obtained orbitals are then projected onto a delocalized SCF manifold. This step, like an extrinsic method of localization, provides a localized SCF-quality set of orbitals.

### Optimization of the guess

In the classical (non Freeze-and-Cut) localization technique, orbitals obtained from the last step are then optimized at CASSCF level. Working on a single reference SCF wavefunction $\Phi_0$, the optimization should follow these steps:

1. define a single excitation configuration interaction $\Psi_{\text{CIS}} = \Phi_0 + \sum_{i,r} C_{ir} a_r^+ a_i \Phi_0$ and obtain the $C_{ir}$ coefficients through diagonalization
2. create the one electron density matrix $R_{ir} = \langle \Psi_{\text{CIS}} | a_r^+ a_i | \Psi_{\text{CIS}} \rangle$ and block diagonalize it
3. use the new set of orbitals to create a new $\Phi_0$, and iterate until $\Psi_{\text{CIS}} = \Phi_0$, satisfying the Brillouin theorem.[5]

Particular care must be applied in the density matrix diagonalization step. A complete diagonalization leads to delocalized natural orbitals. To prevent this behavior, a controlled block diagonalization is performed

$$R_{\text{DP}} = W^+ R W \tag{2.3.4}$$

where $R$ is the density matrix, $R_{\text{DP}}$ is the block diagonalized density matrix, and $W$ is a unitary matrix, given by

$$W = U_{\text{D}} U_{\text{P}}^+ \tag{2.3.5}$$

where $U_{\text{D}}$ is a matrix that fully diagonalizes $R$ and $U_{\text{P}}$ is a matrix that restores the orbital locality, mixing those orbitals inside the same class (occupied or virtual, in the case of a single reference calculation). The $U_{\text{P}}$ matrix is obtained through projection of $U_{\text{D}}$ onto the subspaces and a subsequent orthonormalization.

The same working principle can be applied to multireference CAS wavefunctions: using a CAS space, the CI energy and wavefunction are invariant with respect to rotations inside each block of core, active and virtual orbitals. This invariance allows us to produce localized orbitals equivalent to the delocalized ones, and locality can be maintained preventing as much as possible rotations inside each of the three blocks.

The requirement for convergence is to satisfy the Extended Brillouin Theorem

$$\left\langle \Psi_{\text{CASSCF}} \left| \hat{\mathcal{H}} \left[ a_k^+ a_l - a_l^+ a_k \right] \right| \Psi_{\text{CASSCF}} \right\rangle = 0 \tag{2.3.6}$$

which is always fulfilled if $k$ and $l$ refer both to core, active or virtual orbitals. The previous equation can threfore be decomposed in three equations between

different classes of indexes:

$$\left\langle \Psi_{\text{CASSCF}} \left| \hat{\mathcal{H}} \left[ a_a^+ a_i - a_i^+ a_a \right] \right| \Psi_{\text{CASSCF}} \right\rangle = 0 \tag{2.3.7}$$

$$\left\langle \Psi_{\text{CASSCF}} \left| \hat{\mathcal{H}} \left[ a_r^+ a_i - a_i^+ a_r \right] \right| \Psi_{\text{CASSCF}} \right\rangle = 0 \tag{2.3.8}$$

$$\left\langle \Psi_{\text{CASSCF}} \left| \hat{\mathcal{H}} \left[ a_r^+ a_a - a_a^+ a_r \right] \right| \Psi_{\text{CASSCF}} \right\rangle = 0 \tag{2.3.9}$$

where $i$, $a$ and $r$ refer to core, active and virtual orbitals, respectively. In order to minimize the energy, the non-interaction of singly-excited contracted CAS wavefunctions must be reached iteratively using the Super-CI approach. A prototype for this approach has been developed in Ferrara laboratory, and has been used for the combined Localization-NEVPT2 evaluation, presented later in this thesis.

An alternative solution, followed at Toulouse laboratories, is to consider a decontracted scheme, exploiting the full dimensionality defined by the space of determinants $S = \{\Phi_i\} \cup \{a_r^+ a_i \Phi_i\}$ spanned by the CAS space and their single excitations. The procedure is not a Super-CI optimization and does not converge to the real CASSCF solution. However, as pointed out in Ref. 49, the resulting orbitals are very similar to the CASSCF ones.

The third possibility is to work with a perturbartive contracted approach. The idea is to consider the $\Psi_{\text{CASCI}}$ at a given iteration as a zero-order description of the $\Psi_{\text{CAS+S}}$. We diagonalize a zero-order Hamiltonian inside a perturber space Q, spanned by the singly excited contracted functions $a_a^+ a_i \left|\Psi_{\text{CASCI}}\right\rangle$, $a_r^+ a_i \left|\Psi_{\text{CASCI}}\right\rangle$ and $a_r^+ a_a \left|\Psi_{\text{CASCI}}\right\rangle$. From the diagonalization, we obtain perturber wavefunctions as linear combinations

$$\left|\Psi_\mu\right\rangle = \sum_{ai} c_{ai\mu} a_a^+ a_i \left|\Psi_{\text{CASCI}}\right\rangle + \sum_{ri} c_{ri\mu} a_r^+ a_i \left|\Psi_{\text{CASCI}}\right\rangle + \sum_{ra} c_{ra\mu} a_r^+ a_a \left|\Psi_{\text{CASCI}}\right\rangle \tag{2.3.10}$$

and the first-order correction to the wavefunction is

$$\left|\Psi^{(1)}\right\rangle = \sum_\mu \frac{\left\langle \Psi_\mu \left| \hat{\mathcal{H}} \right| \Psi_{\text{CASCI}} \right\rangle}{E_{\text{CASCI}} - E_\mu^{(0)}} \left|\Psi_\mu\right\rangle \tag{2.3.11}$$

From this correction, it is possible to obtain the correction to the one-particle density matrix

$$\begin{aligned} \rho_{lk}^{(1)} = \sum_\mu \Bigg\{ & \frac{\left\langle \Psi_{\text{CASCI}} \left| a_k^+ a_l \right| \Psi_\mu \right\rangle \left\langle \Psi_\mu \left| \hat{\mathcal{H}} \right| \Psi_{\text{CASCI}} \right\rangle}{E_{\text{CASCI}} - E_\mu^{(0)}} + \\ & + \frac{\left\langle \Psi_{\text{CASCI}} \left| \hat{\mathcal{H}} \right| \Psi_\mu \right\rangle \left\langle \Psi_\mu \left| a_k^+ a_l \right| \Psi_{\text{CASCI}} \right\rangle}{E_{\text{CASCI}} - E_\mu^{(0)}} \Bigg\} \end{aligned} \tag{2.3.12}$$

When the convergence is reached, the correction vanishes. For the choice of the zero-order Hamiltonian, is convenient to make use of a particular Hamiltonian,

the Dyall's Hamiltonian[52]

$$\hat{\mathcal{H}}^D = \hat{\mathcal{H}}_i^D + \hat{\mathcal{H}}_v^D + C \tag{2.3.13}$$

$$\hat{\mathcal{H}}_i^D = \sum_i \epsilon_i a_i^+ a_i + \sum_r \epsilon_r a_r^+ a_r \tag{2.3.14}$$

$$\hat{\mathcal{H}}_v^D = \sum_{ab} h_{ab}^{\text{eff}} a_a^+ a_b + \frac{1}{2} \sum_{abcd} \langle ab|cd\rangle \, a_a^+ a_b^+ a_d a_c \tag{2.3.15}$$

where labels $i, j, \ldots$, $a, b, \ldots$, $r, s, \ldots$ denote core, active and virtual orbitals respectively, and $h_{ab}^{\text{eff}} = \langle a \,|h + \sum_i (J_i - K_i)|\, b\rangle$. This Hamiltonian, which will be presented in more details in the n-Electron Valence state Perturbation Theory approach, leads to a simplification of the computational asset. The zero-order Hamiltonian is defined as the projection of the Dyall inside the interested spaces

$$\hat{\mathcal{H}}_0 = \hat{\mathcal{P}}_{\text{CAS}} \hat{\mathcal{H}}^D \hat{\mathcal{P}}_{\text{CAS}} + \hat{\mathcal{P}}_{\text{Q}} \hat{\mathcal{H}}^D \hat{\mathcal{P}}_{\text{Q}} \tag{2.3.16}$$

The perturbative approach is very straightforward and efficient, and converges to the real CASSCF wavefunction. However, divergent behaviors have been reported when the zero-order wavefunction needs a strong correction.

## 2.4 Freeze-and-Cut method

A fundamental assumption in reducing the computational cost of *ab initio* evaluations is that correlative effects are usually local, given the nature of the studied phenomena. For this reason, the molecular framework not involved in chemical or physical phenomena is an ideal candidate to be frozen at a lower level of theory, like for example SCF, after the creation of the localized guess. This strategy has already proved its efficiency in reducing the computational cost,[44] but has no influence on the basis set size, only the number of molecular orbitals. This is a particularly severe limitation for MRCI methods, which require a two-electron AO/MO integral transformation for all correlated orbitals. This step grows as $N_{\text{AO}}^4 N_{\text{MO}}$, where $N_{\text{AO}}$ is the number of Atomic Orbitals (AO) and $N_{\text{MO}}$ is the number of Molecular Orbitals (MO). If no freezing is performed on the molecular orbitals, the transformation grows as $N^5$. Freezing a part of the orbital set a rectangular transformation is performed and the computational weight is reduced, but the quartic dependence on $N_{\text{AO}}$ still limits calculations on large AO spaces. This problem must be managed in order to perform calculations with a reduced dependency of the basis set.

A possible strategy is to cut out the molecular framework that is not essential for the description of the phenomenon under study. In other words, *all the atoms whose AOs are not fundamental in the description of the non-frozen MOs are eliminated*. The subsequent CASSCF optimization is performed in the preserved

molecular framework only, *under the effect of the eliminated nuclei and electrons as provided by the freeze*. This effect takes into account sterical interaction and charge effects provided by the frozen framework internally (within itself) and externally (onto the non-frozen framework), given that the AO/MO transformation of the integrals retains information about the frozen framework.

Moreover, once the CASSCF optimization has been performed, the method leaves freedom for application of higher levels of theory, still preserving the effects of the frozen molecular framework at a lower level of theory.

### Computational procedure

The procedure starts with the creation of a set of SCF delocalized orbitals, which is subsequently localized to obtain a localized SCF set. The resulting localized orbitals, each one mainly described on a reduced set of atomic functions, are selected for the freezing step. The remaining unfrozen orbitals are then projected onto a reduced AO space, setting to zero the coefficients on those AOs that are not essential for the expression of the orbitals.

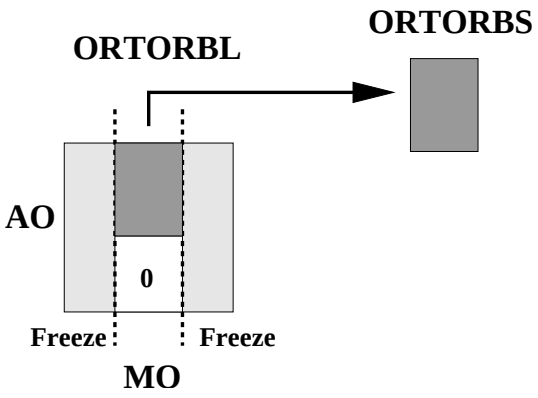


**Figure 2.4.1:** A visual representation of large (ORTORBL) and small (ORTORBS) molecular orbitals matrices. Localized orbitals are first classified as frozen or non-frozen. The non-frozen orbitals are supposed to have negligible coefficients on distant atoms. These coefficients are strictly set to zero by the cut procedure, and a localized optimization is performed in the obtained AO/MO subspace.

This procedure leads to two different orbital matrixes. The first one has the dimension of the original AO basis set, but with some of the coefficients of non-frozen orbitals set to zero. The second one is the submatrix of coefficients that belong to unfrozen orbitals (see Fig. 2.4.1). The latter has a reduced dimension, and corresponds to the effective space on which the optimization procedure will be performed.

The orbitals from the two matrixes are no longer orthogonal, and a hierarchical orthonormalization is performed on the large matrix (non-frozen orbitals first,

then frozen orbitals against the non-frozen ones) and on the small matrix, to obtain two sets of orbitals, named ORTORBL and ORTORBS respectively. An AO/MO integral transformation is performed at each iteration from the improved orbitals. One-electron MO integrals are provided by the full dimensional orbital structure, while two-electron MO integrals are evaluated only on the reduced set described by ORTORBS.

This procedure results in a spatial confinement of the non-frozen orbitals, improving the ORTORBS orbitals only inside this reduced AO space. These MOs have strictly zero coefficients on any AO belonging to the eliminated region. For this reason, the method is affected by the truncation of orthogonalization tails for large basis sets, or in the case of highly delocalized systems. In general this truncation affects the absolute energies, with very little effect on the energy differences.

The computational chain is built to interface `MOLCAS` 5.4,[53] with a set of small home-developed softwares and bash scripting in GNU/Linux[54] environment to provide the complete deployment of this technique.

A scheme for the method is depicted in Fig. 2.4.2. The scheme focuses on the results (depicted as boxes) obtained through processing steps (depicted as arrow lines). Thick arrows represent processes that involves evaluation or use of integrals on the complete system and so represent still a bottleneck in this technique. Improvements are planned to skip this evaluation, by making use of a direct SCF procedure. The formalism can be subdivided into two logical steps:

- creation of the localized guess (upper part of the scheme)
- iterative optimization of the localized guess (lower part, enclosed into dashed frame)

The upper part of the scheme can also be subdivided in two independent calculations: the large calculation, performed on the complete molecular geometry (left side of Fig. 2.4.2) and the small calculation, performed only on the geometry resulting after the cut operation (truncated system, right side of Fig. 2.4.2). The localized guess is obtained through the following strategy:

- evaluation of the AO integrals on the large system, by using standard software components from the `MOLCAS` 5.4 package
- calculation of delocalized SCF orbitals
- creation of a localized guess that describes the bonds or the groups we are interested in
- projection of the localized guess onto the SCF delocalized orbitals

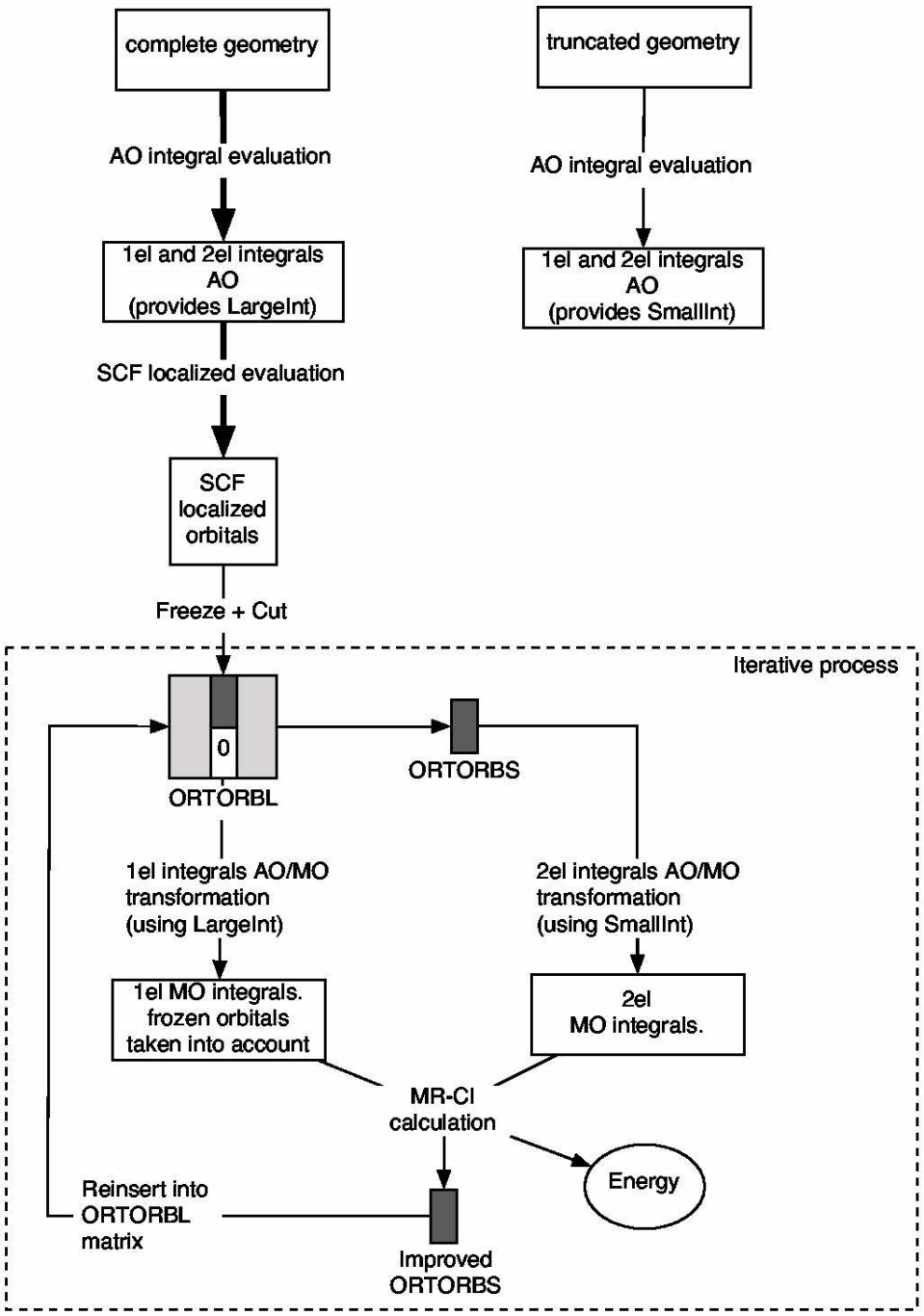


**Figure 2.4.2:** Logical steps (arrows) and obtained entities (boxes) for the presented procedure. Thick lines denote steps that still depend on two electron AO integrals on the complete geometry of the system

These steps describe in more detail the process labeled as "SCF localized evaluation" in Fig. 2.4.2. The result is a set of orthogonal localized SCF orbitals.

Thanks to the localization procedure, each localized SCF orbital is mainly

described on a reduced set of atomic basis functions. These orbitals can now be tagged on two degrees of freedom: the freezing selection of certain localized SCF orbitals, and the cut of a set of atomic basis functions for the non-frozen orbitals. The latter is realized by setting to zero the coefficients that express the non-frozen orbitals on the cut atomic basis set. This operation effectively projects the non-frozen MO in a reduced atomic basis space, and the approximation is good as long as the target coefficients are already close to zero due to the localization.

The large MO matrix (refer to the detailed pictorial representation in Fig. 2.4.1) holds molecular orbitals that are no longer orthonormal, although it must be pointed out that the overlap between them is small. A hierarchical orthonormalization is performed to obtain an orthonormal set: first the non-frozen orbitals among themselves, then the frozen occupied orbitals with respect the non-frozen, and finally the frozen virtual orbitals. Thanks to the hierarchical procedure, the coefficients that have been set to zero remain unchanged. This is important in order to keep the non-frozen orbitals inside a smaller AO space.

Finally, a smaller matrix is obtained by extracting the submatrix of the non-frozen orbitals against the smaller, non-cut atomic set. The final result of this process are two matrixes, named ORTORBL and ORTORBS, which hold the molecular orbitals as described, respectively for the large and the small system (complete system and truncated one).

After an AO/MO integral transformation, the optimization chain is now fed with two-electron MO integrals (which derive only from orbitals that are not frozen) and with the modified one-electron integrals that keep into account the two-electron contribution from the frozen set.

At each iteration, the procedure works as follows:

- calculate an improved density matrix using a Super-CI optimization that preserves orbitals locality. This process is done only on the restricted subset of non-frozen orbitals, and only on the available framework of non-cut atoms, leading to an improved set of orbitals (improved ORTORBS)
- reinsert the improved orbitals inside the frozen framework, thus creating an improved ORTORBL large matrix
- recreate a new set of MO two-electron integrals, transforming the AO integrals from the truncated system using the improved ORTORBS orbitals
- recreate a new set of MO one-electron integrals with the improved ORTORBL set and the complete system
- iterate until convergence is reached (no change in the energy within a given level of approximation)

## 2.5 Evaluations

### 2.5.1 (7Z)-13 ammoniotridec-7-enoate

The first test case presented is relative to (7Z)-13 ammoniotridec-7-enoate, an aminoacid zwitterion specifically designed to test the response of the method to charge interactions (see Fig. 2.5.1). The rigid rotational behavior around the central double bond has been evaluated, describing the absolute CAS+Single excitations (CAS+S) energy curves without geometry relaxation with respect to the torsional dihedral angle.

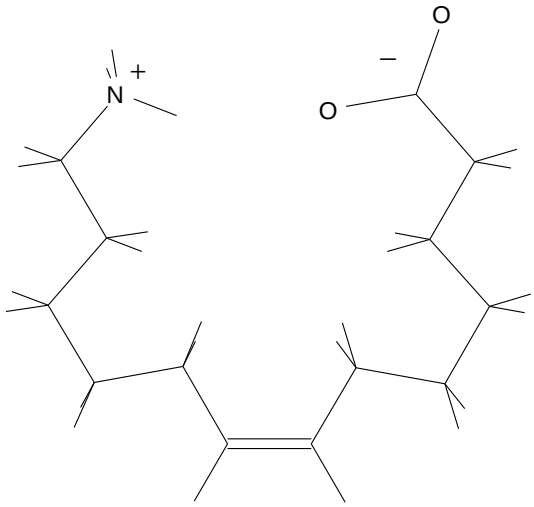


**Figure 2.5.1:** The (7Z)-13 ammoniotridec-7-enoate molecule. The cis-trans rigid interconversion around the central double bond has been performed to test the presented method.

All the calculations have been performed by using ANO basis sets. A minimal basis set ANO-1[51] with $2s1p$ contraction for C,N,O and $1s$ contraction for hydrogen atoms was used. The interatomic distances (in Ångstrom) are r(C-C)=1.450, r(C=C)=1.335, r(C-H)=1.089, r(C-N)=1.440, r(N-H)=1.008, r(C-O)=1.400. The angles are 109.5 degrees except for the C=C group and the $COO^-$ group, where angles of 120 degrees have been used. The dihedral angles were adjusted to obtain the cis and trans molecular skeleton lying completely on the plane.

The CAS space selection consists of 2 electrons in 2 orbitals (C-C $\pi$ and $\pi^*$). This selection has been chosen to keep into account the main correlative effects due to the rotation around the central carbon-carbon double bond.

It must be stressed that, due to geometrical proximity, the interaction between the charged groups affects the rotational behavior, and its contribution must be kept into account. For this reason, the two terminal groups cannot be simply removed and replaced by hydrogen atoms. As a comparison example, Fig. 2.5.2

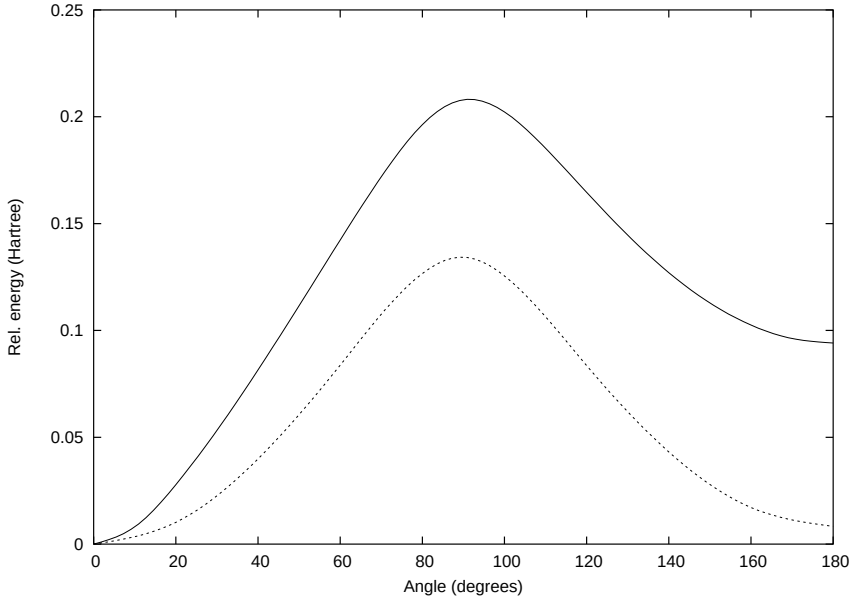


**Figure 2.5.2:** A comparison of the CAS+S energy curves between the zwitterionic (7Z)-13 ammoniotridec-7-enoate (solid line) and the (6Z)-dodec-6-ene, a molecule obtained by replacing the charged $NH_3$ and $COO^-$ groups with hydrogen atoms, dashed line. A common zero in the energy scale of the plot has been obtained taking the trans form as zero of the energy.

shows the behavior obtained by removing these groups and replacing them with hydrogens.

Using the Freeze-and-Cut technique, the evaluation performed on a small system reproduces the expected behavior: the freezing preserves the electronic contribution of the groups, and the cut allows to work with a reduced molecular system.

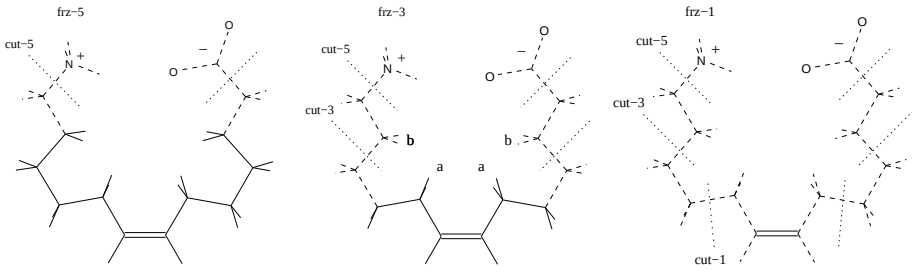


**Figure 2.5.3:** (7Z)-13 ammoniotridec-7-enoate. Localized orbitals expressed on atoms of the dashed line framework are frozen. Orbitals expressed on atoms of the solid line framework are not frozen. Dotted lines depict the cut seam. The cut is always performed on both chains, from "cut-5" (only the charged groups are removed from the molecule, both left and right) to "cut-1" (maximal cut selection, only the central double bond and left and right -$CH_2$- spacers are preserved). The "a" and "b" symbols in frz-3 picture are labels for interesting hydrogen atoms. See text for details.

Different freeze and cut strategies have been chosen in order to evaluate the behavior of our technique with respect to these selections. Three freezing levels, labeled "frz-1," "frz-3" and "frz-5" have been defined (Fig. 2.5.3), and also a "nofrz" level where no freezing has been performed. The $1s$ core atomic orbitals for heavy atoms have been frozen at SCF level regardless of the atom positions.

The cut strategy follows the freezing strategy. Four cut thresholds have been chosen, with labels "cut-1," "cut-3," "cut-5" and "nocut," following the freezing choice, but preserving a -$CH_2$- spacer between the last not frozen orbital and the first cutout atom.

The analysis performed at 0 (cis) and 180 (trans) degrees and their difference are presented in Tab. 2.5.1

| | nofrz | frz-5 | frz-3 | frz-1 |
|---|---|---|---|---|
| | | Cis | | |
| nocut | -709.483006 | -709.482889 | -709.482281 | -709.465687 |
| cut-5 | | -709.373507 | -709.479395 | -709.465638 |
| cut-3 | | | -709.218388 | -709.460918 |
| cut-1 | | | | -709.357068 |
| | | Trans | | |
| nocut | -709.388878 | -709.388760 | -709.388156 | -709.371500 |
| cut-5 | | -709.279814 | -709.385430 | -709.371437 |
| cut-3 | | | -709.123318 | -709.366578 |
| cut-1 | | | | -709.258750 |
| | | Diff | | |
| nocut | 59.0267 | 59.0272 | 59.0251 | 59.0640 |
| cut-5 | | 58.7538 | 58.9248 | 59.0725 |
| cut-3 | | | 59.6173 | 59.1598 |
| cut-1 | | | | 61.6542 |

**Table 2.5.1:** CAS+S absolute energies (Hartree) and energy difference (kcal/mol) between (7Z)-13 ammoniotridec-7-enoate cis and trans structure, with respect to different freeze and cut strategies.

It can be seen that the cut technique produces energy differences between cis and trans that are comparable to the reference value obtained with no cut and freeze.

A different behavior can be evaluated for the difference between the 0 degrees form and the 90 degrees form (Tab. 2.5.2)

| | nofrz | frz-5 | frz-3 | frz-1 |
|---|---|---|---|---|
| nocut | 130.4145 | 130.4953 | 131.2643 | 167.4817 |
| cut-5 | | 130.1703 | 131.1572 | 167.5065 |
| cut-3 | | | 130.8074 | 168.4879 |
| cut-1 | | | | 171.2689 |

**Table 2.5.2:** CAS+S energy difference (kcal/mol) between the (7Z)-13 ammoniotridec-7-enoate cis and the 90 degrees twisted structure with respect to different freeze and cut strategies.

The frz-1 selection presents a large deviation from the expected value. This arises from the fact that the SCF at 90 degrees evaluates very poorly the orbitals involved (directly or indirectly) in the bond breaking.

This leads to an initial bad description of the orbitals, which are frozen at this poor quality level. As a consequence, the optimization guess is affected by this initial description. By relaxing the freeze selection, these orbitals are allowed to improve in the subsequent iterative process, thus drastically reducing the incorrect behavior. In the case of frz-1 strategy, however, the effects are too pronounced to be smoothed out.

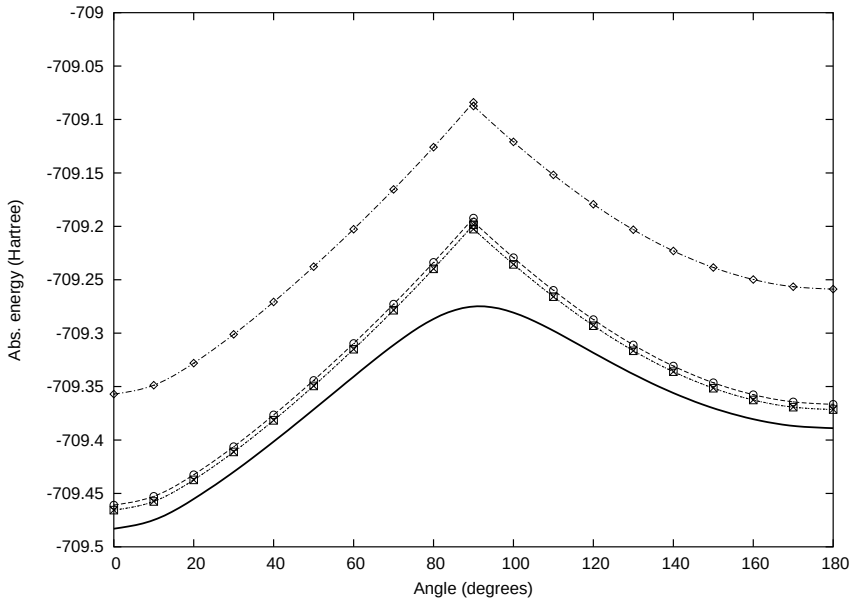


**Figure 2.5.4:** CAS+S energy curves (Hartree) for the cis-trans rigid interconversion of the (7Z)-13 ammoniotridec-7-enoate with different cut strategies and frz-1 freeze strategy. The reference (solid black line) is evaluated on the complete molecular system with no frozen orbitals except the $1s$ core orbital for heavy atoms. frz-1/nocut curve (dotted line, $\square$ symbol) and frz-1/cut-5 (short dash line, $\times$ symbol) appear as superposed on the drawing scale. The other curves are frz-1/cut-3 (long dash line, $\bigcirc$ symbol) and frz-1/cut-1 (dot-dash line, $\Diamond$ symbol).

This is of particular evidence in the curve diagrams depicted in Fig. 2.5.4: the solid line is the nofrz/nocut reference, obtained by interpolating points from 0 to 180 (step 10 degrees) with a cubic spline curve. The other curves represent the incorrect behavior of frz-1 with cut strategies cut-1, cut-3, cut-5 and nocut (these last two curves are superposed). Each point has been generated using as a starting guess the converged orbitals from the previous point on the complete AO space (the ORTORBL matrix at convergence). Depending on the guess (70 or 110 degrees), two different curves can be obtained, giving a spike at 90 degrees. This reflects the dependence on the SCF solution, which affects the frozen orbital framework. It is however important to note that these curves are still parallel to the reference curve when far from the 90 degrees region.

As can be seen from Fig. 2.5.5, this incorrect behavior is drastically reduced by going from frz-1 to frz-3.

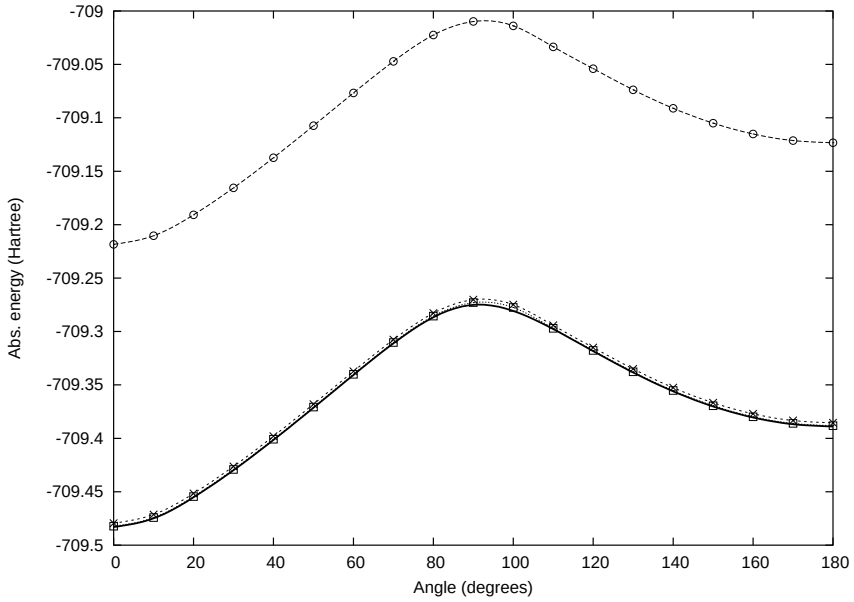


**Figure 2.5.5:** CAS+S energy curves (Hartree) for the cis-trans rigid interconversion of the (7Z)-13 ammoniotridec-7-enoate with different cut strategies and frz-3 freeze strategy. The reference (solid black line, see caption of Fig.2.5.4 for details), frz-3/nocut curve (dotted line, □ symbol) and frz-3/cut-5 (short dash line, × symbol) appear nearly superposed on the drawing scale. The other curve is frz-3/cut-3 (long dash line, ◯ symbol).

The curves are nearly parallel to the reference, but with a strong energy shift between cut-3 and cut-5. This difference is principally due to the four non-frozen C-H bonds (marked with the "a" labels in Fig. 2.5.3): they are described by hydrogen atoms (marked with the "b" labels) that have been cut out from the small system in the cut-3 analysis. This results in a shift in absolute energies, but does not affect the relative behavior.

Treating the same cut-3 system, but also adding the "b" hydrogen atoms gives CAS+S energies in better accord with the frz-3/cut-5 values, providing nearly half of the gap between frz-3/cut-3 and frz-3/cut-5.

| | cis | trans | diff |
|---|---|---|---|
| frz-3/cut-5 | -709.479395 | -709.385430 | 58.9248 |
| frz-3/cut-3+$H_b$ | -709.374870 | -709.280646 | 59.0873 |
| frz-3/cut-3 | -709.218388 | -709.123318 | 59.6173 |

**Table 2.5.3:** CAS+S absolute energies (Hartree) and energy difference (kcal/mol) between (7Z)-13 ammoniotridec-7-enoate cis and trans for frz-3/cut-3, frz-3/cut-5 and the intermediate frz-3/cut-3+$H_b$, where hydrogens labeled "b" in Fig. 2.5.3 have been preserved for the cut-3 strategy.

This confirms the importance of these four hydrogens for the evaluation of the absolute energy, but the relative behavior is unaffected. The same effect arises in all the diagonal values of the absolute energy tables, where a similar situation happens due to the proximity of the cut seam to the non-frozen orbitals.

The frz-3 selection has also been studied with a larger CAS space. This has been obtained by enriching the previous space with the C-C $\sigma$ and $\sigma^*$ orbitals, thus leading to a 4 electron in 4 orbitals space.

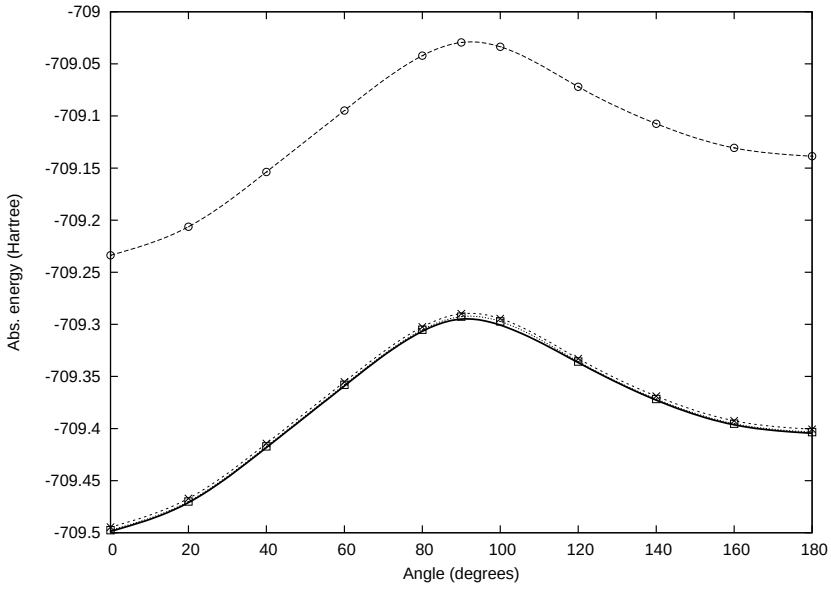


**Figure 2.5.6:** CAS+S energy curves (Hartree) for the cis-trans rigid interconversion of the (7Z)-13 ammoniotridec-7-enoate with an extended CAS defined as 4 electrons in 4 orbitals, with different cut strategies for the frz-3 freeze strategy. The reference (solid black line, see caption of Fig.2.5.4 for details), frz-3/nocut curve (dotted line, □ symbol) and frz-3/cut-5 (short dash line, × symbol) appear nearly superposed on the drawing scale. The other curve is frz-3/cut-3 (long dash line, ◯ symbol).

The evaluation has been performed at 90 degrees and from 0 to 180 degrees with a step of 20 degrees, interpolating the points with a cubic spline curve.

As can be seen from Fig. 2.5.6 the same correct behavior is obtained, the only difference being the shift of the energy due to the larger CAS space. The relative energy for the frz-3/cut-3 is, for example, 59.6149 kcal/mol. This value is in a very good accord with the one obtained using the CAS 2/2 space, 59.6173 kcal/mol.

Fig. 2.5.7 finally depicts the results for frz-5 strategy with the CAS 2/2 space, where frz-5/cut-5 curve lays higher in energy but still parallel to the reference curve, and the frz-5/nocut curve nearly superposed to the reference.

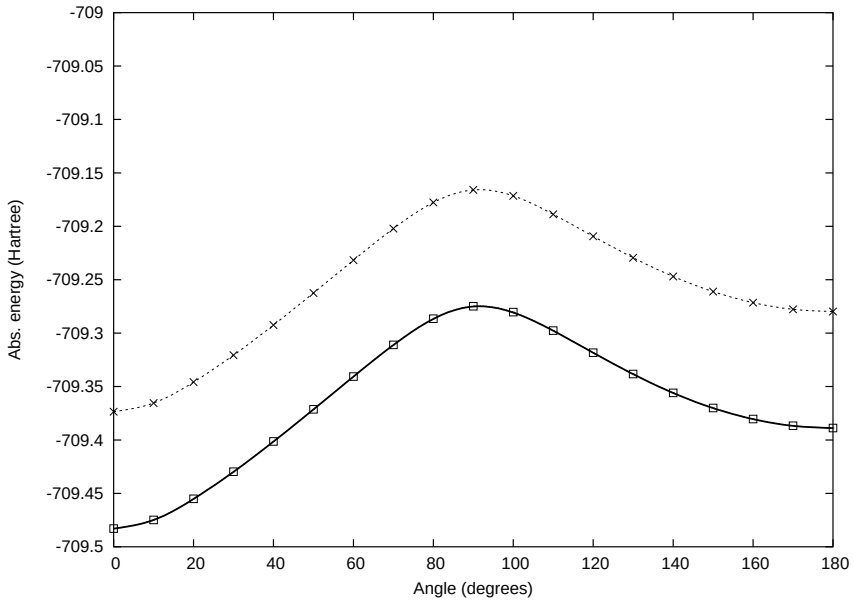


**Figure 2.5.7:** CAS+S energy curves (Hartree) for the cis-trans rigid interconversion of the (7Z)-13 ammoniotridec-7-enoate with different cut strategies and frz-5 freeze strategy. The reference (solid black line, see caption of Fig.2.5.4 for details) and frz-5/nocut curve (dotted line, $\square$ symbol) appear as superposed on the drawing scale. The other curve is frz-5/cut-5 (short dash line, $\times$ symbol)

Fig.2.5.8 shows the optimized $\pi$ orbital for the cis molecule with different cut strategies at frz-1 level of freeze. The plots have been realized with the Molden program[55] with a contour factor of 0.002. We can note that the cut-5 strategy does not change the orbital in a significant way, due to the graphically negligible expression of the orthogonalization tail of the orbital on the removed fragments. The cut-3 and cut-1 strategies show instead the confinement of the optimized orbital inside the group of atoms on which the projection was performed. The small lobes for the $\pi$ description are removed by the cut procedure, and the optimization preserves the locality imposed by the method.

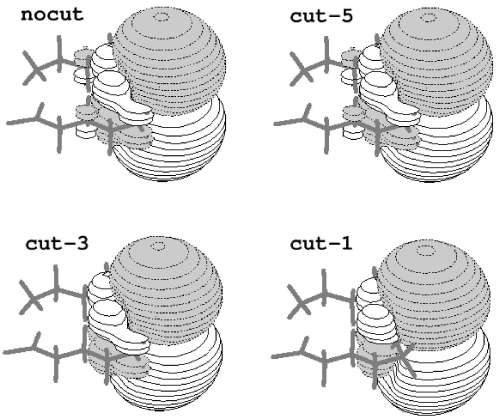


**Figure 2.5.8:** The $\pi$ orbital representation with respect to different cut strategies at frz-1 freeze strategy.

The improvements in timings are presented as an example in Tab. 2.5.4, applied to the (7Z)-13 ammoniotridec-7-enoate with the CAS 2/2 space.

As can be seen, performing the integral evaluation on the reduced system is a lightweight process. Nocut evaluations have a value of zero for the reduced system. The integral evaluation on the complete system can be used in this case, saving the cost of a new evaluation.

The remaining steps are quite fast, but the best compromise between result quality and time is for the frz-3/cut-3 evaluation. In this case a reduction of the time needed to perform the optimization is counterbalanced by the need of an integral evaluation on the small system, but the difference between the frz-3/cut-3 and the frz-3/nocut becomes dramatic when working on systems larger than the presented ones, which are the final targets of this technique.

| | seward-large | seward-small | large | small | CAS+S | total |
|---|---|---|---|---|---|---|
| nofrz/nocut | 1009 | 0 | 50 | 39 | 1062 | 2160 |
| frz-5/nocut | 1009 | 0 | 50 | 19 | 355 | 1433 |
| frz-5/cut-5 | 1009 | 433 | 50 | 6 | 225 | 1723 |
| frz-3/nocut | 1009 | 0 | 50 | 7 | 184 | 1250 |
| frz-3/cut-3 | 1009 | 100 | 50 | 1 | 99 | 1259 |
| frz-1/nocut | 1009 | 0 | 50 | 2 | 43 | 1104 |
| frz-1/cut-1 | 1009 | 8 | 50 | 0 | 18 | 1085 |

**Table 2.5.4:** Timings (in seconds) for evaluations performed on (7Z)-13 ammoniotridec-7-enoate, against different freeze and cut strategies. Test performed on Intel dual Xeon 2.8 GHz 2 GB RAM. "seward-large" and "seward-small" are the timings for the AO integral evaluation. "large" and "small" columns represents the timings needed for the creation of the transformed integrals which will seed the first iteration. "CAS+S" column reports the timings for the complete iterative procedure up to the convergence.

### 2.5.2 $C_{13}$ polyenal

A second test was performed on the highly conjugated $C_{13}$ polyenal

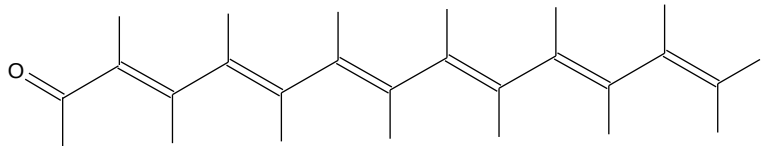


**Figure 2.5.9:** The $C_{13}$ transoid polyenal molecule.

Applications of the localization technique on polyenals have been performed.[41–43] This class of molecules is of particular interest due to their role in photobiology as chromophores.[56] Moreover, they are excellent model systems for studying the interaction between the $\pi$ system of the carbonyl group and the $\pi$ system of the unsaturated chain.[57]

The cisoid-transoid energy difference for the aldehydic group has been studied, by rotating by 180 degrees the C-C single bond. This evaluation should be very weakly affected by the length of the polyene chain. The distances (in Ångstrom) are r(C=O)=1.220, r(C-C)=1.450, r(C=C)=1.350, r(C-H)=1.100. All angles are 120 degrees. The cut and freeze denomination follows the analogy with shorter chain polyenals, as depicted in Fig. 2.5.10.

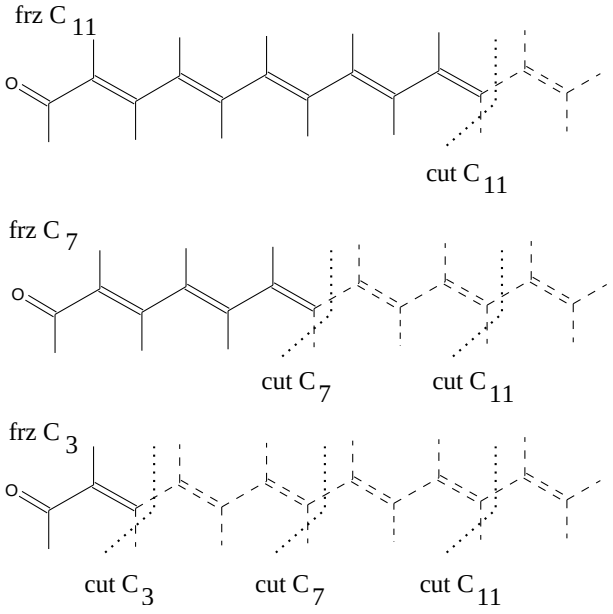


**Figure 2.5.10:** $C_{13}$ polyenal molecule. Freeze and cut strategy and labels follow the analogy with shorter chain polyenals. Localized orbitals expressed on atoms of the dashed line framework are frozen. Dotted lines depict the cut seam.

Two basis sets have been used: a minimal basis of type ANO-1 with $2s1p$ contraction for C and O, and a $1s$ contraction for hydrogen atoms (hereafter named ANO-small), and a larger one ANO-1 with $3s2p$ contraction for C and O, and a $2s$ contraction for hydrogens (ANO-large). As in the previous case, the $1s$ core orbitals for carbon and oxygen atoms have been frozen regardless of the freezing strategy.

Two CAS spaces have been considered. The first one, named CAS-A, contains two electrons in the $\pi/\pi^*$ orbitals of the carbonyl group. The second active space, named CAS-B, is defined as the previous one, further extended with the inclusion of two additional electrons and the $\pi/\pi^*$ orbitals from the double bond near the carbonyl.

A preliminary evaluation with the ANO-small basis set and the CAS-A active space has been performed on smaller polyenals, obtained by replacing the removed part of the molecule with a hydrogen atom. The obtained results are presented in Tab. 2.5.5.

| | cisoid | transoid | diff |
|---|---|---|---|
| $C_3$ | -190.457653 | -190.460042 | 1.4981 |
| $C_7$ | -343.919126 | -343.921129 | 1.2563 |
| $C_{11}$ | -497.381057 | -497.383042 | 1.2446 |
| $C_{13}$ | -574.112029 | -574.114015 | 1.2453 |

**Table 2.5.5:** CAS+S energy (Hartree) and difference (kcal/mol) between cisoid and transoid complete structure for smaller polyenals, using the CAS-A active space and the ANO-small basis set.

Tab. 2.5.6 shows the absolute values for the CAS-A with ANO-small basis set. Values are well reproduced, except when the cut is near the frozen/non-frozen seam. This confirms the preceding statement about this behavior. Also, the relative energies behave as expected. Again, we can see that in the most pronounced Freeze-and-Cut strategy the results are poor even in the relative energy.

Evaluation with the larger CAS-B (Tab. 2.5.7) and with the larger ANO-large basis set (Tab. 2.5.8) show the same behavior.

Finally Fig. 2.5.11 shows the $\pi$ orbital for different cut strategies, on the cisoid structure. Again, the Molden contour factor was 0.002, and the same behavior reported for the (7Z)-13 ammoniotridec-7-enoate can be appreciated. The cut-C11 strategy does not change the orbital in a significant way, due to the graphically negligible expression of the orthogonalization tail of the orbital on the removed part. The other strategies progressively restrict the orbital into the preserved part of the polyenal.

| | nofrz | frz $C_{11}$ | frz $C_7$ | frz $C_3$ |
|---|---|---|---|---|
| | | Cisoid | | |
| nocut | -574.112029 | -574.112025 | -574.111981 | -574.110699 |
| cut $C_{11}$ | | -573.831612 | -574.110383 | -574.110600 |
| cut $C_7$ | | | -573.826403 | -574.108770 |
| cut $C_3$ | | | | -573.837474 |
| | | Transoid | | |
| nocut | -574.114015 | -574.114009 | -574.113945 | -574.112686 |
| cut $C_{11}$ | | -573.833584 | -574.112346 | -574.112589 |
| cut $C_7$ | | | -573.828399 | -574.110797 |
| cut $C_3$ | | | | -573.838137 |
| | | Diff | | |
| nocut | 1.2453 | 1.2438 | 1.2367 | 1.2461 |
| cut $C_{11}$ | | 1.2367 | 1.2314 | 1.2474 |
| cut $C_7$ | | | 1.2513 | 1.2708 |
| cut $C_3$ | | | | 0.4161 |

**Table 2.5.6:** CAS+S absolute energies (Hartree) and energy difference (kcal/mol) between cisoid and transoid $C_{13}$ polyenal using the CAS-A active space and the ANO-small basis set, with respect to different freeze and cut strategies.

| | nofrz | frz $C_{11}$ | frz $C_7$ | frz $C_3$ |
|---|---|---|---|---|
| | | Cisoid | | |
| nocut | -574.173041 | -574.172995 | -574.172423 | -574.158115 |
| cut $C_{11}$ | | -573.892623 | -574.170812 | -574.157975 |
| cut $C_7$ | | | -573.887517 | -574.156091 |
| cut $C_3$ | | | | -573.889043 |
| | | Transoid | | |
| nocut | -574.174560 | -574.174521 | -574.174006 | -574.160153 |
| cut $C_{11}$ | | -573.894131 | -574.172393 | -574.160015 |
| cut $C_7$ | | | -573.889012 | -574.158166 |
| cut $C_3$ | | | | -573.888923 |
| | | Diff | | |
| nocut | 0.9522 | 0.9567 | 0.9919 | 1.2776 |
| cut $C_{11}$ | | 0.9459 | 0.9916 | 1.2795 |
| cut $C_7$ | | | 0.9378 | 1.3015 |
| cut $C_3$ | | | | -0.0756 |

**Table 2.5.7:** CAS+S absolute energies and energy difference (kcal/mol) between cisoid and transoid $C_{13}$ polyenal using the CAS-B active space and the ANO-small basis set with respect to different freeze and cut strategies.

| | nofrz | frz $C_{11}$ | frz $C_7$ | frz $C_3$ |
|---|---|---|---|---|
| | | Cisoid | | |
| nocut | -575.112195 | -575.112194 | -575.112159 | -575.111082 |
| cut $C_{11}$ | | -574.845588 | -575.108852 | -575.110774 |
| cut $C_7$ | | | -574.833700 | -575.106924 |
| cut $C_3$ | | | | -574.850622 |
| | | Transoid | | |
| nocut | -575.115498 | -575.115500 | -575.115451 | -575.114549 |
| cut $C_{11}$ | | -574.848881 | -575.112168 | -575.114263 |
| cut $C_7$ | | | -574.837108 | -575.110644 |
| cut $C_3$ | | | | -574.849650 |
| | | Diff | | |
| nocut | 2.0716 | 2.0738 | 2.0644 | 2.1742 |
| cut $C_{11}$ | | 2.0652 | 2.0792 | 2.1877 |
| cut $C_7$ | | | 2.1369 | 2.3328 |
| cut $C_3$ | | | | -0.6100 |

**Table 2.5.8:** CAS+S absolute energies and energy difference (kcal/mol) between cisoid and transoid $C_{13}$ polyenal using the CAS-A active space and the ANO-large basis set with respect to different freeze and cut strategies.

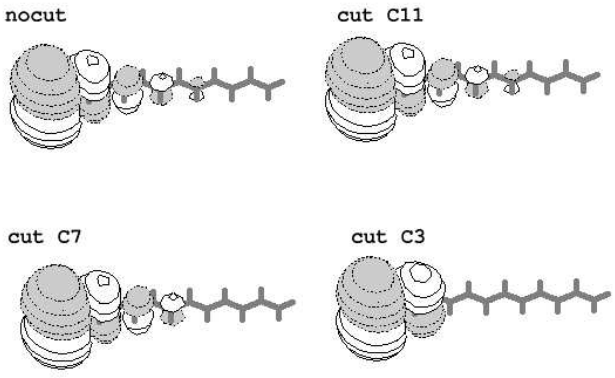


**Figure 2.5.11:** The $\pi$ orbital for the $C_{13}$ cisoid polyenal with respect to different cut strategies at frz-$C_3$ freeze strategy. The Molden contour factor for the plot is 0.002. The progressive neglect of the orbital expression on cut atoms can be appreciated.

### 2.5.3 Acetone + Water

To gain better insights about the sensibility of the technique with respect to the basis set, another test has been performed with the acetone molecule surrounded by six water molecules. Two molecules are directly coordinated to the carbonyl oxygen, while the remaining four coordinate the first two molecules.

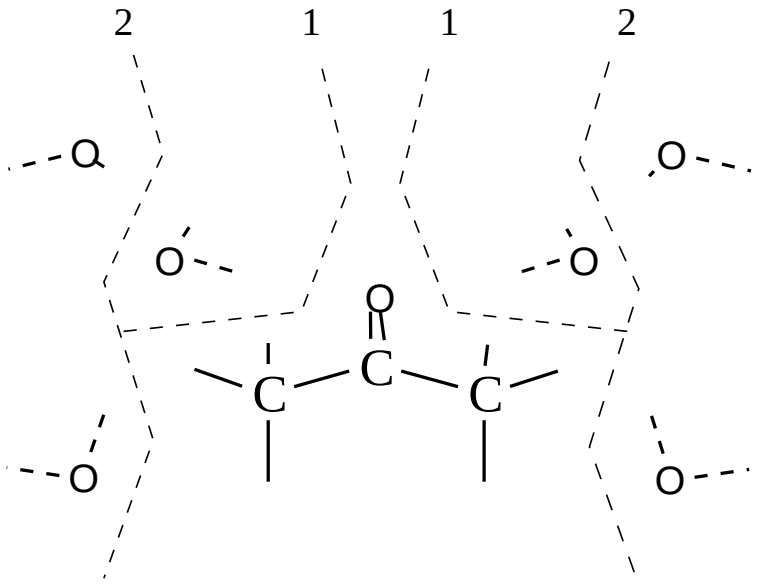


**Figure 2.5.12:** Scheme for freeze and cut of the acetone + 6 $H_2O$ system. Surrounding water molecules have been completely frozen, and two cut seams have been chosen to remove these molecules completely (seam "1") or partially (seam "2").

The freeze and cut scheme focuses on a full freezing of the water molecules. Two cut levels have been implemented: the first level completely removes the water system, and the second level preserves only the two carbonyl-coordinated molecules.

The active space included five orbitals (the oxygen $n_y$ lone pair, and carbonyl $\pi$, $\pi^*$, $\sigma$ and $\sigma^*$ orbitals) with six active electrons. A state average optimization involving the ground state and the $n \rightarrow \pi^*$ has been performed to evaluate this vertical transition energy.

Two ANO-1 basis sets contractions have been used: the minimal set C,O[$2s1p$], H[$1s$], hereafter called Bas-A, and an extended set C,O[$3s2p1d$], H[$2s1p$] (Bas-B). An additional set named Bas-C is further enriched up to C,O[$4s3p1d$], H[$2s1p$], and has been used on a smaller system made up of the acetone molecule and two water molecules.

Tab. 2.5.9 reports the values obtained. The behavior of the absolute energies is as expected by previous experience: when the cut seam is too near to the optimizable region, a difference in absolute energy is obtained.

| | GS | $n \rightarrow \pi^*$ | Diff. |
|---|---|---|---|
| | Bas-A | | |
| nofrz/nocut | -647.40071625 | -647.24175725 | 4.33 |
| frz-1/nocut | -647.39584150 | -647.23311463 | 4.43 |
| frz-1/cut-2 | -647.34998205 | -647.18699040 | 4.43 |
| frz-1/cut-1 | -647.28453399 | -647.11306062 | 4.67 |
| | Bas-B | | |
| frz-1/nocut | -648.52624523 | -648.34449917 | 4.94 |
| frz-1/cut-2 | -645.78542705 | -645.59889274 | 5.07 |
| frz-1/cut-1 | -644.71575515 | -644.51034075 | 5.59 |
| | Bas-B non-ortho | | |
| frz-1/cut-2 | -648.70019443 | -648.52408727 | 4.79 |
| frz-1/cut-1 | -648.93490599 | -648.73207516 | 5.52 |

**Table 2.5.9:** CAS+S absolute energies (Hartree) and energy difference (eV) between ground state and $n \rightarrow \pi^*$ state of acetone + 6 $H_2O$, using the minimal Bas-A basis set and the extended basis set Bas-B (see text). Bas-B non-ortho refers to an evaluation with non-orthonormal orbitals.

The relative behavior, however, is poorly affected. When a large basis set is used, the absolute energy drift is strongly increased. The larger Bas-B basis set produces a very steep variation compared to the minimal Bas-A.

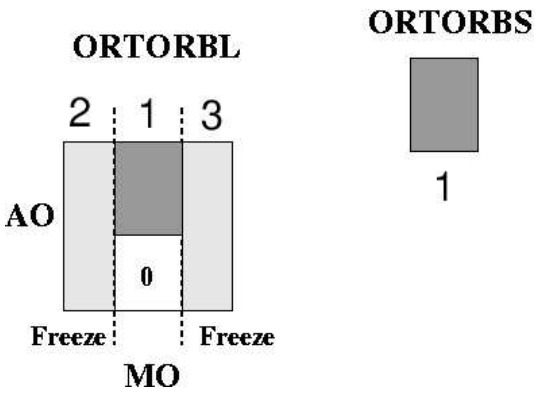


**Figure 2.5.13:** A visual representation of the hierarchical orthonormalization performed on the orbitals. The non-frozen orbitals (marked with "1" in figure) are orthonormalized among themselves. Then, frozen core orbitals (marked with "2" in figure) are orthonormalized among themselves and against the non-frozen ones. Finally, the frozen virtual orbitals (marked with "3") are orthonormalized against "1" and "2".

An additional effect must be considered responsible of the reported behavior. Further investigations, performed with the smaller system acetone + 2 $H_2O$, revealed a problem related to the molecular orbital hierarchical orthonormalization (Fig. 2.5.13). After the cut action, orbitals are no longer orthonormal, and the orthonormalization procedure aims at restoring this orthonormality.

The first class of orbitals performing the orthonormalization is the non-frozen zone (marked "1" in Fig. 2.5.13) which remains described in the smaller AO space. This class must be treated first, in order to preserve itself inside this restricted space.

The second class is the frozen core orbitals class (marked "2" in Fig. 2.5.13). When this class is orthonormalized with respect to the first one, orthonormalization tails are produced, effectively changing the core orbitals. The weight and spatial distribution of these tails is bound to the atomic basis set used to describe the molecular system, therefore larger basis sets produce stronger effects.

This situation is different from the previously reported phenomenon in 7Z-13 ammoniotridec-7-enoate in the frz-3/cut-3 analysis, where hydrogen atoms were removed from non-frozen orbitals. In the current case, frozen orbitals are modified due to orthonormality needs, which impose tails toward nearby atoms. These tails introduce an artificial charge shift toward these atoms, and in particular those involved in the physical process under study, resulting in convergence problems or a high energy drift. The process sometimes makes impossible for the evaluation to converge.

Molden plots have been performed on the smaller system with only two water molecules. The contour factor is 0.03. Fig. 2.5.14 shows a clear difference between the water OH bond with the Bas-C contraction and the Bas-A contraction.

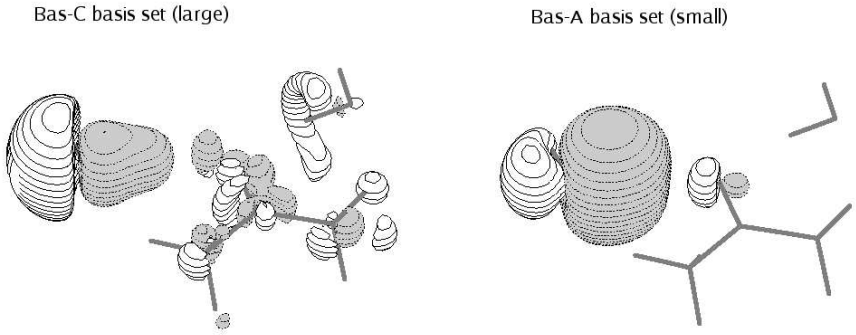


**Figure 2.5.14:** The water OH bond on the small acetone + $2H_2O$ system, using the large basis set Bas-C and the small Bas-A. The difference in spatial distribution of the orthonormalization tails is clearly visible.

The nature of this problem is independent of the frozen spacer between the removed zone and the non-frozen zone, but related to the frozen/non-frozen seam. The problem do not arise when cut is not performed, and consequently the resulting orbitals do not incur in orthonormalization.

At first glance, a possible solution is to keep the frozen/non-frozen seam as much far as possible from the region interested in the chemical phenomenon, leading to absolute energy drifts but prevent tails to act on the active space.

However, this goes in the direction of incrementing the computational cost of the final evaluation, enlarging the non-frozen zone in a large basis regime.

Another possible solution has been attempted working with a non-orthonormal basis set. Frozen orbitals are not orthonormalized with respect to the non-frozen ones, and the results obtained (Tab. 2.5.9 non-ortho) show a cleaner behavior. However, it must be pointed out that this procedure is not strictly formal under the assumptions made for the current theoretical deployment, and the convergence is very slow. An improvement in the theoretical asset is needed to formally work around this problem.

## 2.6 Conclusions

The Freeze-and-Cut technique has proved its usefulness in optimizing local orbitals at CASSCF or quasi-CASSCF level. It combines the freeze of the molecular orbitals that are not relevant for the phenomenon under study, and the cut of those atoms that are not needed to describe the optimized orbitals.

Different levels of theory can be used for the molecular framework and the physically interesting region of the system. The presented test cases demonstrate the ability of the method to give essentially the same results as a localized CASSCF calculation on the complete system, with a significant reduction of the computational cost.

Two problems must be kept into account in order to obtain reliable CASSCF energies: the first problem involves the tails of the unfrozen orbitals, which can have a significant delocalization on the nearest neighbor atoms. The projection of the unfrozen orbitals into a restricted set of atoms has the effect of rising the absolute energy, but the relative energies seems to be poorly affected.

The second problem is relative to the methodological need to reorthonormalize the frozen orbitals set after the cut is performed. This can introduce artificial tails which can create instability when occurring near the active orbitals. Both effects depend on the atomic basis set spatial distribution.

To face these problems, a general strategy can be devised:

- reduce the cut out of the unfrozen orbitals, setting the cut seam far from the unfrozen region. Atoms which are geometrically very distant from the unfrozen set can be usually cut out safely
- reduce the influence of the frozen orbitals orthonormalization tails on the chemically active region (usually, the active orbitals). To minimize this effect, a sufficient spacer of unfrozen orbitals should be kept.

Further evaluations are however needed to gain detailed comprehension of this strategy.

# Resumé chapitre 3
# NEVPT

Les théories perturbatives multiréférence sont des instruments importants pour l'évaluation de l'énergie de corrélation sur molécules petites et moyennes. L'avantage de l'approche perturbative est la réduction du coût computationnel: à parité de qualité des résultats, une approche variationnelle est plus lourde. Seulement en temps récents les approches perturbatives multireference sont utilisées intensivement dans la recherche en chimie quantique.

L'objectif des théories perturbatives multiréférence est la création d'une approche vite et propre, correspondante à l'approche à single déterminant Møller-Plesset (MP). Quand on travaille sur des systèmes bien décrits avec un seul déterminant, la théorie Møller-Plesset au deuxième ordre donne plus de 90 % de l'énergie de corrélation.

Quand le système étudié nécessite d'une approche multiréférence, et donc la fonction d'onde d'ordre zéro est multiréférence, le développement d'une théorie perturbative est plus complexe. Le problème est la définition d'un Hamiltonien d'ordre zéro approprié. Si ce choix n'est pas bon, on obtient une théorie perturbative que

- **n'est pas size consistent**: l'énergie de deux systèmes qui n'interagissent pas entre eux, AB, doit être égale al'énergie qui on obtient en faisant le calcul sur chacune et en faisant l'addition ($E_{AB} = E_A + E_B$)
- **à des états intrus**, quand les fonctions perturbatives ont des énergies presque égales à l'énergie de la fonction d'ordre zéro. Ce problème donne lieu à un comportement divergent.

La théorie n-Electron Valence state Perturbation Theory est une théorie perturbative multiréférence qui travaille sur une fonction d'onde de type Complete Active Space (CAS). Elle a été développée à Ferrara en collaboration avec l'université Paul Sabatier de Toulouse. La méthode adresse à la fois le problème de size consistency et le problème des états intrus, et donne l'approche Møller-Plesset si on effectue le calcul sur une fonction d'onde à un seul déterminant.

Le développement theoretique utilise un Hamiltonien bielectronique, l'Hamiltonien de Dyall, qui donne une théorie propre et efficiente. Les fonctions perturbatives sont de nature multiréférencielle, et peuvent être classifiée dans huit classes d'excitation, selon le schéma de promotion des electrons.

La théorie a deux niveaux de précision, "Strongly Contracted" et "Partially Contracted", selon la contraction de l'espace perturbatif. La Strongly Contracted NEVPT utilise un espace monodimensionnel pour chaque classe, et la Partially

Contracted utilise un espace à dimensionalité plus haute, mais plus petite que l'espace complet des determinants excités. L'approche Strongly Contracted donne des résultats de bonne qualité si comparée avec la Partially Contracted. Une différence entre eux est normalement un symptôme d'une mauvaise déscription à l'ordre-zero.

Dans cette thèse, diffèrent types de NEVPT ont été développé et utilisé:

- Single State NEVPT, qui fait la correction perturbative à un seul état électronique
- Quasi Degenerate NEVPT, qui fait cette correction à un ensemble d'états électroniques simultanément. Cette approche considère les interactions entre les fonctions d'onde après la perturbation, et permit le mélange de ce fonctions.
- Non Canonical NEVPT, est une extension de l'approche single état qui travaille avec des orbitales qui ne sont pas canoniques (elles ne diagonalisent pas la matrice de Fock généralisée). Cette approche est très utile pour l'integration entre Localisation et NEVPT, parce que les resultats sont invariables pour le mélange des orbitales de core et virtuelles, et donc il permit l'utilisation directe avec un ensemble d'orbitales localisées.

L'approche Single-State a été utilisée pour effectuer des calculs sur les molécules de formaldehyde, acetaldehyde et acétone. Tous ces molécules ont un groupe carbonyle, et différents stratégies ont été appliquées pour avoir une compréhension directe des états excités de ce groupe fondamental. En particulier, on a effectué l'évaluation de l'énergié des états excités de valence, soit adiabatiques que verticales, et des états Rydberg. Pour les états adiabatiques, on a aussi évalué l'énergie de point zéro (ZPE) vibrationnel, les géométries et le comportement vibrationnel des états excités de valence.

Avec l'approche Quasi-Degenerate on a étudié la courbe de potentiel du ground state, deux états rydberg et du $\pi \rightarrow \pi^*$ contre le stretching C-O, dans la molécule de formaldehyde. Le courbes obtenues à niveau CASSCF ont des croisements évités, et l'application de la NEVPT Quasi-Degenerate donne des courbes plutôt diffèrentes sur la position e la forme des ces croisements. Les résultats de la théorie QD-NEVPT peuvent être ameliorées avec une procedure iterative, qui peut être une bonne solution à des problémes dans les résultats obtenus.

Enfin, une évaluation préliminaire a été effectué avec la NEVPT Non-Canonique sur un système localisé. Les valeurs obtenues démontrent la validité de cette approche, qui sera fondamentale pour l'intégration directe de la théorie de localisation des orbitals moléculaires et la théorie NEVPT.

# Chapter 3

# NEVPT

Multireference perturbation theories are important tools for the evaluation of the correlation energy in small and medium sized molecules. The main advantage of a perturbative approach is the reduced computational cost when compared to a completely variational treatment, still providing a good level of accuracy. First attempts in deploying theoretical approaches for MRPT are quite old[58,59] but these approaches began to be mainstream only recently.

The goal of multireference perturbation theories is the creation of a quick and accurate tool resembling the single-determinant based Møller-Plesset (MP) approach,[60] extended to a multireference context: when the molecular system under study is well described by a single-reference treatment, second-order Møller-Plesset perturbation theory gives more than 90% of the correlation energy. The zero-order Hamiltonian is defined as the $n$ particles Fock operator

$$\hat{\mathcal{H}}_0 = \sum_{i=1}^{n} \hat{\mathcal{F}}(i) \tag{3.0.1}$$

which leads to the wavefunction correction

$$\Psi_n^{(1)} = -\sum_{k \neq n} \left|\Psi_k^{(0)}\right\rangle \frac{\left\langle \Psi_k^{(0)} \left| \hat{\mathcal{V}} \right| \Psi_n^{(0)} \right\rangle}{E_k^{(0)} - E_n^{(0)}} \tag{3.0.2}$$

and to the well known second-order Møller-Plesset correction for the energy

$$E_0^{(2)} = -\sum_{i,j}^{\text{occ}} \sum_{r,s}^{\text{virt}} \frac{|\langle rs \,||\, ij\rangle|^2}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j} \tag{3.0.3}$$

where $\epsilon$ are the orbital energies associated to each orbital

Unfortunately, with a multireference zero-order wavefunction the application of a perturbative approach is more difficult, due to the complex definition of a zero-order Hamiltonian. When a choice has been made, various kind of problems arise, preventing the deployment of an easy-to-use perturbative scheme for multireference wavefunctions.

These problems can be accounted as the following:

- **lack of size consistency** (strict separability), which requires the energy of two non-interacting systems AB to be equal to the sum of the energies of A and B treated independently ($E_{AB} = E_A + E_B$)
- "**intruder states**", when the perturbers happen to have very similar energies to the zero-order wavefunction, leading to divergent behavior due to the nearly zero denominator in the perturbation theory equations. The Møller-Plesset treatment is not affected by this problem, since the perturbers are the doubly excited determinants $\Phi_{ij}^{rs}$, and the corresponding energies are $E_0 + \epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j$, well distinct from $E_0$.
- **different levels of accuracy for ground and excited states**, making the evaluation of transition energy more complex.
- **no spin purity**, leading for example to a mix between singlet and triplet wavefunctions. This introduces additional complexity for spectroscopical evaluations, and spin-based property evaluations.
- **efficiency**. The perturbative treatment is required to be fast and with reduced disk space and memory footprints.

### Classification of perturbative methods

Multireference perturbation theories can be classified in two categories: "*perturb then diagonalize*", where an effective Hamiltonian is built and then diagonalized,[58,61] and "*diagonalize then perturb*", where the perturbative treatment is applied to a variationally optimized wavefunction.

To the latter category belong for example CIPSI techniques, started by Huron *et al.*[59] in Paris and subsequently developed in Pisa[62] and Ferrara.[63–66] CIPSI uses single determinants as perturbers, using a baricentric Møller-Plesset or a diagonal bielectronic (Epstein-Nesbet[67,68]) zero-order Hamiltonian. Although numerically efficient,[63–65,69] it lacks some formal requirement, like strict separability. Other analogous methods[70–74] suffer from similar problems.

Another "*diagonalize then perturb*" method is the CASPT2 approach.[75,76] It makes use of contracted perturbers obtained from a CASCI under the action of excitation operators, and a one-electron zero-order Hamiltonian. Due to its efficiency, CASPT2 is applicabile to large CAS spaces, providing good results for ground and excited states at a reasonable computational cost. Integration into the *ab initio* suite `MOLCAS` makes CASPT2 one of the most used post Hartree-Fock treatments. The main problem of the method is its sensibility to intruder states, leading to incorrect or divergent behavior.

$n$-Electron Valence Perturbation Theory (NEVPT) approach[77–79] is a "*diagonalize then perturb*" method developed and implemented in Ferrara and Toulouse. The method formally addresses both strict separability and intruder states problems, and also comes as a natural extension to Møller-Plesset. In fact, second-order NEVPT (NEVPT2) formally reproduces the Møller-Plesset equations when applied to a single-reference wavefunction.

In the development of NEVPT two levels of approximations have been considered:

- the degree of contraction of the perturber space, and therefore, of the perturber wavefunctions
- the choice of the zero-order Hamiltonian

## 3.1 Single-State NEVPT

Let $\Psi_m^{(0)}$ be a zero-order CASCI wavefunction

$$\Psi_m^{(0)} = \sum_{I \in \mathrm{CAS}} C_{I,m} \left| I \right\rangle \tag{3.1.1}$$

obtained diagonalizing $\hat{\mathcal{H}}$ inside the CAS

$$\hat{\mathcal{P}}_{\mathrm{CAS}} \hat{\mathcal{H}} \hat{\mathcal{P}}_{\mathrm{CAS}} \left| \Psi_m^{(0)} \right\rangle = E_m^{(0)} \left| \Psi_m^{(0)} \right\rangle \tag{3.1.2}$$

where $\hat{\mathcal{P}}_{\mathrm{CAS}}$ is the projector inside the CASCI space.

We can define perturbers in NEVPT as zero-order wavefunctions of the outer space (external to CAS) where $k$ electrons are removed from the inactive part and added to the valence part. If we decompose the zero-order CASCI wavefunction as an antisymmetrized product of an inactive part (made of both core and virtual orbitals) with $n_c$ electrons, and a valence part (made of active orbitals) with $n_v$ electrons

$$\left| \Psi_m^{(0)} \right\rangle = \left| \Phi_c \Psi_m^v \right\rangle \tag{3.1.3}$$

then the perturber wavefunctions can be expressed as

$$\left| \Psi_{l,\mu}^k \right\rangle = \left| \Phi_l^{-k} \Psi_\mu^{v+k} \right\rangle \tag{3.1.4}$$

where $l$ is a collective index that describes the orbitals involved in the operation and $\mu$ an enumerator index for the different wavefunctions. At second order of perturbation $-2 \leq k \leq 2$.

A graphical representation of the excitation scheme for NEVPT can be seen in Fig. 3.1.1

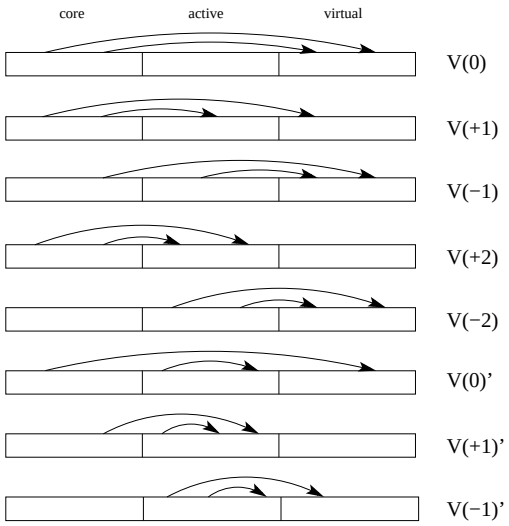


**Figure 3.1.1:** The eight excitation classes defining the perturber functions in NEVPT.

The perturbers functions span a space of determinants with the same inactive part $\Phi_l^{-k}$ and all the possible valence parts $\Psi_I^{+k}$. These spaces are denoted as $S_l^k$

$$S_l^k \equiv \left\{ \Phi_l^{-k} \Psi_I^{+k} \right\} \tag{3.1.5}$$

Perturbers will be defined in these spaces. If the full dimensionality is exploited, the resulting approach is called *totally uncontracted*. The perturbers and their energies can be obtained by diagonalizing the Hamiltonian inside each space

$$\hat{\mathcal{P}}_{S_l^k} \hat{\mathcal{H}} \hat{\mathcal{P}}_{S_l^k} \left| \Phi_l^{-k} \Psi_\mu^{v+k} \right\rangle = E_{l,\mu} \left| \Phi_l^{-k} \Psi_\mu^{v+k} \right\rangle \tag{3.1.6}$$

however, this procedure is impractical given its high computational cost. For each $S_l^k$ space, a diagonalization of the true Hamiltonian must be performed.

We can improve the computational asset by introducing a modified Hamiltonian, proposed by Dyall[52]

$$\hat{\mathcal{H}}^D = \hat{\mathcal{H}}_i^D + \hat{\mathcal{H}}_v^D + C \tag{3.1.7}$$

$$\hat{\mathcal{H}}_i^D = \sum_i^{\text{core}} \epsilon_i E_{ii} + \sum_r^{\text{virt}} \epsilon_r E_{rr} \tag{3.1.8}$$

$$\hat{\mathcal{H}}_v^D = \sum_{ab}^{\text{act}} h_{ab}^{\text{eff}} E_{ab} + \frac{1}{2} \sum_{abcd}^{\text{act}} \langle ab|cd\rangle \left( E_{ac} E_{bd} - \delta_{bc} E_{ad} \right) \tag{3.1.9}$$

$$C = 2 \sum_i^{\text{core}} h_{ii} + \sum_{ij}^{\text{core}} \left( 2 \langle ij|ij\rangle - \langle ij|ji\rangle \right) - 2 \sum_i^{\text{core}} \epsilon_i \tag{3.1.10}$$

where labels $i, j, \ldots$, $a, b, \ldots$, $r, s, \ldots$ denote core, active and virtual orbitals respectively, and this convention will be retained hereafter, $\epsilon_i$ and $\epsilon_r$ are the orbital energies of the involved orbitals, and $E_{mn}$ operators are the spin-traced operators $a^+_{m\alpha}a_{n\alpha} + a^+_{m\beta}a_{n\beta}$. These operators commute with $S^2$ and $S_z$, therefore the application of these operators on a spin-pure function produces again a spin-pure function. Finally, the $h^{\text{eff}}_{ab}$ matrix evaluates as

$$h^{\text{eff}}_{ab} = h_{ab} + \sum_j \left(2\langle aj|bj\rangle - \langle aj|jb\rangle\right) \tag{3.1.11}$$

$\hat{\mathcal{H}}^D$ behaves like the true Hamiltonian inside the CAS space, having the same eigenvalues and eigenvectors of the true Hamiltonian projected onto the CAS space. Also, given the decomposition for the wavefunction defined before, the action of the Dyall's Hamiltonian can be partitioned

$$\hat{\mathcal{H}}^D \left|\Phi_l^{-k}\Psi_\mu^{v+k}\right\rangle = E^k_{l,\mu}\left|\Phi_l^{-k}\Psi_\mu^{v+k}\right\rangle \tag{3.1.12}$$

stripping out the constant contribution of the inactive part and leaving a subsystem to be solved for the valence part

$$\hat{\mathcal{H}}^D_v \left|\Psi_\mu^{v+k}\right\rangle = E^k_\mu \left|\Psi_\mu^{v+k}\right\rangle \tag{3.1.13}$$

The total energy $E^k_{l,\mu}$ is the sum of $E^k_\mu$ and the energies of the orbitals involved in the definition of the inactive part $\Phi_l^{-k}$. This introduces the possibility to perform a single diagonalization of the valence Dyall's Hamiltonian on the CASCI zero-order wavefunction and evaluate the perturber energies using the property depicted above.

### Strongly Contracted approach

A possible choice in the development of the NEVPT approach is to choose a single function for each space $S^k_l$, leading to the *Strongly Contracted (SC)* scheme. A set of perturbative operators are used to produce a single function for each space, according to the following definition

$$\Psi^k_l = \hat{\mathcal{P}}_{S^k_l}\hat{\mathcal{H}}\Psi^{(0)}_m \tag{3.1.14}$$

where $\hat{\mathcal{P}}_{S^k_l}$ is the projector onto the subspace $S^k_l$. The same result can be obtained by applying a specific part of the Hamiltonian to the zero-order wavefunction

$$\Psi^k_l = V_l{}^k\Psi^{(0)}_m \tag{3.1.15}$$

For each space, an appropriate operator can be devised

$$V^0_{ijrs} = \gamma_{ij}\gamma_{rs}\left(\langle rs|ij\rangle E_{ri}E_{sj} + \langle rs|ji\rangle E_{si}E_{rj}\right) \quad i \le j, r \le s \tag{3.1.16}$$

$$V^{-1}_{rsi} = \gamma_{rs}\sum_a \left(\langle rs|ia\rangle E_{ri}E_{sa} + \langle sr|ia\rangle E_{si}E_{ra}\right) \quad r \le s \tag{3.1.17}$$

$$V^1_{ijr} = \gamma_{ij}\sum_a \left(\langle ra|ji\rangle E_{rj}E_{ai} + \langle ra|ij\rangle E_{ri}E_{aj}\right) \quad i \le j \tag{3.1.18}$$

$$V^{-2}_{rs} = \gamma_{rs}\sum_{ab} \langle rs|ba\rangle E_{rb}E_{sa} \quad r \le s \tag{3.1.19}$$

$$V^2_{ij} = \gamma_{ij}\sum_{ab} \langle ba|ij\rangle E_{bi}E_{aj} \quad i \le j \tag{3.1.20}$$

$$V^0_{ir} = \sum_{ab} \left(\langle ra|ib\rangle E_{ri}E_{ab} + \langle ra|bi\rangle E_{ai}E_{rb}\right) + h^{\text{eff}}_{ri}E_{ri} \tag{3.1.21}$$

$$V^{-1}_{r} = \sum_{abc} \langle ra|bc\rangle E_{rb}E_{ac} + \sum_a h^{\text{eff'}}_{ra}E_{ra} \tag{3.1.22}$$

$$V^1_{i} = \sum_{abc} \langle ba|ic\rangle E_{bi}E_{ac} + \sum_a h^{\text{eff}}_{ai}E_{ai} \tag{3.1.23}$$

where $\gamma_{mn} = 1 - \frac{1}{2}\delta_{mn}$.

Auxiliary matrixes have been also defined as

$$h^{\text{eff}}_{mn} = h_{mn} + \sum_j \left(2\langle mj|nj\rangle - \langle mj|jn\rangle\right) \tag{3.1.24}$$

$$h^{\text{eff'}}_{mn} = h^{\text{eff}}_{mn} - \sum_b \langle mb|bn\rangle \tag{3.1.25}$$

The perturber functions are not normalized. The norm of these functions

$$N^k_l = \left\langle \Psi^k_l \middle| \Psi^k_l \right\rangle = \left\langle \Psi^{(0)}_m \middle| \left(V^k_l\right)^+ V^k_l \middle| \Psi^{(0)}_m \right\rangle \tag{3.1.26}$$

plays an important role in the Strongly Contracted development. The values of the norms for each function are

$$N^0_{ijrs} = 4\gamma_{ij}\gamma_{rs}\left(\langle rs|ij\rangle^2 + \langle rs|ji\rangle^2 - \langle rs|ij\rangle\langle rs|ji\rangle\right) \tag{3.1.27}$$

$$N^{-1}_{rsi} = \gamma_{rs}\sum_{ab}\left[2\left(\langle rs|ib\rangle\langle rs|ia\rangle + \langle sr|ib\rangle\langle sr|ia\rangle\right) - \langle rs|ib\rangle\langle sr|ia\rangle - \langle sr|ib\rangle\langle rs|ia\rangle\right] R^{(1)}_{ba} \tag{3.1.28}$$

$$N^1_{ijr} = \gamma_{ij}\sum_{ab}\left[2\left(\langle rb|ji\rangle\langle ra|ji\rangle + \langle rb|ij\rangle\langle ra|ij\rangle\right) - \langle rb|ji\rangle\langle ra|ij\rangle - \langle rb|ij\rangle\langle ra|ji\rangle\right] \tilde{R}^{(1)}_{ba} \tag{3.1.29}$$

$$N^{-2}_{rs} = \gamma_{rs}\sum_{abcd}\langle rs|ba\rangle\langle rs|dc\rangle R^{(2)}_{abcd} \tag{3.1.30}$$

$$N^2_{ij} = \gamma_{ij}\sum_{abcd}\langle ba|ij\rangle\langle dc|ij\rangle \tilde{R}^{(2)}_{abcd} \tag{3.1.31}$$

$$
\begin{aligned}
N_{ir}^{0} &= \sum_{abcd}\Big[\left(2\left\langle ra|ib\right\rangle\left\langle rc|id\right\rangle-\left\langle ra|ib\right\rangle\left\langle rc|di\right\rangle-\left\langle ra|bi\right\rangle\left\langle rc|id\right\rangle\right)\left\langle\Psi_m^{(0)}\left|E_{ba}E_{cd}\right|\Psi_m^{(0)}\right\rangle \\
&\quad+\left\langle ra|bi\right\rangle\left\langle rc|di\right\rangle\left(\left\langle\Psi_m^{(0)}\left|E_{bd}\tilde{E}_{ac}\right|\Psi_m^{(0)}\right\rangle+\delta_{ab}R_{ba}^{(1)}\right)\Big] \\
&\quad+2\sum_{ab}\left(2\left\langle ra|ib\right\rangle-\left\langle ra|bi\right\rangle\right)h_{ri}^{\text{eff}}R_{ba}^{(1)}+2\left(h_{ri}^{\text{eff}}\right)^2 \qquad (3.1.32)\\
N_{r}^{-1} &= \sum_{abcdef}\left\langle rd|ef\right\rangle\left\langle ra|bc\right\rangle\left\langle\Psi_m^{(0)}\left|E_{fd}E_{eb}E_{ac}\right|\Psi_m^{(0)}\right\rangle \\
&\quad+2\sum_{abcd}\left\langle ra|bc\right\rangle h_{ra}^{\text{eff'}}\left\langle\Psi_m^{(0)}\left|E_{ca}E_{bd}\right|\Psi_m^{(0)}\right\rangle \\
&\quad+\sum_{ab}h_{ra}^{\text{eff'}}h_{rb}^{\text{eff'}}R_{ba}^{(1)} \qquad (3.1.33)\\
N_{i}^{1} &= \sum_{abcdef}\left\langle ed|if\right\rangle\left\langle ba|ic\right\rangle\left\langle\Psi_m^{(0)}\left|E_{fd}\tilde{E}_{eb}E_{ac}\right|\Psi_m^{(0)}\right\rangle \\
&\quad+2\sum_{abcd}\left\langle ba|ic\right\rangle h_{di}^{\text{eff}}\left\langle\Psi_m^{(0)}\left|E_{ca}\tilde{E}_{bd}\right|\Psi_m^{(0)}\right\rangle \\
&\quad+\sum_{ab}h_{ai}^{\text{eff}}h_{bi}^{\text{eff}}\tilde{R}_{ba}^{(1)} \qquad (3.1.34)
\end{aligned}
$$

To evaluate these norms, the spinless density matrix of rank not higher than three between the $\Psi_m^{(0)}$ functions are needed. A short notation has been used for the hole density matrixes

$$
\begin{aligned}
\tilde{R}_{ab}^{(1)} &= 2\delta_{ab}-R_{ba}^{(1)} \qquad (3.1.35)\\
\tilde{R}_{abcd}^{(2)} &= R_{abcd}^{(2)}+\delta_{ad}R_{cb}^{(1)}+\delta_{bc}R_{da}^{(1)}-2\delta_{ac}R_{db}^{(1)} \qquad (3.1.36)\\
&\quad-2\delta_{bd}R_{ca}^{(1)}-2\delta_{ad}\delta_{bc}+4\delta_{ca}\delta_{db} \qquad (3.1.37)
\end{aligned}
$$

and for the operator $\tilde{E}_{ab}=2\delta_{ab}-E_{ba}$

An important property of the $\Psi_l^k$ is that any other function of the space $S_l^k$ which is orthogonal to $\Psi_l^k$ do not interact with the zero-order wavefunction through the true Hamiltonian. It is possible to use the $\Psi_l^k$ functions as a basis set for the expansion of the first-order correction to the wavefunction, and also for the expression of the zero-order Hamiltonian by means of a spectral decomposition

$$
\hat{\mathcal{H}}_0=\sum_{lk}\left|\Psi_l^{k\prime}\right\rangle E_l^k\left\langle\Psi_l^{k\prime}\right|+\sum_m\left|\Psi_m^{(0)}\right\rangle E_m^{(0)}\left\langle\Psi_m^{(0)}\right| \qquad (3.1.38)
$$

where $\left|\Psi_l^{k\prime}\right\rangle$ are the normalized $\left|\Psi_l^k\right\rangle$.

The expression for the first-order correction to the wavefunction is therefore

$$
\begin{aligned}
\Psi_m^{(1)} &= \sum_{kl}\left|\Psi_l^{k\prime}\right\rangle\frac{\left\langle\Psi_l^{k\prime}\left|\hat{\mathcal{H}}\right|\Psi_m^{(0)}\right\rangle}{E_m^{(0)}-E_l^k} \qquad (3.1.39)\\
&= \sum_{kl}\left|\Psi_l^{k\prime}\right\rangle\frac{\sqrt{N_l^k}}{E_m^{(0)}-E_l^k} \qquad (3.1.40)
\end{aligned}
$$

and for the energy is

$$E_m^{(2)} = \sum_{kl} \frac{\left\langle \Psi_l^{k\prime} \middle| \hat{\mathcal{H}} \middle| \Psi_m^{(0)} \right\rangle^2}{E_m^{(0)} - E_l^k} = \sum_{kl} \frac{N_l^k}{E_m^{(0)} - E_l^k} \tag{3.1.41}$$

however, we still miss a definition of the perturber energies $E_l^k$. These energies can be defined in a computationally advantageous approach, by means of the Dyall's Hamiltonian.

$$E_l^k = \frac{1}{N_l^k} \left\langle \Psi_l^k \middle| \hat{\mathcal{H}}^D \middle| \Psi_l^k \right\rangle \tag{3.1.42}$$

which leads to

$$N_l^k E_l^k = \left\langle \Psi_m^{(0)} \middle| V_l^{k+} \hat{\mathcal{H}}^D V_l^k \middle| \Psi_m^{(0)} \right\rangle \tag{3.1.43}$$

$$= \left\langle \Psi_m^{(0)} \middle| V_l^{k+} V_l^k \hat{\mathcal{H}}^D \middle| \Psi_m^{(0)} \right\rangle + \left\langle \Psi_m^{(0)} \middle| V_l^{k+} \left[ \hat{\mathcal{H}}^D, V_l^k \right] \middle| \Psi_m^{(0)} \right\rangle \tag{3.1.44}$$

Developing the braket and extracting the inactive part of the Dyall's Hamiltonian we obtain

$$E_l^k = E_m^{(0)} + \Delta\epsilon_l + \frac{1}{N_l^k} \left\langle \Psi_m^{(0)} \middle| V_l^{k+} \left[ \hat{\mathcal{H}}_v, V_l^k \right] \middle| \Psi_m^{(0)} \right\rangle \tag{3.1.45}$$

with $\Delta\epsilon_l$ is the sum of the orbital energies of the newly occupied virtual orbitals minus the orbital energies of the unoccupied core orbitals.

The term that still need to be evaluated is the braket involving the commutator. After development of each $V$ operator and substitution, we obtain the contributions to the energy. Here we present the result for $V_{rsi}^{(-1)}$:

$$\begin{aligned} \left\langle \Psi_m^{(0)} \middle| V_{rsi}^{(-1)} \left[ \hat{\mathcal{H}}_v, V_{rsi}^{(-1)} \right] \middle| \Psi_m^{(0)} \right\rangle &= \gamma_{rs} \sum_{ab} \left[ 2 \left( \langle rs|ia\rangle \langle rs|ib\rangle + \langle sr|ia\rangle \langle sr|ib\rangle \right) \right. \\ &\quad \left. - \left( \langle rs|ia\rangle \langle sr|ib\rangle + \langle sr|ia\rangle \langle rs|ib\rangle \right) \right] K_{ab} \end{aligned} \tag{3.1.46}$$

where $K_{ab}$ is analogous to a spin-free formulation of the Koopmans matrix, defined only in the active space

$$K_{ab} = \frac{1}{2} \left\langle \Psi_m^{(0)} \middle| E_{as} E_{ir} \left[ \hat{\mathcal{H}}_v, E_{ri} E_{sb} \right] \middle| \Psi_m^{(0)} \right\rangle \tag{3.1.47}$$

$$= \sum_c h_{ac}^{\text{eff}} R_{bc}^{(1)} - \sum_{cde} \langle cb|ed\rangle \, R_{acde}^{(2)} \tag{3.1.48}$$

The contribution of the $\Psi_{rsi}^{-1}$ is therefore given by

$$E_m^{(2)} \left( S_{rsi}^{-1} \right) = -\frac{N_{rsi}^{-1}}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_{rsi}^{-1}} \tag{3.1.49}$$

with

$$\epsilon_{rsi}^{-1} = \frac{1}{N_l^k} \left\langle \Psi_m^{(0)} \middle| V_{rsi}^{-1+} \left[ \hat{\mathcal{H}}_v, V_{rsi}^{-1} \right] \middle| \Psi_m^{(0)} \right\rangle \tag{3.1.50}$$

An interesting case is the $V^{(0)}_{ijrs}$ case, which is identical to the Møller-Plesset contribution

$$E^{(2)}_m \left( S^0_{rsij} \right) = -\frac{N^0_{rsij}}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j} \tag{3.1.51}$$

NEVPT2 can therefore be seen as a generalized form of MP2 to multireference wavefunctions, and this result is also found in the completely uncontracted and Partially Contracted approaches.[80]

The cases involving the $V^{-2}_{rs}$ and $V^{2}_{ij}$ operators require the knowledge of the three-particles density matrix, and the $V^{-1}_{r}$ and $V^{1}_{i}$ require auxiliary matrixes depending on the four-particles density matrixes. All these matrixes always involve only the active indexes, therefore represent only a minor issue for nowadays computers.

### Partially Contracted approach

An alternative approach, named *Partially Contracted (PC)*, is to use a subspace $\overline{S}^k_l$ of $S^k_l$. To define this subspace, a set of functions $\Phi$ is generated by means of the $V^k_l$ operators. These functions must be orthonormalized and purged of linear dependencies which may arise. The resulting set spans the $\overline{S}^k_l$ space.

Once the $\overline{S}^k_l$ spaces have been defined, we can obtain a set of perturbers

$$\hat{\mathcal{P}}_{\overline{S}^k_l} \hat{\mathcal{H}} \hat{\mathcal{P}}_{\overline{S}^k_l} \left| \Psi^k_{l\mu} \right\rangle = E^k_{l,\mu} \left| \Psi^k_{l\mu} \right\rangle \tag{3.1.52}$$

where we can substitute $\hat{\mathcal{H}}$ with $\hat{\mathcal{H}}^D$ to simplify the evaluation. Perturbers are generated by the application of the excitation operators without contraction. For example, in the case of the $V^{-1}_{rsi}$ operator

$$V^{-1}_{rsi} = \gamma_{rs} \sum_a \left( \langle rs|ia\rangle \, E_{ri} E_{sa} + \langle sr|ia\rangle \, E_{si} E_{ra} \right) \quad r \leq s \tag{3.1.53}$$

the Partially Contracted approach makes use of functions $\Phi_{risa} = E_{ri} E_{sa} \Psi^{(0)}_m$ and $\Phi_{risa} = E_{si} E_{ra} \Psi^{(0)}_m$. We can group these functions in two row vectors $\mathbf{\Phi}_{ris}$ and $\mathbf{\Phi}_{sir}$ to write the expression of the overlap and $\hat{\mathcal{H}}^D$ matrixes

$$\langle \mathbf{\Phi}_{ris}, \mathbf{\Phi}_{sir} | \mathbf{\Phi}_{ris}, \mathbf{\Phi}_{sir} \rangle = \begin{bmatrix} 2R^{(1)} & -R^{(1)} \\ -R^{(1)} & 2R^{(1)} \end{bmatrix} \tag{3.1.54}$$

$$\begin{aligned} \left\langle \mathbf{\Phi}_{ris}, \mathbf{\Phi}_{sir} \left| \hat{\mathcal{H}}^D \right| \mathbf{\Phi}_{ris}, \mathbf{\Phi}_{sir} \right\rangle \;=\; & \begin{bmatrix} 2\mathbf{K} & -\mathbf{K} \\ -\mathbf{K} & 2\mathbf{K} \end{bmatrix} \\ & + \left( E^{(0)}_m + \epsilon_r + \epsilon_s - \epsilon_i \right) \begin{bmatrix} 2\mathbf{R}^{(1)} & -\mathbf{R}^{(1)} \\ -\mathbf{R}^{(1)} & 2\mathbf{R}^{(1)} \end{bmatrix} \end{aligned} \tag{3.1.55}$$

where $\mathbf{R}^{(1)}$ is the one-particle density matrix in the spinless formulation and the $\mathbf{K}$ matrix is defined as 3.1.47 The diagonalization of the $\hat{\mathcal{H}}^D$ matrix is quickly obtained diagonalizing $\mathbf{K}$ in the metric $\mathbf{R}^{(1)}$ to obtain $\mathbf{c}_\mu$ coefficients and $\epsilon_\mu$ energies

$$\mathbf{K}\mathbf{c}_\mu = -\epsilon_\mu \mathbf{R}^{(1)}\mathbf{c}_\mu \tag{3.1.56}$$

to obtain two distinct orthonormal functions

$$\Psi^{-1}_{ris\mu} = \frac{1}{\sqrt{2}} \sum_a c_{a\mu} \left(\Phi_{risa} + \Phi_{sira}\right) \tag{3.1.57}$$

$$\Psi^{-1\ \prime}_{ris\mu} = \frac{1}{\sqrt{6}} \sum_a c_{a\mu} \left(\Phi_{risa} - \Phi_{sira}\right) \tag{3.1.58}$$

These functions will be used in the evaluation of the interaction with the zero-order wavefunction. We define

$$\begin{aligned} (rs, \mu i) &= \left\langle \Psi^{-1}_{ris\mu} \middle| \hat{\mathcal{H}} \middle| \Psi^{(0)}_m \right\rangle && (3.1.59)\\ &= \frac{1}{\sqrt{2}} \sum_a \left(\langle rs|ia\rangle + \langle sr|ia\rangle\right) S_{a\mu} && (3.1.60) \end{aligned}$$

$$\begin{aligned} (rs, \mu i)' &= \left\langle \Psi^{-1\ \prime}_{ris\mu} \middle| \hat{\mathcal{H}} \middle| \Psi^{(0)}_m \right\rangle && (3.1.61)\\ &= \sqrt{\frac{3}{2}} \sum_a \left(\langle rs|ia\rangle - \langle sr|ia\rangle\right) S_{a\mu} && (3.1.62) \end{aligned}$$

where

$$S_{a\mu} = \sum_{a'} c^*_{a\mu} R^{(1)}_{a'a} \tag{3.1.63}$$

We can now finally write the contribution for this subspace to the second-order energy

$$E^{(2)}(\overline{S}^{-1}_{rsi}) = -\sum_\mu \frac{(rs, \mu i)^2 + (rs, \mu i)'^2}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_\mu} \tag{3.1.64}$$

## 3.2 Evaluation: carbonyl valence excited states

*This section is based on the articles Refs. 81 and 82*

Small carbonyl molecules formaldehyde, acetaldehyde and acetone attracted both experimental and theoretical chemists since long time, due to their particular properties and roles.

For instance the behavior of the carbonyl is important in order to describe photochemical reactions such as Norrish type-I reactions in aliphatic ketones and especially acetone,[83–85] where the cleavage of one of the $\alpha$ CC bonds is related to an

excitation to the $S_1$ $n \rightarrow \pi^*$ state. A similar kind of reaction is the formaldehyde decomposition in $H_2$ + CO or H + HCO,[86] mainly interesting for high-atmosphere chemistry and astronomy.

From the computational point of view, the small size of these systems has allowed for a long time the application of high-level theoretical approaches. For these reasons, a consistent number of evaluations are available in bibliography, defining an interesting testcase for comparisons and discussions.

Despite these accurate insights, doubts still exist on some photochemical features of such small molecules. For example, the absence of a $\pi \rightarrow \pi^*$ transition from the experimental spectrum is still unsolved. The first hypothesis for its absence,[87] was that it should lie over the first ionization limit of 10.88 eV[88,89] but it was discarded thanks to refinements in theoretical evaluations, producing a value for this transition in the 9.5-9.8 eV range, well below 10.88 eV.

A more probable explanation calls for a combined action of mixing between valence states and Rydberg states (electronically excited states where the promoted electron is described by a spatially diffuse orbital), and a very long progression due to geometrical distortion of the excited state. The first effect reduces the absorption due to intensity borrowing from the $\pi \rightarrow \pi^*$ to Rydberg states. The second effect produces Franck-Condon factors with a very long and broad progression, mainly due to the change in bond length and strength after the excitation. The double-bond nature changes to single-bond in the excited state, and the resulting Potential Energy Surface along the CO stretching coordinate is broad and with a longer CO distance with respect to the ground state minimum. The molecule in the excited state is also highly bent along the out-of-plane coordinate, further increasing the effect.

As a final result, the $\pi \rightarrow \pi^*$ probably expresses itself in the experimental spectrum as a broad and low-intensity band, hardly detectable in a spectral region dominated by $n \rightarrow$ Ryd bands. Most of the research performed on these molecules during this thesis was finalized to gather a deeper insight into these problems.

From a historical point of view, the most largely evaluated electronic transition for simple carbonyl molecules is the $n_y \rightarrow \pi^*$ from the oxygen lone pair. This transition is computationally very accessible, and experimentally detectable. For two of the three molecules under investigation, formaldehyde and acetone, the $C_{2v}$ symmetry group forbids this excitation, but the transition is observed due to the intensity borrowing performed by other states interacting via vibronic coupling. For acetaldehyde, the symmetry group is $C_s$ and the transition is allowed by symmetry, but the local symmetry of the carbonyl determines the selection rule and reduces the intensity.[90]

Among the properties of excited states, the equilibrium geometrical structures

and the vibrational frequencies are also of particular interest, since they help in the interpretation of the photochemical behavior of the molecules and of their electronic spectra. Within the harmonic approximation, only the energy Hessian at the equilibrium geometry is required for the calculation of the vibrational frequencies. The comparison of the theoretical frequencies so obtained with the measured ones however is not straightforward: the anharmonicity of the potential, which is not taken into account in the harmonic approximation, can have a sizable effect even on the first vibrational energy difference. In the particular cases when a series of vibrational levels for a given normal mode is experimentally known, the "harmonic experimental frequencies" can be computed (see for instance Refs. 91 and 92). These quantities can be directly compared with the theoretically computed ones.

### Evaluation details

In this section we present various analyses performed on the valence excited states of formaldehyde, acetaldehyde, and acetone:

- single state NEVPT2 study of the vertical $^1(n \rightarrow \pi^*)$ and $^3(n \rightarrow \pi^*)$transitions
- a detailed single state NEVPT2 study of $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ adiabatic singlet and triplet states
- geometry optimization of ground and excited states at CASSCF level
- vibrational analysis for ground and excited states at CASSCF level to evaluate the Zero Point Energy correction

Particular care has been devoted to maintain the computational parameters as homogeneous as possible during the evaluation, in order to perform a systematic analysis for similarities and differences, evaluate trends and finally allowing a deep comparison between NEVPT2 and other theoretical methods.

Formaldehyde, acetaldehyde and acetone have been studied using an ANO-1 basis set,[51] contracted with the C,O (14s9p4d)/[4s3p1d], H (8s4p)/[2s1p] scheme. No diffuse functions have been added in order to simplify the evaluation, removing Rydberg functions from the orbital scheme.

Rydberg states are expected to have a marginal influence at the equilibrium geometry of the ground and excited valence states.[93] This is due to various factors: first, the $n \rightarrow$ Ryd excitations remove an electron from a non-bonding orbital, producing Potential Energy Surfaces nearly parallel to the Ground State surface. At the equilibrium geometries of the valence excited states (whose minima are very different in geometry from that of the Ground State) Rydberg states are expected to have very high energies, making valence/Rydberg mixing negligible.

Second, as clearly shown by Hachey *et al*,[94,95] the perturbation of the Rydberg states originates from a series of avoided crossings along the CO stretching coordinate where valence states are almost degenerate with the Rydberg ones. The $n \rightarrow \pi^*$ vertical singlet excitation is expected around 4 eV, lower than the 8-9.5 range for the $n \rightarrow$ Ryd states. Valence vertical excitations laying higher in energy could be more sensitive to Rydberg coupling in the Franck-Condon region, but the strong dependence of valence/Rydberg mixing with respect to the geometry makes Potential Surface alterations almost negligible when not explicitly looking for avoided crossing situations.

We can therefore conclude that the electronic valence/Rydberg mixing is not very influent both at the GS equilibrium geometry and at the valence excited states equilibrium geometries, acting poorly on vertical energies, adiabatic energies, equilibrium geometries, and harmonic vibrational frequencies.

The equilibrium geometry of the ground state (GS) and of the first three valence singlet [$^1(n \rightarrow \pi^*)$, $^1(\pi \rightarrow \pi^*)$ and $^1(\sigma \rightarrow \pi^*)$] and triplet [$^3(n \rightarrow \pi^*)$, $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$] excited states have been evaluated at CASSCF level using the `DALTON` program.[96] At the time of the evaluation, the development version used for this study included an experimental direct merge of NEVPT2 into the `DALTON` codebase. This study also confirms the effectiveness of this interface and its simplicity for the user.

For all compounds, the molecular skeleton in the ground state (H-CO-H for formaldehyde, C-CO-H for acetaldehyde and C-CO-C for acetone) lies in the $yz$ plane, with the carbonyl group along the $z$ axis. For acetaldehyde and acetone, the hydrogen atoms lying on the $yz$ plane are at the maximum distance from the $z$ axis.

The symmetry group is $C_{2v}$ for formaldehyde and acetone, and $C_s$ for acetaldehyde. For excited state equilibrium geometries, where the molecular skeleton is non-planar, the $C_s$ (formaldehyde and acetone) and $C_1$ (acetaldehyde) symmetry groups have been used. The motion of the CO bond out of the $CR_2$ plane (wagging) is described by the out-of-plane bending angle $\theta$ between the CO axis and the $CR_2$ plane.

The active space is the same for all the states under investigation, consisting of the involved orbitals: $n_y$ oxygen lone-pair, $\sigma$, $\pi$ and $\pi^*$ CO orbitals, and $\sigma^*$ orbital to keep into account the expected elongation of the carbonyl. This results in a six electrons, five orbitals active space, hereafter called CAS(A). The term "$\sigma$ CO orbital" refers to the highest occupied molecular orbital with $\sigma$ symmetry with respect to the CO axis: at the equilibrium geometry of the excited states this orbital can be seen as a mixing of the $n_z$ lone pair of the oxygen atom and of the bonding $\sigma_{CO}$ orbital of the GS (with a larger component on the latter).

In the case of formaldehyde and acetone, the $xz$ symmetry plane simplifies the definition of the active space, and preserves the nature of this space during the symmetry break required for the geometry optimization of the excited states. This is not true for acetaldehyde: in the evaluation of GS and $\pi \rightarrow \pi^*$, the $n_y$ orbital is replaced by $n_z$, since the latter brings a larger correlative contribution and is therefore preferred for inclusion during the orbital optimization procedure. The problem does not occur on evaluating the $n_y \rightarrow \pi^*$ excited state, where the intrinsic electronic nature imposes the presence of the $n_y$ orbital inside the active space.

As we saw in the previous chapter, a localized orbital approach is able to preserve the correct nature of the active space. With a delocalized approach this is not possible, and a different strategy is needed: to enlarge the active space including a higher number of orbitals until a consistent space is obtained. Adding a single doubly-occupied orbital turns out to add a $\sigma_{CC}$ to the active space, and not the $n_y$ orbital. A further increase finally includes the requested orbital in the active set and a consistent ten electrons in seven orbitals space (named CAS(B)) is obtained.

Finally, for comparison against the results by Dallos et al.[97] an enriched active space has been defined for formaldehyde, containing all the valence orbitals and electrons $[(3-7)a_1$, $(1-2)b_1$ and $(1-3)b_2]$ and defining a space made of 12 electrons in ten orbitals (named CAS(C)).

As pointed out in Refs. 98 and 99, for a given choice of the active space multiple different final solutions exist. The difference resides in the nature of the orbitals included inside the active space, and it hardly affects the equilibrium geometries obtained after a geometry optimization. In order to produce a consistent behavior, during the optimization of the ground and excited states a proper nature of the active space has been kept. It is important to note that this choice does not always match the minimum energy criterion.

Both Strongly and Partially contracted NEVPT have been applied for the evaluation of the excitation energies, involving all electrons and orbitals in the perturbative treatment. Vertical excitation energies have been evaluated computing the electronic energy of all states at the CASSCF equilibrium geometry for the ground state. For adiabatic excitation energies, the single state CASSCF equilibrium geometry for the appropriate excited state has been considered.

The vibrational frequencies have been computed analytically using the harmonic approximation and are here reported without any scaling factor. The normal mode assignment has been done looking at the dominant contributions of the atom displacements. From the knowledge of the harmonic frequencies, the Zero Point Energy (ZPE) has been obtained at CASSCF level, and kept into ac-

count in this form for both CASSCF and NEVPT evaluations. This introduces the approximation that NEVPT acts on the states with a pure translation of the CASSCF energy surface, not altering the shape (which in turn acts on the Zero Point Energy).

## Results and discussion

### Ground state

CASSCF optimized geometries for the ground state are reported in Tab. 3.2.1.

| Method | $R_{CO}$ | $R_{CH}$[a] | $R_{CC}$ | ∠XCY[b] | ∠OCC |
|---|---|---|---|---|---|
| | | Formaldehyde | | | |
| CASSCF[CAS(A)][c] | 1.220 | 1.090 | | 117.3 | |
| CASSCF[CAS(C)][c] | 1.210 | 1.118 | | 115.1 | |
| Exp.[100] | 1.208 | 1.116 | | 116.5 | |
| | | Acetaldehyde | | | |
| CASSCF[CAS(B)][c] | 1.222 | 1.092 | 1.501 | 116.3 | 124.2 |
| Exp.[101] | 1.213 | 1.106 | 1.504 | 114.9 | 124.0 |
| | | Acetone | | | |
| CASSCF[CAS(B)][c] | 1.224 | | 1.509 | 117.2 | 121.4 |
| Exp.[102] | 1.222 | | 1.507 | 117.1 | 121.4 |

**Table 3.2.1:** Formaldehyde, acetaldehyde and acetone ground state equilibrium geometries. Distances in Å, angles in degrees.

[a] In acetaldehyde, H is the aldehydic hydrogen.

[b] Formaldehyde X=H, Y=H; acetaldehyde X=H, Y=C; acetone X=C, Y=C.

[c] This work.

An overall agreement with the experimental data can be appreciated. For formaldehyde, the CAS(A) bond distances show a discrepancy with respect to the experiment of 0.012 ($R_{CO}$) and 0.026 ($R_{CH}$). The error for the HCH angle is of 0.8°. The CAS(C) produces values in even better agreement, with the exception of the HCH angle value which is higher (1.4°) than with CAS(A) but remains small.

In the case of acetaldehyde and acetone, it is possible to note that the carbonyl group maintains the same structure found in formaldehyde, and both CASSCF values agree with the experimental data.

### General consideration on the excited states

An extended review of the experimental works on excited states of formaldehyde has been presented by Moule and Walsh[103] and Clouthier and Ramsay.[104] Many theoretical works have been dedicated to formaldehyde, starting from the works

of Whitten and Hackmeyer[105] and of Buenker and Peyerimhoff.[106] In 1982 a review of these pioneering theoretical works was published.[107] More recently the EOM-CCSD,[108,109] EOM-MBPT,[109] TDDFT/B3LYP,[110] TDDFT/PBE0,[111] OO-CCD,[112] CASPT2,[93] MRD-CI,[94,113] MRCISD+Q,[114] MRAQCC/LRT,[114] MRPT2,[115] MR-CI,[116] $(SC)^2$-MR-SDCI,[117] CCSD,[117] and NEVPT2[118] have been applied to formaldehyde, mainly focusing on vertical transitions.

Acetone has been less studied from the theoretical point of view. Most of the existent investigations concentrated on the $^1(n\rightarrow\pi^*)$ state,[119–123] or on the vertical transition energies.[108,110,118,124,125] The $S_1$ and $T_1$ surfaces have been also investigated[83,126–128] because of the already cited Norrish type-I reaction, which takes place on these surfaces. In this reaction, after the excitation to the $S_1$ state, the molecule dissociates leading to the cleavage of one of the $\alpha$-CC bonds. For the $^1(\pi \rightarrow \pi^*)$ state, only one determination of the equilibrium geometry exists.[128] Concerning the $\sigma \rightarrow \pi^*$, no studies have been performed.

Finally, for acetaldehyde the theoretical studies of the valence states are even more scarce than for acetone. From the theoretical point of view, CIS-MP2,[129] UMP2,[130,131] CIS(D),[132] EOM-CCSD,[108] P-EOM-MBPT,[109] TDDFT[110,133] studies have been recently published, but only in few cases the geometrical structure and the frequencies of the excited states have been investigated.

Despite the large number of theoretical works on these three molecules, various valence states have never been characterized and even for the states considered a study of all the molecules based on a common strategy appears to be missing. The use of different theoretical approaches, for which different approximation levels have been employed, makes it difficult to compare the data reported in the literature.

### The vertical $n\rightarrow\pi^*$ states

The calculation of the $n\rightarrow\pi^*$ singlet and triplet vertical energies are presented in Tabs. 3.2.2, 3.2.3 and 3.2.4 for formaldehyde, acetaldehyde and acetone, respectively. All the NEVPT2 vertical excitation energies agree with the previously published high-level theoretical results.

In the case of the singlet state of formaldehyde a fairly large number of different theoretical works has been published, and experimental values exist. For this state the NEVPT2 results (4.00-4.08 eV) are very close to the best available theoretical data, which are in the range 3.90-4.10 eV and with one of the experimental values (4.07 eV).

The valence triplet state has been less studied computationally, while experimental information exists. The NEVPT2 computed values for this state agree well with the available data, the only relevant differences being with TDDFT/B3LYP[110]

and RMS MRPT2/IVO[115] which, however, show too low and too high values, respectively. The tendency of TDDFT to give a too low transition energy for the $^3(n\rightarrow\pi^*)$ state has been found also in a TDDFT/PBE0 study[111] (3.18 and 3.20 eV with two basis sets).

| Method | $^1(n\rightarrow\pi^*)$ | $^3(n\rightarrow\pi^*)$ |
|---|---|---|
| CASSCF (CAS(A))[a] | 4.37 | 3.99 |
| NEVPT2 SC (CAS(A))[a] | 4.02 | 3.59 |
| NEVPT2 PC (CAS(A))[a] | 4.00 | 3.56 |
| CASSCF (CAS(C))[a] | 4.51 | 4.25 |
| NEVPT2 SC (CAS(C))[a] | 4.08 | 3.69 |
| NEVPT2 PC (CAS(C))[a] | 4.06 | 3.66 |
| NEVPT2 SC[118] | 3.93 | |
| NEVPT2 PC[118] | 3.91 | |
| CASPT2[93] | 3.91 | 3.48 |
| MRD-CI[94] | 4.05 | |
| MRD-CI[113] | | 3.69 |
| EOM-CCSD[108] | 3.98 | |
| P-EOM-MBPT[109] | 4.31 | |
| P-EOM-CCSD[109] | 4.41 | |
| $(SC)^2$-MR-SDCI[117] | 4.04 | 3.59 |
| CCSD[117] | 4.02 | 3.58 |
| OO-CCD[112] | 3.91 | |
| TDDFT/B3LYP[110] | 3.93 | 3.25 |
| MRCISD+Q[114] | 4.07 | |
| MRAQCC/LRT[114] | 3.98 | |
| RMS MRPT2/IVO[115] | 4.25 | 3.79 |
| Experimental | 3.79[b], 4.07[c] | 3.50[b], 3.75[d] |

**Table 3.2.2:** Vertical transition energies (eV) for the $n\rightarrow\pi^*$ singlet and triplet excited states of formaldehyde. Two active spaces considered: CAS(A) is defined by six active electrons and five active orbitals, while CAS(C) has twelve active electrons and ten active orbitals (see text).

[a] This work.

[b] Electron-impact data from Ref. 90.

[c] Ref. 87.

[d] Trapped electron spectrum, Ref. 134.

The use of a larger active space (CAS(C)) does not introduce strong variations at CASSCF level for both states: the $^1(n\rightarrow\pi^*)$ state remains almost unchanged, while the $^3(n\rightarrow\pi^*)$ state shows a moderate variation. This indicates that the electronic distribution for the latter states induces a small modification (polarization) of the $n_z$ and $\sigma_{\mathrm{CH}}$ orbitals which is taken into account if these orbitals are added to the active space together with the $\sigma^*_{\mathrm{CH}}$ orbitals.

| Method | $^1(n\rightarrow\pi^*)$ | $^3(n\rightarrow\pi^*)$ |
|---|---|---|
| CASSCF (CAS(B))[a] | 4.59 | 4.21 |
| NEVPT2 SC (CAS(B))[a] | 4.31 | 3.99 |
| NEVPT2 PC (CAS(B))[a] | 4.29 | 3.97 |
| TDDFT/BLYP[133] | 3.83 | |
| CISMP2[129] | 5.27 | 4.93 |
| CIS(D)[132] | 4.28 | |
| CCSD[132] | 4.26 | |
| VOO-CCD[112] | 3.46 | |
| P-EOM-MBPT[109] | 4.65 | |
| P-EOM-CCSD[109] | 4.71 | |
| EOM-CCSD[108] | 4.32 | |
| TDDFT/B3LYP[110] | 4.49 | 3.67 |
| Experimental | 4.28[b], 4.27[c] | 3.97[b], 3.91[d], 3.75[e] |

**Table 3.2.3:** Vertical transition energies (eV) for the $n\rightarrow\pi^*$ singlet and triplet excited states of acetaldehyde. CAS(B) is defined by ten active electrons and seven active orbitals.

[a] This work.
[b] Ref. 87.
[c] Electron-impact spectroscopy, Ref. 90.
[d] Ref. 135.
[e] Trapped electron spectrum, Ref. 134.

For acetaldehyde (Tab. 3.2.3) the agreement of the NEVPT2 values with experimental data for the $^1(n\rightarrow\pi^*)$ transition is excellent. For this transition a number of theoretical determinations exists: apart from the CIS(D),[132] CCSD[132] and EOM-CCSD[108] methods, which give results of the same quality as NEVPT2, the other methods give too low (TDDFT/BLYP[133] and VOO-CCD[112]) and too high (CISMP2,[129] TDDFT/B3LYP,[110] P-EOM-MBPT[109] and P-EOM-CCSD[109]) values. For the triplet $n\rightarrow\pi^*$ state also, the NEVPT2 results are in good agreement with the experimental data, apart from the value obtained in Ref. 134 (3.75 eV) which is quite far from the other determinations. In this case only two theoretical estimations are available: the CISMP2[129] value, as occurs for other states, is too large, while the TDDFT/B3LYP[110] is slightly too low.

For the acetone molecule (Tab. 3.2.4) the $n\rightarrow\pi^*$ singlet NEVPT2 results (both SC and PC) agree well with the experimental data and, with very few exceptions, with other theoretical results. For the $^3(n\rightarrow\pi^*)$ state only three theoretical estimations have been published and all of them give an energy slightly lower (3.86-3.98 eV) than the experimental values (4.15-4.3), while NEVPT2 (4.11 and 4.14 eV) shows a better agreement with experiments.

As to the comparison of the three molecules, one notes that the experimental study on acetone shows an increment, with respect to formaldehyde, of the excitation energy of the $^1(n\rightarrow\pi^*)$ and $^3(n\rightarrow\pi^*)$ states of $\simeq$ 0.4-0.5 eV, acetaldehyde

being between formaldehyde and acetone: these results are well reproduced by NEVPT2 and probably indicate a destabilization of the $\pi^*$ orbital due to the presence of the methyl group.

| Method | $^1(n\rightarrow\pi^*)$ | $^3(n\rightarrow\pi^*)$ |
|---|---|---|
| CASSCF (CAS(A))[a] | 4.86 | 4.54 |
| NEVPT2 SC (CAS(A))[a] | 4.49 | 4.14 |
| NEVPT2 PC (CAS(A))[a] | 4.47 | 4.11 |
| NEVPT2 SC[118] | 4.42 | |
| NEVPT2 PC[118] | 4.22 | |
| CASPT2[124] | 4.18 | 3.90 |
| EOM-CCSD[b] | 4.47 | |
| TDDFT[c] | 4.46 | |
| MRCI[122] | 4.44 | |
| TDDFT[d] | 4.40 | |
| CAS/MP2[e] | 4.04 | |
| MR-MP[f] | 4.27 | 3.98 |
| TDDFT[136] | 4.29 | |
| TDDFT(B3LYP)[110] | 4.44 | 3.86 |
| EOM-CCSD[108] | 4.48 | |
| TDDFT/BLYP[133] | 3.93 | |
| Experimental[g] | 4.5 | 4.3 |
| Experimental | 4.38[h],4.43[i] | 4.16[j],4.15[k] |

**Table 3.2.4:** Vertical transition energies (eV) for the $n\rightarrow\pi^*$ singlet and triplet excited states of acetone. CAS(A) is defined by six active electrons and five active orbitals.

[a] This work.

[b] Computed using the 6-311(2+,2+)G** basis set, Ref. 125. The reported values correspond to the $1^1A_2$, $4^1A_1$ and $2^1B_1$ states.

[c] cc-pVDZ+aug(CO) basis set with the PBE0 functional, Ref. 137.

[d] 6-31+G(d) basis set with the B3LYP functional, Ref. 128.

[e] State-average CASSCF plus MR MP2 calculation using the 6-311+G(d,p) basis set, Ref. 128.

[f] CASSCF with 8 electrons in 7 orbitals using a C[$3s2p1d$]/H[$2s1p$] basis set, Ref. 138.

[g] Electron-energy-loss spectroscopy in condensed phase at 18 K, Ref. 139.

[h] Electron-impact data from Ref. 90.

[i] Ref. 140.

[j] Electron-impact data from Refs. 90 and 141.

[k] Trapped electron spectrum, Ref. 134.

### Excited state equilibrium geometries

For the calculation of the adiabatic excitation energies, the equilibrium geometries for the $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ states have been computed at the CASSCF level.

### The $n \rightarrow \pi^*$ states

The equilibrium geometries for the $n \rightarrow \pi^*$ triplet and singlet states are reported in Tabs. 3.2.5, 3.2.6 and 3.2.7 for the formaldehyde, acetaldehyde and acetone respectively.

| Method | $R_{CO}$ | $R_{CH}$ | $\angle$HCH | $\theta^a$ |
|---|---|---|---|---|
| $^1(n \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.387 | 1.078 | 119.9 | 37.5 |
| CASSCF(CAS(C))[b] | 1.358 | 1.109 | 116.8 | 37.7 |
| CASSCF(6e5o)/6-311++G**[142] | 1.382 | 1.076 | 120.3 | 36.8 |
| CASSCF(10e/9o)[143] | 1.364 | 1.107 | 117.5 | 37.2 |
| MRD-CI[95] | 1.335 | 1.116 | 120.2 | 34.5 |
| MR-CISD[97] | 1.338 | 1.080 | 119.5 | 34.3 |
| MR-AQCC[97] | 1.325 | 1.090 | 117.4 | 34.9 |
| P-EOM-MBPT2[121] | 1.340 | 1.093 | | |
| EOM-CCSD[144] | 1.324 | 1.096 | | 30.3 |
| EOM-CCSD[125] | 1.311 | 1.095 | 118.8 | 32.5 |
| CASPT2[93] | 1.352 | 1.092 | 120.6 | 30.8 |
| EOM-CCSD[145] | 1.321 | 1.105 | 116.6 | 28.0 |
| DFT/BLYP[133] | 1.329 | 1.102 | 115 | 38 |
| Experimental[146] | 1.323 | 1.098 | 118.8 | 34.0 |
| Experimental[147] | 1.323 | 1.103 | 118.1 | 34.0 |
| $^3(n \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.365 | 1.079 | 119.3 | 39.3 |
| CASSCF(CAS(C))[b] | 1.335 | 1.112 | 113.9 | 43.7 |
| CASSCF(6e5o)/6-311++G**[142] | 1.360 | 1.077 | 119.7 | 38.5 |
| CCSD(T)[148] | 1.314 | 1.090 | 115.6 | 41.4 |
| MRD-CI[113] | 1.313 | 1.100 | 116.3 | 40.0 |
| MP2[113] | 1.318 | 1.092 | 117.3 | 40.8 |
| CIS[129] | 1.248 | 1.096 | 111.5 | 51 |
| CIS[149] | 1.256 | 1.092 | 112.8 | 43.1 |
| Experimental[c] | 1.307 | 1.084 | 121.8 | 41.1 |

**Table 3.2.5:** Equilibrium geometries for the $^1(n \rightarrow \pi^*)$ and $^3(n \rightarrow \pi^*)$ states of formaldehyde. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.

[b] This work.

[c] Refs. 104 and 150 as cited in Ref. 113.

The $^1(n \rightarrow \pi^*)$ state of formaldehyde is experimentally well characterized and many spectroscopical parameters are known.[147,151,152] As expected, the population of the $\pi^*$ antibonding orbital, weakening the CO bond, induces a lengthening

of the C-O distance and a pyramidalization on the carbon atom. The equilibrium geometry obtained with the CAS(A) and CAS(C) active spaces are in reasonable agreement with the experimental determinations: as for the bond lengths, $R_{CO}$ has an appreciable error (0.064 and 0.035 Å for the two spaces respectively) while $R_{CH}$ is well described with CAS(C) and slightly too short with CAS(A). This behavior for $R_{CO}$ was already found at CASSCF level in Refs. 142 and 143 and is common to all the excited states of formaldehyde here considered: the CASSCF(CAS(A)) description gives too large values, only partially improved with the CAS(C) active space. Dallos *et al*[97] found that increasing the basis set size, the $R_{CO}$ distance decreases and is therefore expected an improvement using a larger basis set. Also, one can expect that performing the geometry optimization including the dynamical correlation energy could further decrease the CO distance. The out–of–plane bending angle is found to be 37.5 and 37.7 degrees in the present work, 3.5 and 3.7 degrees higher than the experimental value.

For the $^3(n\rightarrow\pi^*)$ state similar considerations can be done. In this case there are fewer theoretical works in literature. The obtained results confirm that the $n\rightarrow\pi^*$ triplet state is more bent than the singlet[103,129] and has a shorter CO bond length.

| Method | $R_{CO}$ | $R_{CC}$ | $R_{CH_{ald}}$ | ∠OCC | ∠$CCH_{ald}$ | $\theta^a$ |
|---|---|---|---|---|---|---|
| | $^1(n\rightarrow\pi^*)$ excited state | | | | | |
| CASSCF(CAS(B))[b] | 1.383 | 1.501 | 1.080 | 114.5 | 120.3 | 38.4 |
| DFT/BLYP[133] | 1.334 | 1.524 | 1.102 | 118 | 117 | 35 |
| P–EOM–MBPT2[121] | 1.359 | 1.494 | 1.095 | 116.0 | | |
| CIS[129] | 1.256 | 1.531 | 1.089 | 118.3 | 117.0 | |
| Exp.[153] | | | | | | 35.7 |
| Exp.[130] | | | | | | 37.8 |
| Exp.[154] | 1.32 | | | | | 26 |
| | $^3(n\rightarrow\pi^*)$ excited state | | | | | |
| CASSCF(CAS(B))[b] | 1.360 | 1.506 | 1.082 | 114.8 | 119.1 | 40.2 |
| UMP2[131] | 1.323 | 1.509 | 1.094 | 114.5 | 127.6 | 39.7 |
| UMP2[129] | 1.338 | 1.505 | 1.094 | 114.2 | 119.1 | |
| UMP2[130] | | | | | | 39.7 |
| CIS[129] | 1.254 | 1.544 | 1.097 | 116.8 | 113.1 | ≃48 |
| Exp.[153] | | | | | | 42.2 |

**Table 3.2.6:** Equilibrium geometries for the $^1(n\rightarrow\pi^*)$ and $^3(n\rightarrow\pi^*)$ states of acetaldehyde. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.

[b] This work.

In the case of acetaldehyde, the state is much less characterized experimentally: only the out–of–plane bending angle has been determined.[130,153] The values here obtained compare well with these results. The theoretical knowledge of the

equilibrium structure of the two $n \rightarrow \pi^*$ states is less complete than for formaldehyde, being essentially based on single reference methods. The comparison with the analogous states in formaldehyde shows that the CO and $CH_{ald}$ bond lengths and the out-of-plane angle are very similar for the two molecules.

| Method | $R_{CO}$ | $R_{CC}$ | $\angle$CCC | $\theta^a$ |
|---|---|---|---|---|
| $^1(n \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.401 | 1.499 | 121.3 | 39.6 |
| CASSCF[c] | 1.355 | 1.521 | 117.3 | 43.1 |
| CASSCF[d] | 1.399 | 1.507 | | 42.3 |
| CASSCF[e] | 1.394 | 1.504 | 120.3 | 41.7 |
| CASSCF[f] | 1.394 | 1.504 | 120.7 | 41.2 |
| CASSCF[g] | 1.401 | 1.500 | 121.5 | 36.3 |
| P-EOM-MBPT2[121] | 1.376 | 1.502 | 122.3 | 35.6 |
| DFT/BLYP[133] | 1.341 | 1.524 | 119 | 36 |
| $^3(n \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.381 | 1.502 | 120.4 | 40.2 |
| CASSCF[d] | 1.376 | 1.512 | | 49.7 |
| CASSCF[e] | 1.370 | 1.510 | 119.4 | 41.8 |
| CASSCF[f] | 1.371 | 1.509 | 119.5 | 43.6 |

**Table 3.2.7:** Equilibrium geometries for the $^1(n \rightarrow \pi^*)$ and $^3(n \rightarrow \pi^*)$ states of acetone. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.
[b] This work.
[c] CASSCF 10 electrons in 11 orbitals using the 6-311G** basis set, Ref. 122.
[d] CASSCF 8e/7o using a C[$3s2p1d$]/H[$2s1p$] basis set, Ref. 138.
[e] CASSCF 8e/7o using the 6-31G* basis set, Ref. 127.
[f] CASSCF 10e/8o using the 6-31G(d) basis set, Ref. 83.
[g] CASSCF 6e/5o using the 6-31G(d,p) basis set, Ref. 119.

For acetone no experimental data exist but some CASSCF geometry optimizations for both the $n \rightarrow \pi^*$ triplet and singlet have been published. For the singlet also P-EOM-MBPT2[121] and DFT/BLYP[133] studies have been performed. The results here obtained agree with these calculations, the main difference being for the $^1(n \rightarrow \pi^*)$ state where the DFT/BLYP[133] and the CASSCF (Ref. 122) results give a shorter CO bond length, a longer CC bond length and a smaller $\angle$CCC angle.

Concerning the comparison with the other molecules, one can note that all the geometrical parameters show modest variations changing the carbonyl substituents (considering CAS(A) for formaldehyde). The only exception is the CO bond length in acetone which is slightly larger than in formaldehyde and acetaldehyde for both states.

The present results show that the excitation process, apart from the lengthening of the CO bond and the pyramidalization, leads to a moderate increase of

the ∠RCR′ (R, R′=H, $CH_3$) and to a modest shortening of the $CH_{ald}$ bond length while the CC bond length does not change significantly.

### The $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ states

The CASSCF equilibrium geometry for the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ states of formaldehyde, acetaldehyde and acetone are reported in Tabs. 3.2.8, 3.2.9 and 3.2.10, respectively.

| Method | $R_{CO}$ | $R_{CH}$ | ∠HCH | $\theta$[a] |
|---|---|---|---|---|
| $S_2$ excited state | | | | |
| CASSCF(CAS(A))[b] (I) | 1.564 | 1.082 | 116.2 | 48.3 |
| CASSCF(CAS(C))[b] (I) | 1.535 | 1.112 | 113.2 | 44.6 |
| CASSCF(CAS(A))[b] (II) | 1.356 | 1.140 | 91.0 | 76.5 |
| CASSCF(CAS(C))[b] (II) | 1.257 | 1.264 | 55.7 | 66.8 |
| MRD-CI[155] | 1.528 | 1.080 | 111.3 | ∼44 |
| EOM-CCSD[125] | 1.583 | 1.095 | 119.6 | 0.0 |
| CASPT2[93] | 1.492 | 1.094 | 118.4 | 46.2 |
| CIS[149] | 1.460 | 1.073 | 124.4 | 0.0 |
| MR-CISD[97] | 1.505 | 1.082 | 117.5 | 45.8 |
| MR-AQCC[97] | 1.495 | 1.088 | 118.9 | 45.4 |
| $^3(\pi \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.474 | 1.077 | 120.0 | 35.2 |
| CASSCF(CAS(C))[b] | 1.465 | 1.102 | 119.0 | 32.2 |
| MRD-CI[113] | 1.476 | 1.075 | 119.9 | 38.1 |
| MP2[113] | 1.451 | 1.082 | 121.3 | 28.9 |
| CIS[149] | 1.408 | 1.072 | 119.2 | 39.4 |
| Experimental[c] | 1.423 | | | |
| $^1(\sigma \rightarrow \pi^*)$ excited state | | | | |
| CIS[129] | 1.489 | 1.079 | 121.9 | 46 |
| MRD-CI[95] | 1.529 | 1.080 | 115.1 | 42.1 |
| $^3(\sigma \rightarrow \pi^*)$ excited state | | | | |
| CASSCF(CAS(A))[b] | 1.499 | 1.073 | 129.4 | 46.9 |
| CASSCF(CAS(C))[b] | 1.483 | 1.101 | 128.7 | 45.2 |
| MRD-CI[113] | 1.455 | 1.090 | 133.5 | 35.5 |

**Table 3.2.8:** Equilibrium geometries for the $S_2$ (the $^1(\pi \rightarrow \pi^*)$ and $^1(\sigma \rightarrow \pi^*)$ configurations are mixed), $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states of formaldehyde. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.

[b] This work. I and II indicate two different minima found on the $S_2$ energy surface.

[c] Electron-impact spectroscopy, Ref. 156 as cited in Ref. 113.

The equilibrium geometry of the $^1(\pi \rightarrow \pi^*)$ state of formaldehyde has been subject of debate. The Walsh rules[157] indicate that all the $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ states should have bent equilibrium geometry, as actually has been experimentally and theoretically demonstrated for the $n \rightarrow \pi^*$ state (see Tab. 3.2.5). For the $\pi \rightarrow \pi^*$ state no experimental data exist regarding the equilibrium ge-

ometry. Excited state SCF (SCFX),[158] two configuration SCF (TCSCF),[158] CI of singles (CIS),[129,149] multireference singles and doubles CI (MRD-CI)[94,159] and CASPT2[93] calculations have found that the equilibrium geometry for the $^1(\pi \rightarrow \pi^*)$ state is planar. In a recent work[97] Dallos *et al*, performing full geometry optimization using high-level MR methods, showed that this state has a pyramidal equilibrium geometry and that the structure optimized under planar constraints is a saddle point, rather than a minimum, therefore confirming the predictions of the Walsh rules. Dallos *et al* also found that the $^1(\pi \rightarrow \pi^*)$ and $^1(\sigma \rightarrow \pi^*)$ energy surfaces exhibit a conical intersection and that the $^1(\pi \rightarrow \pi^*)$ state at its equilibrium geometry has a strong component on the $n^2$ GS electronic configuration (this is the origin of the predissociated character of all VUV absorption bands), and a large component of the $^1(\sigma \rightarrow \pi^*)$ state. A pyramidal minimum in the 2 $^1A'$ energy surface with a mixed $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ character was also found by other authors.[93,94]

In the present calculations, two minima have been found on the $S_2$ potential energy surface of formaldehyde: they are reported in Tab. 3.2.8 and indicated with the notation I and II. In other cases, multiple solutions have been found which differ only for the nature of the active orbitals with occupation number close to two or zero and describe therefore the same geometrical minimum, giving almost the same equilibrium geometry. On the contrary, in this case the two solutions are related to different geometrical structures.

Structure I confirms the results of Dallos *et al*: the molecule has a pyramidal structure with an out-of-plane bending angle of 45-48 degrees and the wavefunction has a $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ mixed character. This mixing makes the label here used, $\pi \rightarrow \pi^*$, less meaningful than for the other states and this state would be better termed $S_2$. The CASSCF $R_{CO}$ bond distance for the I minimum with both the CAS(A) and CAS(C) active spaces is slightly too long with respect to the MR-CISD[97] ones, while the $R_{CH}$ is slightly too large with CAS(C) and is well described with CAS(A). The $\angle$HCH angle is always lower than in the MR-CISD calculations.

The second minimum (II) has a peculiar structure: the $R_{CO}$ distance is shorter than for the I structure (with CAS(C) it is close to the value found for the GS), the $R_{CH}$ bond length is the longest found for all the states here studied and the $\angle$HCH angle is the smallest. The molecule is strongly bent with an out-of-plane angle of 76.5 degrees (CAS(A)) and 66.8 degrees (CAS(C)). This structure, which has never been characterized, can be interesting because its energy is close to the energy of the I structure, with CAS(C) giving a lower energy. The II geometry can be a good candidate as a key structure in photochemical dynamical processes but it can be reasonably excluded as a relevant structure in the interpretation

of the absorption spectrum. Its actual importance can be established only by a complete scanning of the GS, $S_1$ and $S_2$ potential energy surfaces, a study which is beyond the scope of this thesis. Given its probably marginal interest in the description of the absorption spectrum, it will not be further discussed here.

The triplet $\pi \rightarrow \pi^*$ state has also a pyramidal equilibrium geometry with a shorter $R_{CO}$ bond length and a less bent structure than the singlet state. The geometrical parameters agree with previous published results, apart from the already discussed characteristic of the $R_{CO}$ bond distance which turns out to be slightly too long with respect to the experimental value. In this case no mixing is noted with the $\sigma \rightarrow \pi^*$ configuration and the electronic configuration $\pi \rightarrow \pi^*$ provides a good representation.

The $\sigma \rightarrow \pi^*$ singlet and triplet states of formaldehyde have received much less attention than the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ states. In this work no minimum has been found on the $S_3$ energy surface, thus confirming the results of Dallos *et al.*[97] Also in this case, the triplet is exempt from the $\pi \rightarrow \pi^*/\sigma \rightarrow \pi^*$ mixing. It has a pyramidal structure with a larger out-of-plane bending angle than the $^3(\pi \rightarrow \pi^*)$ state.

| Method | $R_{CO}$ | $R_{CC}$ | $R_{CH_{ald}}$ | $\angle OCC$ | $\angle CCH_{ald}$ | $\theta$[a] |
|---|---|---|---|---|---|---|
| | $S_2$ excited state | | | | | |
| CASSCF(CAS(B))[b] | 1.624 | 1.448 | 1.083 | 110.7 | 118.2 | 32.8 |
| CIS[129] | 1.486 | 1.472 | 1.079 | 112.8 | 121.8 | |
| | $^3(\pi \rightarrow \pi^*)$ excited state | | | | | |
| CASSCF(CAS(B))[b] | 1.478 | 1.489 | 1.078 | 114.3 | 120.4 | 37.3 |
| | $^3(\sigma \rightarrow \pi^*)$ excited state | | | | | |
| CASSCF(CAS(B))[b] | 1.523 | 1.482 | 1.074 | 110.9 | 129.9 | 45.5 |

**Table 3.2.9:** Equilibrium geometries for the $S_2$ (the $^1(\pi \rightarrow \pi^*)$ and $^1(\sigma \rightarrow \pi^*)$ configurations are mixed), $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states of acetaldehyde. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.

[b] This work.

In the case of acetaldehyde and acetone, for the $\pi \rightarrow \pi^*$ states only two theoretical determinations of the equilibrium geometry have been published: the first regards the $^1(\pi \rightarrow \pi^*)$ state of acetaldehyde[129] while the second regards the $^1(\pi \rightarrow \pi^*)$ state of acetone.[128] The $^3(\pi \rightarrow \pi^*)$ and the $\sigma \rightarrow \pi^*$ states have not been studied. No experimental data exist in these cases.

The results here obtained show that, as in formaldehyde, the second singlet excited state has a mixed $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ nature in both molecules and it is called hereafter $S_2$. In acetaldehyde the CASSCF approach gives for this state a geometrical structure quite similar to the CIS one,[129] apart from the CO bond length for which CIS is known to give too short values.

| Method | $R_{CO}$ | $R_{CC}$ | ∠CCC | $\theta^a$ |
|---|---|---|---|---|
| | $S_2$ excited state | | | |
| CASSCF(CAS(A))[b] | 1.648 | 1.453 | 123.8 | 44.6 |
| CASSCF[c] | 1.642 | 1.453 | 123.7 | 45.4 |
| | $^3(\pi \rightarrow \pi^*)$ excited state | | | |
| CASSCF(CAS(A))[b] | 1.483 | 1.493 | 121.1 | 40.4 |
| | $^3(\sigma \rightarrow \pi^*)$ excited state | | | |
| CASSCF(CAS(A))[b] | 1.445 | 1.494 | 120.6 | 40.1 |

**Table 3.2.10:** Equilibrium geometries for the $S_2$ (the $^1(\pi \rightarrow \pi^*)$ and $^1(\sigma \rightarrow \pi^*)$ configurations are mixed), $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states of acetone. Distances in Å, angles in degrees.

[a] Out-of-plane bending angle.

[b] This work.

[c] CASSCF with six electrons in six orbitals using the 6-311+G(d,p) basis set, Ref. 128.

The equilibrium geometry here obtained for acetone is very close to the one published in Ref. 128 and confirms that the molecule is bent with an out-of-plane angle similar to the one found for formaldehyde. The CO bond length is remarkably longer and the ∠CCC angle is larger than in formaldehyde. In both triplets no mixing has been found between the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ electronic configurations. The structure of the $^3(\sigma \rightarrow \pi^*)$ state for acetaldehyde and acetone is here reported for the first time.

As for the $S_3$ state, our calculations confirm the results on acetone of Diau *et al*,[128] who have found that no stationary point exists on the $S_3$ surface, any attempt to find it always leads to the $S_2$ minimum.

The comparison of the three molecules reveals that the the $^3(\pi \rightarrow \pi^*)$ state has a geometrical structure similar for all molecules, only the out–of–plane bending angle showing a moderate variation. This behavior was also found for the $n \rightarrow \pi^*$ states. On the contrary, the $S_2$ and $^3(\sigma \rightarrow \pi^*)$ states present different structures when changing the carbonyl substituents. In the case of the $S_2$ state the CO bond length and the ∠RCR′ (R, R′=H, $CH_3$) angle increase when the hydrogen atoms are substituted with the methyl group, while the CC and $CH_{ald}$ bond lengths remain almost unchanged. The pyramidalization angle does not have a clear trend, being large for formaldehyde, small for acetaldehyde and intermediate for acetone. For the $^3(\sigma \rightarrow \pi^*)$ state the behavior is less clear: one can again note that the CC and $CH_{ald}$ bond lengths show small variations and that the ∠RCR′ and pyramidalization angles are very close for formaldehyde and acetaldehyde.

In all cases the excitation process leads to a CO bond lengthening larger than for the $n \rightarrow \pi^*$ states: this can be easily understood noting that in the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ states the electron is promoted from a bonding to an antibonding orbital.

### Adiabatic excitation energies

The adiabatic excitation energies for the states here considered are reported in Tabs. 3.2.11, 3.2.12 and 3.2.13. In order to account for the vibrational contribution of the $0 \leftarrow 0$ adiabatic transition, the Zero Point Energy has been estimated in the harmonic approximation at the CASSCF level, computing the harmonic vibrational frequencies of all the states here considered. Corrected values are in parentheses.

| Method | $^1(n\rightarrow\pi^*)$ | $^3(n\rightarrow\pi^*)$ | $S_2$ | $^3(\pi\rightarrow\pi^*)$ | $^3(\sigma\rightarrow\pi^*)$ |
|---|---|---|---|---|---|
| CASSCF (CAS(A)) [a] | 3.63 (3.56) | 3.36 (3.30) | 7.90 (7.82) | 4.36 (4.29) | 7.77 (7.71) |
| NEVPT2 SC (CAS(A)) [a] | 3.52 (3.46) | 3.17 (3.10) | 7.67 (7.59) | 4.38 (4.31) | 7.28 (7.22) |
| NEVPT2 PC (CAS(A)) [a] | 3.54 (3.47) | 3.17 (3.10) | 7.66 (7.58) | 4.41 (4.34) | 7.28 (7.21) |
| CASSCF (CAS(C)) [a] | 3.92 (3.82) | 3.64 (3.54) | 8.25 (8.15) | 4.75 (4.67) | 7.89 (7.79) |
| NEVPT2 SC (CAS(C)) [a] | 3.57 (3.48) | 3.21 (3.11) | 7.65 (7.55) | 4.48 (4.40) | 7.20 (7.10) |
| NEVPT2 PC (CAS(C)) [a] | 3.57 (3.48) | 3.18 (3.08) | 7.62 (7.52) | 4.49 (4.42) | 7.17 (7.07) |
| CASSCF(10e/9o)[143] | 3.71 (3.80) | | | | |
| CIS[129] | 4.54 | 3.63 | 8.65 | 3.64 | |
| MRD-CI[113] | | 3.22 | | 4.43 | 7.15 |
| MR-CISD+Q[116] | 3.42 | | | | |
| CASPT2[93] | 3.37 | | 7.43 | | |
| MRD-CI[94] | 3.64 | | 7.95 | | |
| EOM-CCSD[125] | 3.70 | | 8.43 | | |
| CCSD(T)[b] | | 3.17 (3.08) | | | |
| MR-AQCC[c] | 3.60 | | 7.45 | | |
| MR-CISD+Q[c] | 3.63 | | 7.56 | | |
| Experimental | 3.50[d] | 3.12[d] | | 4.70[e],4.83[f] | |

**Table 3.2.11:** Adiabatic transition energies (eV) (all states computed at the CASSCF equilibrium geometry) for the $n\rightarrow\pi^*$, $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ singlet and triplet excited states of formaldehyde. The values in parentheses are the ZPE corrected values.

[a] This work.

[b] CCSD(T) value using the TZ2P(f,d) basis, Ref. 148.

[c] Extrapolated to basis set limit using cc-pVXZ basis sets (X=D, T and Q), indicated with TQ in Ref. 97.

[d] Optical spectroscopy results from Ref. 104, 150.

[e] Estimated in Ref. 113 from a polynomial fit of the electron-impact data for the $\nu_2$ progression in Ref. 156.

[f] Lowest observed peak in the $\nu_2$ progression assigned to $\nu' = 1$ in Ref. 156.

As a general consideration, it is possible to note that the inclusion of the correlation energy *via* the NEVPT2 approach leads in general to a lowering of the excitation energy. This lowering is of the order of some tenths of eV, not exceeding 0.6-0.7 eV apart from the particular case of the $S_2$ state of formaldehyde with CAS(A) ($\simeq$ 2 eV, reduced to $\simeq$ 0.6 eV with CAS(C)). This shows that the zero-order description appears to be qualitatively correct, but correlation effects must be taken into account for a quantitative estimation of the excitation energies.

Moreover, in all the cases here studied both Strongly and Partially Contracted NEVPT2 variants give almost the same results, the difference being always of the order of some hundredths of eV. The agreement between the SC and PC variants of NEVPT2 is another indication of the good quality of the zero-order wavefunctions.

For formaldehyde the $^1(n \rightarrow \pi^*)$ NEVPT2 values range in the narrow interval 3.52-3.57 eV (3.46-3.48 with ZPE) in very good agreement with the experimental determination (3.50 eV) and previously published high quality theoretical estimations. Analogous considerations can be done for the $^3(n \rightarrow \pi^*)$ state, where the NEVPT2 values (3.17-3.21 eV, 3.08-3.11 eV with ZPE) agree with the experimental value (3.12 eV) and with the MRD-CI[113] and CCSD(T)[148] results (3.22 and 3.17 eV, respectively), while the CIS[129] determination (3.63 eV) is too high.

For the $S_2$ state, at the I minimum (the one involved in the excitation process) NEVPT2 gives an adiabatic excitation energy in the range 7.62-7.67 eV (7.52-7.59 eV with ZPE), only slightly higher than the high-level MR-AQCC,[97] MR-CISD+Q[97] and CASPT2 values[93] (7.45, 7.56 and 7.43 eV respectively). For this state no experimental measurements have been published. The value here found for the excitation energy of the $S_2$ adiabatic transition, close to the adiabatic $^1(n \rightarrow 3s)$ transition[155] (7.08 eV), confirms its role as a perturber state for this transition.

The $^3(\pi \rightarrow \pi^*)$ state is interesting because experimental determinations of the adiabatic excitation energy exist. For this state considerations have been published[156] regarding the hypothesis that the lowest observed peak (4.83 eV) could not be assigned to the $0 \leftarrow 0$ vibrational transition, but rather to the $1 \leftarrow 0$ one (the vibrational levels regard the $\nu_2$ CO stretch normal coordinate). On the basis of computed Franck-Condon factors, Bruna *et al*[113] showed that the $0 \leftarrow 0$ and $1 \leftarrow 0$ transitions are predicted to be so weak that probably they have not been detected experimentally and propose to assign the 4.83 eV observed peak to the $2 \leftarrow 0$ transition. Using their predicted difference between the $\nu_2 = 2$ and $\nu_2 = 0$ CO stretch levels (0.27 eV), the $0 \leftarrow 0$ transition can be estimated at 4.56 eV. The NEVPT2 results in this case range in a slightly wider interval (4.38-4.49 eV, 4.31-4.42 eV with ZPE) than previously and confirm that the observed peak cannot be assigned to the $0 \leftarrow 0$ transition. Given the high level of accuracy found for the NEVPT2 results for the $^1(n \rightarrow \pi^*)$ and $^3(n \rightarrow \pi^*)$ states, one can suppose the 4.56 eV estimation of Bruna *et al*[113] to be too high and that the 4.83 eV observed peak could be assigned to the $\nu_2$ $3 \leftarrow 0$ transition. With this hypothesis and the estimation of 0.39 eV for the difference between the $\nu_2 = 3$ and $\nu_2 = 0$ CO stretch levels,[113] the new assignment for the $0 \leftarrow 0$ transition is 4.44 eV.

The $^3(\sigma \rightarrow \pi^*)$ transition has not been observed experimentally. The NEVPT2 values (7.17-7.28 eV, 7.07-7.21 eV with ZPE) agree with the theoretical work[113]

concerning this state.

For acetaldehyde the theoretical knowledge of the valence excited states is modest. All the available studies are of single reference type and the NEVPT2 results of this work are the first obtained with a multireference approach. For the $n \rightarrow \pi^*$ states, where experimental data exist, the agreement of the NEVPT2 excitation energies with experiment is good and they show a sizable improvement with respect to the other theoretical estimations (the CIS-MP2[129] and UMP2[129] values are too high while the TDDFT[133] one is too low).

| Method | $^1(n \rightarrow \pi^*)$ | $^3(n \rightarrow \pi^*)$ | $S_2$[a] | $^3(\pi \rightarrow \pi^*)$ | $^3(\sigma \rightarrow \pi^*)$ |
|---|---|---|---|---|---|
| CASSCF (CAS(B))[b] | 3.79 (3.73) | 3.52 (3.46) | 7.61 (7.50) | 4.49 (4.43) | 7.54 (7.46) |
| NEVPT2 SC (CAS(B))[b] | 3.67 (3.61) | 3.45 (3.39) | 7.16 (7.05) | 4.47 (4.41) | 7.06 (6.97) |
| NEVPT2 PC (CAS(B))[b] | 3.68 (3.62) | 3.45 (3.39) | 7.13 (7.02) | 4.49 (4.43) | 7.04 (6.96) |
| TDDFT/BLYP[133] | 3.35 | | | | |
| CIS-MP2[129] | 4.79 | 3.94 | 8.30 | | |
| UMP2[129] | | 3.75 | | | |
| Experimental[90] | 3.56 | | | 5.08 | |
| Experimental | 3.69[c,d], 3.58[e] | 3.29[c], 3.38[e] | | | |

**Table 3.2.12:** Adiabatic transition energies (eV) (all states computed at the CASSCF equilibrium geometry) for the $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ singlet and triplet excited states of acetaldehyde. The values in parentheses are the ZPE corrected values.

[a] The $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ electronic configurations are mixed. No minima have been found on the $S_3$ energy surface.

[b] This work.

[c] Ref. 160

[d] Ref. 161

[e] Ref. 154

For the $S_2$ state no experimental assignment and only one theoretical study have been published. The NEVPT2 values (7.13-7.16 eV, 7.02-7.05 with ZPE) are much lower than the CIS-MP2[129] one (8.30 eV). This again confirms the tendency of the CIS-MP2 method to give too high excitation energies. Finally for the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ triplets only an experimental measurement exists for the $^3(\pi \rightarrow \pi^*)$ state and those here presented are the first theoretical values. For this state the disagreement of NEVPT2 with experiment ($\simeq$ 0.6-0.7 eV) is slightly larger than in formaldehyde and could have the same origin, i.e., the observed peak should be assigned to a $n \leftarrow 0$ vibrational transition (with $n > 0$).

In the case of the acetone molecule, for the $n \rightarrow \pi^*$ singlet and triplet states the NEVPT2 results are in a better agreement with the experimental data than the other available theoretical estimations (apart from the MRCI[122]).

It must be noted that in all the calculations for the excited states of acetone, the geometry optimization has been performed in the $C_s$ symmetry group: this allows the molecular skeleton to pyramidalize, but imposes a constraint for the

rotation of the methyl groups. Nevertheless the optimized geometries have been confirmed to be true minima (possibly not the global minimum) by a vibrational analysis. One can however reasonably suppose that the possible global minima arising from a rotation of the methyl group should have energy properties very similar to those here reported.

| Method | $^1(n\rightarrow\pi^*)$ | $^3(n\rightarrow\pi^*)$ | $S_2$[a] | $^3(\pi\rightarrow\pi^*)$ | $^3(\sigma\rightarrow\pi^*)$ |
|---|---|---|---|---|---|
| CASSCF (CAS(A))[b] | 3.86 (3.81) | 3.66 (3.61) | 7.28 (7.17) | 4.47 (4.42) | 7.84 (7.77) |
| NEVPT2 SC (CAS(A))[b] | 3.74 (3.70) | 3.50 (3.45) | 6.63 (6.52) | 4.46 (4.41) | 7.25 (7.18) |
| NEVPT2 PC (CAS(A))[b] | 3.75 (3.70) | 3.49 (3.44) | 6.61 (6.51) | 4.48 (4.42) | 7.20 (7.14) |
| CASPT2[c] | 3.64 | | 7.19 | | |
| MRCI[122] | (3.73) | | | | |
| MRMP[138] | 3.54 (3.47) | 3.27 (3.21) | | | |
| CASPT2[127] | 4.40 (4.35) | 4.05 (4.01) | | | |
| CASSCF[d] | 4.15 (4.11) | 3.93 (3.89) | | | |
| TDDFT/B3P86[83] | 4.11 (4.07) | 3.69 (3.65) | | | |
| TDDFT/BLYP[133] | 3.41 | | | | |
| Experimental | 3.75[e], 3.77[f] | 3.47[h] | | 5.15[e], 5.3[g] | |
| Experimental | 3.8[g] | 3.44–3.54[i] | | | |

**Table 3.2.13:** Adiabatic transition energies (eV) (all states computed at the CASSCF equilibrium geometry) for the $n\rightarrow\pi^*$, $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ singlet and triplet excited states of acetone. The values in parentheses are the ZPE corrected values.

[a] The $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ electronic configurations are mixed. No minima have been found on the $S_3$ energy surface.

[b] This work.

[c] Theoretical values from Ref. 124. For $^1(n\rightarrow\pi^*)$ the value is an estimation based on the hypothesis that the lowering passing from the vertical to adiabatic transition is equal to the one found in Ref. 93 for formaldehyde. For the $^1(\pi\rightarrow\pi^*)$ state the molecular skeleton is maintained planar and the $CH_3$ groups have the GS geometry.

[d] CASSCF with 10 electrons in 8 orbitals using the 6-31G(d) basis set, Ref. 83.

[e] Electron-impact data from Ref. 90.

[f] Fluorescence excitation spectra in Ar supersonic nozzle beam, Ref. 162.

[g] Electron-energy-loss spectroscopy in condensed phase at 18 K, Ref. 139.

[h] Experimental determination of Refs. 163,164, as reported in Ref. 85.

[i] Experimental estimation of the 0-0 band, Ref. 165.

The $S_2$ state is found to be 6.61-6.63 eV above the ground state, in partial disagreement with the CASPT2 value[124] which however has been obtained performing a restricted optimization (the molecular skeleton is planar and the methyl groups have the GS geometry).

For the $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ triplets the results here reported are, to the best of our knowledge, the first theoretical estimations. The disagreement between the NEVPT2 and the experimental values for the $^3(\pi\rightarrow\pi^*)$ state (0.7-0.9 eV) could be explained, as previously described in the case of formaldehyde and acetaldehyde, supposing that the experimental value, assigned to the $0\leftarrow 0$ vibrational

transition, is actually due to a transition to higher vibrational levels in the excited state, the first vibrational levels having too low Franck-Condon factors. In the case of formaldehyde and acetaldehyde the difference between the experimental and the NEVPT2 values was smaller (0.4-0.5 and 0.6-0.7 eV). This point can be clarified only with a calculation of the Franck-Condon factors.

It is worth noting that if the NEVPT2 description is correct, an interesting feature distinguishes acetone from formaldehyde and acetaldehyde: in acetone the $^3(\pi \rightarrow \pi^*)$ state at its equilibrium geometry is very close in energy to the $^1(n \rightarrow \pi^*)$ state in the Franck-Condon region. This feature could play a role in the Norrish type-I reaction, which is known to proceed *via* an intersystem crossing[85] from the $^1(n \rightarrow \pi^*)$ state to the triplet manifold. In formaldehyde the minimum of $^3(\pi \rightarrow \pi^*)$ was predicted to be $\simeq$ 0.4 eV above the vertical excitation of the $^1(n \rightarrow \pi^*)$ state, thus excluding its participation in the photophysical processes taking place after the $^1(n \rightarrow \pi^*)$ excitation. In the case of acetaldehyde the difference is less pronounced than in formaldehyde, but the $^3(\pi \rightarrow \pi^*)$ minimum remains higher than the $^1(n \rightarrow \pi^*)$ at the vertical excitation geometry.

Finally, one notes that the inclusion of the ZPE in the calculation of the adiabatic excitation energies makes in general the computed values closer to the experimental data. The use of the harmonic approximation at the CASSCF level is found to be a reasonable approximation in the present applications.

### Harmonic vibrational frequencies

#### Ground state

The harmonic vibrational frequencies for the ground state are reported in Tabs. 3.2.14, 3.2.15 and 3.2.16 for the three molecules, respectively, together with the most meaningful published values.

For the ground state of the three molecules here considered the experimental fundamental and harmonic vibrational frequencies have been published and the assignment of the normal modes is known without ambiguity.

The comparison of the results here obtained with these data allows some general considerations to be drawn. It is well known that for HF calculations the computed harmonic frequencies are consistently too high and they must be scaled by a factor less than one for comparisons with experimental harmonic frequencies, due to the lack of electronic correlation (static and dynamic) of the HF wavefunction.

In the CASSCF wavefunction part of the static electronic correlation is in principle taken into account, so a closer agreement with experiment could be expected. Actually this is true for the normal modes involving bonds for which the valence bonding and antibonding orbitals are included in the active space. In the

calculations here reported, this is always true only for the CO stretching. In the particular case of formaldehyde with CAS(C), where all the valence bonding and antibonding orbitals are active, the above consideration holds true for all bonds. For instance the CH bonds in formaldehyde with CAS(A) are treated in a "single reference" (similar to HF) manner therefore neglecting the $\sigma/\sigma^*$ correlation.

| Method | asym CH str. ($b_2$) | sym CH str. ($a_1$) | CO str. ($a_1$) | $CH_2$ sciss. ($a_1$) | $CH_2$ rock ($b_2$) | out–of–pl. bend ($b_1$) | ZPE |
|---|---|---|---|---|---|---|---|
| CAS(A)[a] | 3236 | 3140 | 1755 | 1590 | 1302 | 1223 | 0.759 |
| CAS(C)[a] | 3031 | 2795 | 1788 | 1558 | 1294 | 1214 | 0.724 |
| CAS(C)[b] | 3006 | 2789 | 1791 | 1564 | 1304 | 1224 | 0.724 |
| CAS(C)[c] | 2999 | 2780 | 1797 | 1567 | 1311 | 1224 | 0.724 |
| MR-CISD[d] | 3102 | 3024 | 1778 | 1564 | 1285 | 1212 | 0.742 |
| MR-AQCC[d] | 3011 | 2946 | 1773 | 1543 | 1272 | 1170 | 0.726 |
| CCSD[e] | 3042 | 2963 | 1787 | 1531 | 1259 | 1167 | 0.728 |
| DFT[f] | 2947 | 2890 | 1810 | 1531 | 1264 | 1200 | 0.722 |
| Experiment[g] | 2997 | 2978 | 1778 | 1529 | 1299 | 1191 | 0.730 |
| Experiment[h] | 3009 | 2944 | 1764 | 1563 | 1288 | 1191 | 0.729 |
| Experiment[ij] | 2843 | 2782 | 1746 | 1500 | 1249 | 1167 | 0.700 |

**Table 3.2.14:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the ground state of formaldehyde.

[a] Theoretical harmonic frequencies, ANO basis set (C,O $(14s9p4d)/[4s3p1d]$, H $(8s4p)/[2s1p]$ contraction scheme), this work.

[b] Theoretical harmonic frequencies, ANO basis set (C,O $(14s9p4d3f)/[5s4p2d1f]$, H $(8s4p3d)/[3s2p1d]$ contraction scheme), this work.

[c] Theoretical harmonic frequencies, cc-pVQZ basis set (C,O $(5s9p5d3f1g)/[8s7p5d3f1g]$, H $(6s3p2d1f)/[4s3p2d1f]$ contraction scheme), this work.

[d] Theoretical harmonic frequencies, Ref. 97.

[e] Theoretical harmonic frequencies, Ref. 145.

[f] Theoretical harmonic frequencies, Ref. 166.

[g] Experimental harmonic frequencies from stimulated emission spectroscopy, Ref. 91.

[h] Experimental harmonic frequencies, Ref. 167.

[i] Experimental fundamental frequencies, Ref. 167.

[j] FT-IR gas phase measurement of fundamental frequencies, Ref. 168.

For formaldehyde (Tab. 3.2.14) the CAS(A) frequencies are in acceptable agreement with the experimental harmonic ones, apart from the symmetric and asymmetric stretchings of the CH bonds for which too high ($\simeq 200$ $cm^{-1}$) values are found. The use of a larger active space has the effect to improve the agreement with experimental data, in particular lowering the two CH stretchings, but whereas the asymmetric stretching is now very close to experiment, the symmetric one shows a considerable error ($\simeq 150$ $cm^{-1}$). The source for such an error can be located in the incompleteness of the basis set or in the limited account of the correlation effects brought by the CASSCF wavefunction. In order to clarify this point, larger basis sets (ANO with more orbitals and cc-pVQZ) have been

used: the results show quite modest variations with respect to those obtained with the C,O $(14s9p4d)/[4s3p1d]$, H $(8s4p)/[2s1p]$ contraction of the ANO basis sets. The completeness of the basis set is therefore excluded as the source of the discrepancy on the symmetric CH stretching frequency, which can be reasonably ascribed to the lack of dynamical correlation in the CAS(C) wavefunction. This result is confirmed by the comparison with other theoretical data: indeed all methods take into account the dynamic correlation and their agreement with experiment is higher than in the present case.

| Symm. | Approx. mode | CAS(B)[a] | DFT[b] | DFT[c] | Exp.[d] | Exp.[e] | Exp. |
|---|---|---|---|---|---|---|---|
| A′, 1 | $CH_3$ str | 3272 | 3142 | 3136 | 3138 | 3014 | |
| 2 | $CH_3$ str | 3169 | 3030 | 3021 | 3056 | 2923 | |
| 3 | CH str | 3146 | 2878 | 2871 | 2842 | 2716 | |
| 4 | CO str | 1777 | 1802 | 1808 | 1774 | 1743 | 1742[f] |
| 5 | $CH_3$ bend | 1583 | 1466 | 1460 | 1457 | 1433 | |
| 6 | $CH_3$ bend | 1520 | 1424 | 1420 | 1412 | 1395 | |
| 7 | CH bend | 1487 | 1384 | 1377 | 1400 | 1352 | |
| 8 | $CH_3$ rock | 1206 | 1132 | 1133 | 1146 | 1114 | |
| 9 | CC str | 955 | 886 | 886 | 907 | 867 | |
| 10 | CO bend | 525 | 511 | 510 | 521 | 509 | |
| A″, 11 | $CH_3$ str | 3220 | 3079 | 3075 | 3073 | 2964 | |
| 12 | $CH_3$ bend | 1595 | 1475 | 1469 | 1470 | 1448 | |
| 13 | CH bend | 1203 | 1136 | 1128 | 1130 | 1111 | |
| 14 | $CH_3$ rock | 837 | 778 | 776 | 781 | 764 | 764[f] |
| 15 | $CH_3$ torsion | 167 | 158 | 152 | 187 | 150 | 142-144[g] |
| ZPE | | 1.591 | 1.505 | 1.501 | 1.506 | 1.457 | |

**Table 3.2.15:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the ground state of acetaldehyde.

[a] CASSCF harmonic frequencies, this work.
[b] DFT B3LYP/6-311++G(2d,2p) harmonic frequencies, Ref. 166.
[c] DFT B3LYP/6-311++G** harmonic frequencies, Ref. 169.
[d] Experimental harmonic frequencies, Ref. 92.
[e] Experimental fundamental frequencies, Ref. 170.
[f] Fundamental frequencies, Ref. 154.
[g] Fundamental frequencies, Ref. 171. 142 and 144 $cm^{-1}$ correspond to $\nu''_{15}$(E) and $\nu''_{15}$(A) respectively.

For acetaldehyde and acetone similar considerations can be done. The CASSCF frequencies are all too high with respect to the experimental harmonic frequencies with the exception of the CO stretching, for which a good agreement is found: comparing the CAS(A) results for acetone and those obtained at the HF level[166] is possible to note that the two methods give very close results, apart from the CO stretching. As expected, the difference between theory and experiment is even larger if theoretical harmonic and experimental fundamental frequencies are compared, due to the anharmonicity of the potential energy surfaces.

| Symm. | Approx. mode | CAS(A)[a] | HF[b] | DFT[b] | DFT[c] | Exp.[d] | Exp.[e] |
|---|---|---|---|---|---|---|---|
| $A_1$, 1 | $CH_3$ s str | 3274 | 3286 | 3146 | 3020 | 3004 | 3132 |
| 2 | $CH_3$ s str | 3173 | 3180 | 3041 | 2914 | 2937 | 3062 |
| 3 | CO str | 1785 | 1975 | 1778 | 1733 | 1731 | 1760 |
| 4 | $CH_3$ a bend | 1587 | 1592 | 1476 | 1414 | 1435 | 1471 |
| 5 | $CH_3$ s bend | 1519 | 1524 | 1398 | 1330 | 1355 | 1389 |
| 6 | $CH_3$ rock | 1164 | 1182 | 1090 | 1041 | 1072 | 1099 |
| 7 | CC s str | 837 | 840 | 783 | 750 | 777 | 797 |
| 8 | CCC s bend | 397 | 404 | 383 | 364 | 385 | 395 |
| $A_2$, 9 | $CH_3$ a str | 3217 | 3221 | 3086 | 2960 | 2972 | 3097 |
| 10 | $CH_3$ a bend | 1589 | 1591 | 1472 | 1410 | 1426 | 1462 |
| 11 | $CH_3$ rock | 978 | 972 | 892 | 848 | 872 | 894 |
| 12 | $CH_3$ torsion | 71 | 51 | 59 | 22 | 112 | 115 |
| $B_2$, 13 | $CH_3$ a str | 3273 | 3285 | 3145 | 3018 | 3018 | 3147 |
| 14 | $CH_3$ a str | 3166 | 3172 | 3034 | 2907 | 2920 | 3044 |
| 15 | $CH_3$ a bend | 1584 | 1584 | 1465 | 1404 | 1410 | 1446 |
| 16 | $CH_3$ a bend | 1528 | 1533 | 1394 | 1332 | 1364 | 1398 |
| 17 | CC a str | 1329 | 1343 | 1237 | 1183 | 1216 | 1247 |
| 18 | $CH_3$ rock | 982 | 973 | 892 | 848 | 891 | 914 |
| 19 | CO bend | 551 | 578 | 537 | 515 | 530 | 543 |
| $B_1$, 20 | $CH_3$ a str | 3225 | 3231 | 3093 | 2967 | 2972 | 3099 |
| 21 | $CH_3$ a bend | 1612 | 1614 | 1496 | 1431 | 1454 | 1491 |
| 22 | $CH_3$ rock | 1202 | 1232 | 1122 | 1075 | 1090 | 1118 |
| 23 | CO bend | 511 | 535 | 495 | 471 | 484 | 496 |
| 24 | $CH_3$ torsion | 156 | 150 | 141 | 133 | 130 | 133 |
| ZPE | | 2.400 | 2.421 | 2.273 | 2.175 | 2.204 | 2.278 |

**Table 3.2.16:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the ground state of acetone.

[a] CASSCF harmonic frequencies, this work.

[b] HF and DFT B3LYP 6–311++G(2d,2p) harmonic frequencies, Ref. 166.

[c] DFT B3LYP/6–311G** harmonic frequencies scaled by a factor 0.9614, Ref. 122.

[d] Experimental fundamental frequencies, Ref. 172.

[e] Experimental harmonic frequencies estimated using the ratio harmonic/observed values of acetaldehyde, Ref. 172.

On the contrary, if correlation effects on the energy potential are considered (albeit partially as in the CASSCF method) the theoretical estimation is of good quality, as indicated in the present calculation by the CO stretching: in acetaldehyde the CAS(B) value (1777 $cm^{-1}$) compares well with the experimental one (1774 $cm^{-1}$), while in acetone the agreement is only slightly worse (1785 $cm^{-1}$ with CAS(A), 1760 $cm^{-1}$ experimental) but much better than that given by the HF method (1975 $cm^{-1}$).

The DFT method, taking partially into account the correlation effect all over the molecule, gives results[122,166] in closer agreement with experiment when compared with both CASSCF and HF for all the normal coordinates, apart from the CO stretching for which CASSCF gives good results.

### The $n\rightarrow\pi^*$, $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ states

The vibrational frequencies computed in this work for the excited states, together with the experimental and some relevant theoretical data are reported in Tabs. 3.2.17, 3.2.18 and 3.2.19.

For the excited states experimental information on the vibrational frequencies is reduced in number and the theoretical approach can be of great help for the interpretation and assignment of the vibrational structure of spectra. In the case of acetaldehyde and acetone the few published experimental data are not reported in respective tables but they are commented in the text for the sake of clarity.

For excited states, especially in the case of acetaldehyde and acetone, the assignment of a description label to the normal modes is not an easy task (and in some cases meaningless): the molecules have a lower symmetry than the GS and a distorted geometry with elongated bonds, therefore the identification of a normal mode with a group motion can be difficult. In the following the labels used for the ground state have been maintained, but they must be intended in general only as an approximate description of the nuclear motion.

For formaldehyde, the effect of the $n\rightarrow\pi^*$ excitation is to lower the CO stretching, the $CH_2$ rocking and the out-of-plane bending while the CH stretchings show modest variations (the symmetric stretching is slightly increased). This behavior is common to both the singlet and the triplet state, and the agreement with the experimental value is good. Moreover let us note that the singlet and triplet states have very similar theoretical harmonic frequencies.

For the $CH_2$ scissoring mode in the singlet state, particular care must be exerted in analyzing the experimental data. As described by Dallos *et al*,[97] the original value of 887 $cm^{-1}$ obtained analyzing hot bands,[152] has been modified to 1290 $cm^{-1}$ by Hardwick and Till[173] using a tentative assignment of a cold band and to 1429 $cm^{-1}$ by van Dijk *et al*[174] on the basis of theoretical transition intensities. The problem of the experimental assignment of the frequency of such normal mode is therefore still open.

The comparison of theoretical and experimental data is made even more complex if one considers that the theoretically computed values are harmonic frequencies, while experimentally the fundamental frequencies are measured. The CAS(A) value cannot be considered a good estimation given the large change shown when CAS(C) is used.

The CAS(C) value here obtained (1356 $cm^{-1}$) is between the two experimental values, but considering that the anharmonicity of the potential usually gives fundamental frequencies lower than the harmonic ones, the CAS(C) result is in a better agreement with the 1290 $cm^{-1}$ value. The same consideration has been provided by Dallos *et al.*[97]

| Method | asym CH str. ($a''$) | sym CH str. ($a'$) | CO str. ($a'$) | $CH_2$ sciss. ($a'$) | $CH_2$ rock ($a''$) | out-of-pl. bend ($a'$) | ZPE |
|---|---|---|---|---|---|---|---|
| | | | $^1(n\rightarrow\pi^*)$ excited state | | | | |
| CAS(A)[a] | 3392 | 3254 | 1098 | 1514 | 1059 | 839 | 0.692 |
| CAS(C)[a] | 3018 | 2919 | 1128 | 1356 | 976 | 771 | 0.630 |
| MR-AQCC[97,b] | 3134 | 3025 | 1184 | 1353 | 936 | 691 | 0.640 |
| EOM-MBPT[121,b] | 3223 | 3102 | 1187 | 1413 | 933 | 565 | 0.646 |
| EOM-CCSD[175,b] | 3190 | 3073 | 1249 | 1408 | 934 | 720 | 0.656 |
| Exp.[c] | 2968 | 2847 | 1173 | 887 | 904 | | |
| Exp.[d] | | | | 1290[d] | 683[e] | | |
| Exp.[f] | | | | 1429 | | | |
| | | | $^3(n\rightarrow\pi^*)$ excited state | | | | |
| CAS(A)[a] | 3377 | 3243 | 1153 | 1509 | 1031 | 857 | 0.692 |
| CAS(C)[a] | 2974 | 2868 | 1194 | 1322 | 929 | 842 | 0.628 |
| CIS[g] | 2765 | 2700 | 1251 | 1405 | 778 | 721 | 0.596 |
| CCSD(T)[h] | 3127 | 3019 | 1249 | 1333 | 895 | 767 | 0.644 |
| Exp. | | 2871[i] | 1218[j] | | | 524[k] | |
| | | | $S_2$ (mixed $\pi\rightarrow\pi^*$and $\sigma\rightarrow\pi^*$) excited state | | | | |
| CAS(A)[a] | 3383 | 3238 | 717 | 1443 | 985 | 1241 | 0.682 |
| CAS(C)[a] | 3037 | 2911 | 690 | 1366 | 960 | 1113 | 0.625 |
| MR–AQCC[l] | 3266 | 3121 | 815 | 1329 | 987 | 1141 | 0.661 |
| CIS[m] | 3075 | 2932 | 818 | 1351 | 963 | 2204 | 0.703 |
| CISD[158] | 3328 | 3191 | 737 | 1495 | 938 | ··· | |
| | | | $^3(\pi\rightarrow\pi^*)$ excited state | | | | |
| CAS(A)[a] | 3406 | 3261 | 998 | 1548 | 1160 | 832 | 0.695 |
| CAS(C)[a] | 3125 | 2999 | 1022 | 1460 | 1121 | 693 | 0.646 |
| CIS[m] | 3044 | 2938 | 1164 | 1464 | 1159 | 765 | 0.596 |
| Exp. | | | 887[n] | | | | |
| | | | $^3(\sigma\rightarrow\pi^*)$ excited state | | | | |
| CAS(A)[a] | 3501 | 3271 | 807 | 1422 | 1147 | 961 | 0.689 |
| CAS(C)[a] | 3157 | 2954 | 756 | 1306 | 1048 | 876 | 0.626 |

**Table 3.2.17:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the excited states of formaldehyde.

[a] Theoretical harmonic frequencies, this work.
[b] Theoretical harmonic frequencies.
[c] Experimental fundamental frequencies from Ref. 152.
[d] Exp. fundamental frequencies by tentative assignment of a cold band, Ref. 173.
[e] Exp. fundamental frequencies defined as the energy between the average of $\nu(0^+,0^-)\leftrightarrow\nu(1^+,1^-)$, Refs. 129,176.
[f] Exp. fundamental frequencies by assignment of the bands recorder in Ref. 152 using calculated transition intensities, Ref. 174.
[g] Theoretical harmonic frequencies scaled using ground state scaling factors (0.875-0.908), Ref. 129. The two frequencies 1251 and 1405 $cm^{-1}$ are assigned in Ref. 129 to the $CH_2$ scissor and CO stretch respectively.
[h] CCSD(T) theoretical harmonic frequencies using TZ2P(f,d) basis.
[i] Ref. 177 as cited in Ref. 129.
[j] Ref. 178 as cited in Ref. 129.
[k] Ref. 156 as cited in Ref. 129.
[l] Theoretical harmonic frequencies, Ref. 97.
[m] Theoretical harmonic frequencies scaled using ground state scaling factors (0.875-0.908), Ref. 129.
[n] Ref. 156 as cited in Ref. 129.

For the $\pi\rightarrow\pi^*$ and $\sigma\rightarrow\pi^*$ triplet and $S_2$ states only one frequency has been experimentally measured (CO stretching for the $^3(\pi\rightarrow\pi^*)$ state). The $^3(\sigma\rightarrow\pi^*)$

state has never been studied theoretically and for the $^3(\pi \rightarrow \pi^*)$ state only one CIS[129] calculation has been published. The results obtained with CAS(A) and CAS(C) are reported for such states in Tab. 3.2.17. The following comments are based on the CAS(C) values. One promptly notes the sizable lowering of the CO stretching in the $S_2$ (690 $cm^{-1}$) and $^3(\sigma \rightarrow \pi^*)$ (756 $cm^{-1}$) with respect to the $n \rightarrow \pi^*$ states (1128 $cm^{-1}$ for singlet, 1194 $cm^{-1}$ for triplet) and with respect to the GS (1788 $cm^{-1}$). The lowering is less pronounced for the $^3(\pi \rightarrow \pi^*)$ state (1022 $cm^{-1}$).

A common characteristic of all the states reported in Tab. 3.2.17 is the quite high value for the CH stretchings, Another interesting feature of the data reported in Tab. 3.2.17 is the value of the out-of-plane bending frequency for the $S_2$ singlet (1113 $cm^{-1}$): it is closer to the ground state value (1214 $cm^{-1}$) than to the $n \rightarrow \pi^*$ one (771 $cm^{-1}$ for singlet, 842 $cm^{-1}$ for triplet) while for the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ triplets the opposite occurs. The particular behavior of this singlet could be a manifestation of the avoided crossing of the $\pi \rightarrow \pi^*$ and $\sigma \rightarrow \pi^*$ surfaces which leads to the mixing of the two electronic configurations. Finally in all states the $CH_2$ scissoring has a frequency similar to the one found for the $n \rightarrow \pi^*$ states, only the $^3(\pi \rightarrow \pi^*)$ state showing a difference larger than 100 $cm^{-1}$.

| | | $^1(n \rightarrow \pi^*)$ | | $^3(n \rightarrow \pi^*)$ | | | $S_2$ | | $^3(\pi \rightarrow \pi^*)$ | $^3(\sigma \rightarrow \pi^*)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Symm. | Approx. mode | CAS(B)[a] | CIS[b] | CAS(B)[a] | UHF[c] | CIS[b] | CAS(B)[a] | CIS[b] | CAS(B)[a] | CAS(B)[a] |
| A′, 1 | CH str | 3282 | 3313 | *3259* | 3316 | 3319 | *3284* | 3302 | 3305 | 3349 |
| 2 | $CH_3$ str | 3249 | 3271 | *3246* | 3279 | 3283 | *3261* | 3330 | 3241 | 3267 |
| 3 | $CH_3$ str | 3218 | 3202 | 3221 | 3247 | 3199 | 3209 | 3190 | 3206 | 3223 |
| 4 | $CH_3$ str | 3154 | 3187 | 3156 | 3180 | 3055 | 3023 | 3120 | 3138 | 3154 |
| 5 | $CH_3$ bend | 1607 | 1672 | 1609 | 1663 | 1660 | 1591 | 1615 | 1597 | 1600 |
| 6 | $CH_3$ bend | 1596 | 1617 | 1598 | 1651 | 1616 | 1519 | 1543 | 1587 | 1566 |
| 7 | $CH_3$ bend | 1521 | 1607 | 1518 | 1576 | 1612 | 1462 | 1506 | 1529 | 1503 |
| 8 | CH bend | 1393 | 1508 | 1388 | 1422 | 1497 | 1330 | 1265 | 1420 | 1291 |
| 9 | $CH_3$ rock | *1178* | 1277 | *1200* | 1196 | 1240 | 1232 | 1210 | 1190 | 1119 |
| 10 | CC str | *1140* | 1123 | *1121* | 1159 | 1128 | 1063 | 1093 | 1141 | 1174 |
| 11 | $CH_3$ rock | *1092* | 1092 | *1098* | 1114 | 1012 | 891 | 1046 | 1093 | 1036 |
| 12 | CO str | *943* | 968 | *960* | 939 | 924 | 601 | 811 | 890 | 764 |
| 13 | CH bend | 698 | 556 | 711 | 673 | 670 | 697 | 763 | 692 | 669 |
| 14 | CO bend | 399 | 408 | 397 | 388 | 375 | *339* | 404 | 404 | 402 |
| 15 | $CH_3$ torsion | 203 | 202 | 200 | 184 | 202 | *292* | 234 | 220 | 212 |
| ZPE | | 1.530 | 1.550 | 1.530 | 1.549 | 1.537 | 1.481 | 1.515 | 1.528 | 1.508 |

**Table 3.2.18:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the excited states of acetaldehyde. The numbers in italic indicate strongly coupled modes.

[a] CASSCF harmonic frequencies, this work.

[b] Ref. 129.

[c] Ref. 179.

For the acetaldehyde molecule, the previously published calculations of the excited state harmonic frequencies regard only the $^1(n \rightarrow \pi^*)$, $^3(n \rightarrow \pi^*)$ and $S_2$ states and are all performed with single reference based methods. For these states, those here reported are the first example of multireference values, while for the $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states this work presents the first theoretical study. First of all, one notes that the occurrence of quasi-degeneracy of the vibrational levels

(and therefore their mixing) appears in various cases: a similar result was also obtained by Hadad *et al.*[129] The aldehydic CH stretching is the highest for all states, with an increase, with respect to the GS, in the range of 100-200 $cm^{-1}$. In two cases ($^3(n\rightarrow\pi^*)$ and $S_2$) this mode becomes quasi degenerate with one of the $CH_3$ stretchings, leading to a mixing of the two modes. The $CH_3$ stretchings show very modest variations with respect to the GS: the only exception is given by the $S_2$ state, where the lowest frequency is $\simeq$ 120 $cm^{-1}$ lower than in the GS.

In the $n\rightarrow\pi^*$ states, the CO stretching is strongly mixed with other modes ($CH_3$ rocking and CC stretching): the resulting frequencies are in the range 950-1200 $cm^{-1}$. For the other states, the frequency associated to such mode is very low (601, 890 and 764 $cm^{-1}$ for the $S_2$, $^3(\pi\rightarrow\pi^*)$ and $^3(\sigma\rightarrow\pi^*)$ states, respectively) with respect to the GS value (1777 $cm^{-1}$). The CH "in plane" bend (1487 $cm^{-1}$ in the GS) is lowered by $\simeq$ 100 $cm^{-1}$ in the $n\rightarrow\pi^*$ states, by $\simeq$ 160 $cm^{-1}$ in the $S_2$ state and by $\simeq$ 60 and 200 $cm^{-1}$ in the $^3(\pi\rightarrow\pi^*)$ and $^3(n\rightarrow\pi^*)$ states, while for the "out-of-plane" bending (1203 $cm^{-1}$) the lowering is much larger (frequencies in the range 669-711 $cm^{-1}$ are found according to the state). As for the various $CH_3$ bending and rocking modes, one notes that they do not show sizable variations, with the exception of the lowest $CH_3$ bending for the $S_2$ state (where a lowering of almost 60 $cm^{-1}$ is found) and from the lowest $CH_3$ rocking which shows an increase in all states (from 837 $cm^{-1}$ to 1036-1098 $cm^{-1}$) apart from $S_2$ (where the increase is only of $\simeq$ 50 $cm^{-1}$). The CC stretching also shows an increase of the vibrational frequency passing from the GS (955 $cm^{-1}$) to the excited states: the increase is in a close range (170-220 $cm^{-1}$) for all states apart from the $S_2$ one for which it is 108 $cm^{-1}$. It is important to note that, for the $n\rightarrow\pi^*$ states, the lowest $CH_3$ rocking and the CC stretching are badly defined due to the large mixing with other modes, and therefore the above considerations have to be regarded as only indicative. Finally, for the $CH_3$ torsion, for all the states there is a modest increase from 167 $cm^{-1}$ in the GS to 200-220 $cm^{-1}$, apart again the case of the $S_2$ state (292 $cm^{-1}$).

From the experimental point of view, some information is available for the $n\rightarrow\pi^*$ states. For the singlet, one of the $CH_3$ deformations has been assigned[180] to 1254 $cm^{-1}$, the CO stretching to 1119[154] and 1173[180] $cm^{-1}$, the CO out-of-plane bend to 374[160] and 482[180] $cm^{-1}$ and finally the $CH_3$ torsion to 202,[171] 187[154] and 186[181] $cm^{-1}$. The comparison of these data with the results of this work leads to the following considerations:

- for the $CH_3$ deformation the CAS(B) value is 1521 $cm^{-1}$, with a difference of 267 $cm^{-1}$ with the experimental one. In the GS the difference was smaller (100-150 $cm^{-1}$) for the three values indicated in Ref. 180 in the analogous modes (1433, 1352 and 1114 $cm^{-1}$)

- the experimental CO stretching seems to be far from the present assignment. It is worth noticing again that in the calculation on this state and on the $n \rightarrow \pi^*$ triplet, the CO stretching has a component over four different normal modes. Three of them are closer to the experimental values
- for the out–of–plane CO bending the comparison is difficult: it is well known that along this normal coordinate the system shows a double minimum potential and the vibrational wavefunction is delocalized in both minima. The values here reported, obtained using the harmonic approximation in one minimum, appear to be consistent with the experimental assignment
- for the $CH_3$ torsion the agreement is good

For the triplet $n \rightarrow \pi^*$ state four vibrational frequencies have been measured experimentally:[182] 905, 518, 364, and 179 $cm^{-1}$. These values agree well with the present 1098, 711, 397 and 200 $cm^{-1}$, remembering that experimental fundamental and theoretical harmonic frequencies are compared.

As for the acetone molecule, let us note that in the calculations of the excited states, the geometry optimization has been performed in the $C_s$ symmetry group: this allows the molecular skeleton to pyramidalize, but imposes a constraint for the rotation of the methyl groups. Nevertheless the optimized geometries have been confirmed to be true minima (possibly not the global minimum) by a vibrational analysis. Let us note that while for the $^1(n \rightarrow \pi^*)$, $^3(n \rightarrow \pi^*)$ and $S_2$ states CASSCF calculations have been published, for the $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states the one here reported is the first theoretical study.

Some general considerations can be drawn for all the states. First of all, as in acetaldehyde, in some cases difficulties have been found in the assignment of the vibrational modes. In one case the mixing of two modes of the GS was strong, due to the large coupling among them. The CH stretching and the $CH_3$ bending modes show higher frequencies than in the ground state, while the $CH_3$ bending frequencies do not show sizeable variations. The asymmetric and symmetric CC stretchings remain almost unchanged. The $CH_3$ torsions have higher frequencies for all states, especially for the $S_2$ singlet, probably due to a sterical effect originated by the pyramidalization of the carbonyl carbon atom. In all the excited states the CCC bend has a slightly lower frequency than in the GS.

As far as the CO bending modes are concerned, remarkably lower frequencies are found for all states apart from $^1(n \rightarrow \pi^*)$ for the "out-of-plane" ($A''$) mode. This behavior is particularly pronounced for the $S_2$ state where a lowering of more than 200 $cm^{-1}$ is found for the "out-of-plane" bend. On the other hand, for the "in-plane" ($A'$) bend the lowering is larger than 150 $cm^{-1}$ for several states.

| | | $^1(n\rightarrow\pi^*)$ | | | $^3(n\rightarrow\pi^*)$ | | $S_2$ | | $^3(\pi\rightarrow\pi^*)$ | $^3(\sigma\rightarrow\pi^*)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Symm. | Approx. mode | CAS(A)[a] | CAS[b] | CAS[c] | CAS(A)[a] | CAS[c] | CAS(A)[a] | CAS[d] | CAS(A)[a] | CAS(A)[a] |
| A′ 1 | $CH_3$ str | 3245 | 3019 | 3313 | 3245 | 3317 | 3265 | 3269 | 3248 | 3253 |
| 2 | $CH_3$ str | 3212 | 2986 | 3278 | 3212 | 3282 | 3206 | 3209 | 3210 | 3215 |
| 20 | $CH_3$ str | 3142 | 2928 | 3186 | 3142 | 3189 | 3062 | 3048 | 3132 | 3136 |
| 4 | $CH_3$ bend | 1615 | 1400 | 1735 | 1617 | 1745 | 1599 | 1526 | 1608 | 1617 |
| 21 | $CH_3$ bend | 1604 | 1481 | 1734 | 1606 | 1745 | 1533 | 1482 | 1596 | 1584 |
| 5 | $CH_3$ bend | 1531 | 1408 | 1544 | 1530 | 1539 | 1478 | 1219 | 1531 | 1530 |
| 3 | CO str | 1337 | 1226 | 1053 | 1352 | 1077 | 1263 | 1604 | 1287 | 1234 |
| 6 | $CH_3$ rock | 1192 | 1113 | 1455 | 1197 | 1465 | 1122 | 1174 | 1174 | 1196[e] |
| 22 | $CH_3$ rock | 1008[e] | 942 | 1153 | 1024 | 1147 | 640[e] | 906 | 1004 | 981[e] |
| 7 | CCC str | 816 | 768 | 838 | 826 | 847 | 881 | 639 | 793 | 791 |
| 23 | CO bend | 401 | 395 | 407 | 411 | 398 | 393 | 440 | 389 | *400* |
| 8 | CCC bend | 350 | 320 | 371 | 348 | 370 | 366 | 352 | 369 | *348* |
| 24 | $CH_3$ torsion | 201 | 196 | 234 | 203 | 235 | 292 | 261 | 212 | 141 |
| A″, 13 | $CH_3$ str | 3242 | 3021 | 3307 | 3242 | 3314 | 3261 | 3263 | 3243 | 3247 |
| 9 | $CH_3$ str | 3210 | 2992 | 3269 | 3211 | 3276 | 3202 | 3204 | 3205 | 3207 |
| 14 | $CH_3$ str | 3138 | 2927 | 3178 | 3139 | 3184 | 3058 | 3043 | 3128 | 3132 |
| 15 | $CH_3$ bend | 1597 | 1483 | 1577 | 1597 | 1578 | 1600 | 1606 | 1592 | 1600 |
| 16 | $CH_3$ bend | 1590 | 1470 | 1576 | 1591 | 1578 | 1523 | 1524 | 1584 | 1565 |
| 10 | $CH_3$ bend | 1520 | 1399 | 1531 | 1517 | 1524 | 1483 | 1481 | 1527 | 1515 |
| 17 | CCC str | 1308 | 1294 | 1349 | 1292 | 1323 | 1344 | 1341 | 1341 | 1356 |
| 18 | $CH_3$ rock | 1047 | 962 | 1360 | 1042 | 1360 | 1057 | 1046 | 1043 | 1030 |
| 11 | $CH_3$ rock | 1017 | 952 | 1200 | 1023 | 1199 | 782 | 798 | 995 | 984 |
| 19 | CO bend | 530 | 360 | 434 | 400 | 446 | 307 | 336 | 417 | 397 |
| 12 | $CH_3$ torsion | 181 | 170 | 192 | 179 | 192 | 287 | 238 | 190 | 202 |
| ZPE | | 2.358 | 2.183 | 2.435 | 2.353 | 2.438 | 2.294 | 2.294 | 2.345 | 2.335 |

**Table 3.2.19:** Vibrational frequencies ($cm^{-1}$) and ZPE (eV) for the excited states of acetone. The numbers in italic indicate strongly coupled modes.

[a] CASSCF harmonic frequencies, this work.

[b] CASSCF/6–311G** harmonic frequencies scaled by a factor 0.93, Ref. 122.

[c] CASSCF with 8 electrons in 7 orbitals using a C[$3s2p1d$]/H[$2s1p$] basis set, Ref. 138. The authors have not indicated the symmetry of the vibration, here a tentative labeling is given.

[d] CASSCF(6,6)/6–311+G(d,p) harmonic frequencies, Ref. 128.

[e] The harmonic frequency partially describe the CO stretch.

The most interesting mode is however the CO stretching for which a marked lowering is expected given the nature of the excited states. The lowering with respect to the GS (1785 $cm^{-1}$) is of the order of 400-600 $cm^{-1}$ according to the state and is in general less pronounced than in formaldehyde and acetaldehyde, particularly for the $S_2$, $^3(\pi\rightarrow\pi^*)$ and $^3(\sigma\rightarrow\pi^*)$ states where in both molecules frequencies lower than 1000 $cm^{-1}$ have been found (for instance in formaldehyde the reported values were 717, 998 and 807 with CAS(A) for the three states respectively). In acetone this frequency ranges from 1234 $cm^{-1}$ in the $^3(\sigma\rightarrow\pi^*)$ state to 1352 $cm^{-1}$ in the $^3(n\rightarrow\pi^*)$ one. Some disagreement with previously published results is observed: the values obtained by Sakurai and Kato[138] for the two $n\rightarrow\pi^*$ states (1053 and 1077 $cm^{-1}$) are lower by almost 300 $cm^{-1}$ with respect to those here computed. Note that the value obtained by Liao *et al*[122] for the $^1(n\rightarrow\pi^*)$ state (1226 $cm^{-1}$) is scaled by a factor 0.93 (1318 $cm^{-1}$ before scaling) and therefore is in excellent agreement with our result (1337 $cm^{-1}$).

In the case of the $^1(n\rightarrow\pi^*)$ state, refined analysis of the vibrational structure of the spectra have allowed the determination of some vibrational frequencies. In 1985 Baba *et al*[180] gave a detailed vibronic analysis of the fluorescence excitation

spectrum in a supersonic nozzle beam assigning the following frequencies: CO stretching, 1121 $cm^{-1}$; CO out-of-plane bending, 373 and 578 $cm^{-1}$ for the 1+ and 1- levels, respectively; CCC stretching, 757 (sym) and 1294 (asym) $cm^{-1}$; $CH_3$ rocking or deformation, 955 $cm^{-1}$ (tentative assignment). In 1992 Zuckermann *et al*[183] performed a high resolution (0.01 $cm^{-1}$) laser-induced fluorescence study giving the following assignment: $CH_3$ torsions, 156 and 172 $cm^{-1}$; CCC bending, 373 $cm^{-1}$; CO out-of-plane bending, 333 $cm^{-1}$; CO in-plane bending, 177 or 465 $cm^{-1}$ with two different assignment strategies.

The comparison of these values with the ones obtained in this work shows that the CCC symmetric and asymmetric stretching and the CO out-of-plane modes agree with the assignment of Baba *et al.*[180] Moreover the tentative assignment for the 955 $cm^{-1}$ frequencies is confirmed ($CH_3$ rocking). The CASSCF CO stretching is notably higher than the experimental values ($\simeq$ 200 $cm^{-1}$), indicating that the CASSCF potential energy surface is not well described along the CO stretch and/or large anharmonic effects can be expected for this mode. With respect to the results of Zuckermann *et al*,[183] while the $CH_3$ torsions are in good agreement, the assignment of the CCC bending and of the CO out-of-plane bending are probably reversed and for the CO in-plane bending the "second possibility" discussed in Ref. 183 (465 $cm^{-1}$) better agrees with the value here obtained (530 $cm^{-1}$). Gwaltney and Bartlett[121] support the same conclusion on the basis of P-EOM-MBPT(2) calculations.

## Conclusions

This section reported a CASSCF/NEVPT2 evaluation for the adiabatic excitation energies for six valence states (three triplets and three singlets) of three carbonyl molecules (formaldehyde, acetaldehyde and acetone), together with a CASSCF study of the equilibrium geometrical parameters and of the harmonic vibrational frequencies. Particular attention has been paid to deploy a common strategy for the different states and molecules, in order to allow a comparison of the results and to identify, if possible, trend behaviors.

The results show that the introduction of the dynamical correlation through the NEVPT2 theory leads to an improvement of the excitation energies with respect to the CASSCF values. The NEVPT2 values are in general in good agreement with the available experimental data.

As far as the equilibrium geometries are concerned, the results obtained compare well with the available experimental and high-level theoretical data, and allow some controversial assignments to be clarified. For some states the results here presented are the first theoretical characterizations.

The main contributions of this part of the thesis can be summarized as follows:

- the first theoretical description of the harmonic frequencies of the $^3(\sigma \rightarrow \pi^*)$ state of formaldehyde
- the first theoretical description of the geometrical structure and of the harmonic frequencies for the $^3(\pi \rightarrow \pi^*)$ and $^3(\sigma \rightarrow \pi^*)$ states of acetaldehyde and acetone
- the first multireference study of the vibrational frequencies of the $^3(\pi \rightarrow \pi^*)$ state of formaldehyde and of the $n \rightarrow \pi^*$ states (singlet and triplet) of acetaldehyde
- the first example of a common approach to all the valence states of three small carbonyl molecules, thus allowing a meaningful comparison of the results

## 3.3 Creation of a Rydberg augmented basis set

As already anticipated briefly, a Rydberg state is a particular kind of electronically excited state where the electron is promoted into a spatially diffuse orbital. Given its high distance from the remaining molecular electronic system, the spatial "Rydberg" orbital describing the excited electron resembles an atomic orbital, where the ionized molecule is seen as a positive pointcharge. Almost all the structural differences are lost due to distance, therefore slight changes in geometry does not affect the Rydberg orbitals description. For these reasons, the naming convention for Rydberg orbitals resembles the one used for atomic orbitals, starting at quantic number 3: Rydberg orbitals are named $3s$, $3p$, $3d$ and so on. Also, it must be noted that the promoted electron has a very low correlation with the remaining valence electrons.

Description of Rydberg states is complex: a Rydberg orbital is not localized on an atom of the molecule, but on the molecule as a whole. Standard basis sets are not able to correctly describe Rydberg orbitals, therefore an enrichment with diffuse functions is needed to address these cases. Two strategies exist for this task

- use an atom-centered basis set with both low-diffusion functions for the description of the valence and high-diffusion functions for the Rydberg orbitals. Linear combinations emerging from this basis set are able to describe both types of valence and diffuse orbitals, left size of Fig. 3.3.1
- a standard basis set is used for the molecule, and a dummy X atom is added holding a diffuse basis, right side of Fig. 3.3.1

The first solution is simpler to use, but it produces a high number of not interesting combinations. For the description of the 3s Rydberg, for example, an "all positive" linear combination is needed, but this approach also produces $p$, $d$, $f$... functions, arising from other combinations with different sign. These combinations increase the computational cost and introduce problems in the interpretation of data.

The second technique reduces both problems, at the cost of introducing complexity in the creation of the enriched basis set. In the subsequent works, the second strategy has been used.

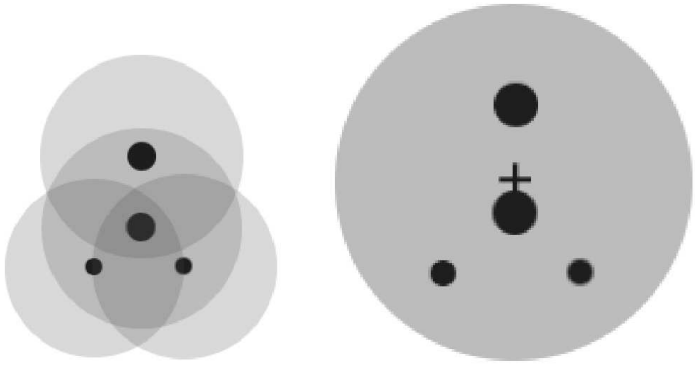

**Figure 3.3.1:** Two possible schemes to define an augmented basis set for the description of diffuse Rydberg orbitals: adding diffuse functions on the single atoms (left side of the picture) or adding diffuse functions supported by a single dummy atom with no charge. In this thesis, the development of the Rydberg orbitals followed the second approach.

### Position of the charge centroid

The position of the X dummy atom is usually chosen to coincide with the position of the positive charge centroid, which depends on the molecular geometry and the excitation performed. Various strategies could be devised in the evaluation of these coordinates, but practical evaluations performed on test molecules confirmed the very reduced dependency of the absolute energy with respect to the dummy atom position. A possible strategy is represented by the procedure described by Roos and coworkers.[184]

Another possible strategy is to choose the center of mass. This approach is simpler, introducing a negligible change in the results, and removes the need to evaluate the molecular dipole. In the evaluations performed later in this thesis both approaches have been used, however as already stated the choice of the position does not produce sizable effects on the final result.

Once placed, the dummy X atom is provided with an uncontracted basis set.

The exponents for the gaussian functions can be obtained through a universal algorithm by Kaufmann *et* al.[185] `DALTON` provides an interface to generate these coefficients by means of the `.CM FUN` keyword. The keyword must be provided with three numbers as defined in Ref. 185: the first number indicates the maximal L quantum number desired for the generated functions.

The second number is related to the starting value for the exponents. The chosen value gives exponents compatible with small molecules, such as formaldehyde or acetone. Kaufmann suggests using a value which provides exponents in smooth continuity with the core basis set.

The last number must be equal to or greater than the second, and produces a new gaussian coefficient for every increment of $\frac{1}{2}$.

In the performed evaluations the choice (2 2 5.5) has been made. The first value 2 will generate $s$, $p$ and $d$-type Rydberg diffuse functions, and the remaining two values 2 and 5.5 produce eight functions. The final basis set on the X dummy atom is therefore the uncontracted $8s8p8d$. Exponents can be extracted from the output file of the `DALTON` run.

### Contraction of the Rydberg basis set

The previous section demonstrated how to produce an uncontracted basis set of diffuse gaussians. These gaussians can be used as-is to enrich the valence basis set, providing a high number of degrees of freedom in the description of high-energy Rydberg orbitals, but the resulting basis set is heavily extended. This introduces a high computational cost and, in the case of CASPT2 evaluations, convergence problems due to intruder states.

For these reasons, it is often useful to contract the Rydberg basis set with appropriate coefficients. A contraction from $8s8p8d$ to $1s1p1d$, for example, leads to a reduction of the number of additional function from 72 to 9. A more flexible choice is the contraction to $3s3p3d$, which gives a good compromise between degrees of freedom and basis set reduction.

An explorative study has been performed to evaluate the influence of Rydberg contraction on the energy of valence and Rydberg states of formaldehyde. Tab. 3.3.1 reports an evaluation performed on various excited valence and Rydberg states of the formaldehyde. Details of the evaluation will be presented in the next section, the only difference being the basis set, in this case cc-pVTZ.[186]

As can be seen, the reported values do not change significantly by decontracting the Rydberg basis set. We can therefore conclude that a contraction from $8s8p8d$ to $1s1p1d$ is mostly sufficient for obtaining valuable results.

The procedure to obtain the contraction coefficients has been implemented *via* both `MOLCAS` and `DALTON`. The `MOLCAS` suite provides the `GENANO` program, while

for the DALTON program an analogous program GENANODAL has been implemented by our research group.

| | | 1 Ryd | 3 Ryd | 8 Ryd |
|---|---|---|---|---|
| | $A_1$ | | | |
| $n_y \rightarrow 3p_y$ | | 8.14 | 8.14 | 8.14 |
| $n_y \rightarrow 3d_{yz}$ | | 9.29 | 9.30 | 9.30 |
| $\pi \rightarrow \pi^*$ | | 9.98 | 9.86 | 9.86 |
| | $A_2$ | | | |
| $n_y \rightarrow \pi^*$ | | 4.05 | 4.03 | 4.04 |
| $n_y \rightarrow 3p_x$ | | 8.47 | 8.48 | 8.49 |
| $n_y \rightarrow 3d_{xz}$ | | 9.47 | 9.50 | 9.50 |
| | $B_2$ | | | |
| $n_y \rightarrow 3s$ | | 7.38 | 7.38 | 7.38 |
| $n_y \rightarrow 3p_z$ | | 8.29 | 8.24 | 8.24 |
| $n_y \rightarrow 3d_{x^2-y^2}$ | | 9.26 | 9.22 | 9.22 |
| $n_y \rightarrow 3d_{z^2}$ | | 9.43 | 9.39 | 9.39 |
| | $B_1$ | | | |
| $n_y \rightarrow 3d_{xy}$ | | 9.38 | 9.40 | 9.40 |
| $\sigma \rightarrow \pi^*$ | | 9.37 | 9.33 | 9.33 |

**Table 3.3.1:** An evaluation of valence and Rydberg excited states transition energies of formaldehyde, using the cc-pVTZ basis set and different diffuse basis set contractions. It can be noted that the contraction $8s8p8d$ to $1s1p1d$ (1 Ryd) produces stable results.

## Using Molcas

The procedure using the MOLCAS suite can be found in the manual,[53] and is also reported by Roos et al.[184] Here a brief discussion will be presented. The needed steps are:

- add the diffuse, uncontracted basis set to the molecule (modifying the Seward input file). The contraction coefficients matrix in this case is a unit matrix
- perform a CASSCF evaluation for the system with one electron removed
- in the resulting output, identify the Rydberg orbitals described by one or (probably) more functions on the X dummy atom
- modify the .RasOrb file, containing the orbital coefficients and their occupation numbers, setting the core occupation numbers to zero. For the Rydberg orbitals, their occupation numbers must be set to small, non-zero values, decreasing of one order of magnitude inside each symmetry. The usual choice is 0.1, 0.01, 0.001 and so on. This prevents mixing of different Rydberg orbitals inside the same symmetry

- run the program GENANO with the modified RasOrb file. This will produce contraction coefficients
- finally, replace the contraction matrix into the Seward input file with the new coefficients

### Using Dalton

The DALTON suite of programs does not implement a strategy for the generation of contracted basis sets, therefore an external program has been developed, named GENANODAL, to accomplish this task. The strategy is as follows:

- perform a Hartree-Fock open shell evaluation for the molecular ion
- use the program IJKLDALI to perform orbital and integral reordering
- run the program GENANODAL with an appropriate input, built using the informations provided by IJKLDALI

An example input file is provided:

```
&LEGGI FILE11='/scr1/stef/$NAME/FILE11',
       FILE25='/scr1/stef/$NAME/FILE25',
  ZANO=T,NOCCUP=8*0.0,
  0.1,0.01,0.001,0.0001,45*0.0,
  0.0,0.1,0.01,21*0.0,
  0.1,0.01,24*0.0,
  0.1,
  NAOX=72,
  IAOX=23,24,25,26,27,28,29,30,
  64,65,66,67,68,69,70,71,
  92,93,94,95,96,97,98,99,
  31,32,33,34,35,36,37,38,
  39,41,43,45,47,49,51,53,
  40,42,44,46,48,50,52,54,
  72,73,74,75,76,77,78,79,
  100,101,102,103,104,105,106,107,
  111,112,113,114,115,116,117,118,
  MAXL=2,NPRIML=8,8,8,NANOL=1,2,3, &END
```

NOCCUP have to be set to dummy occupation numbers for each orbital, honoring the reordering performed by IJKLDALI. As in the case of MOLCAS, these numbers must be set to values in descending order for the Rydberg orbitals of the shell we are interested in.

`NAOX` and `IAOX` represent how many and which atomic functions belong to the X dummy atom, repectively. These values must be extracted from the `IJKLDALI` output. The ordering of the functions must be kept into account: first $s$-type functions, then $p$-type functions, finally $d$-type functions.

`MAXL` is the maximal quantum number L to extract. In this case, up to $d$-type functions are requested.

`NPRIML` is the number of primitives for each quantum number. For contracting a $8s8p8d$ Rydberg basis, these values are 8,8,8.

Finally, `NANOL` are the number of contracted functions generated for each quantum number L. In this case, provided only as an example, 1 $s$-type contracted function will be produced, along with 2 $p$-type functions and 3 $d$-type functions. From the output, only the first column of coefficients will be used.

The `GENANODAL` code provides exactly the same results as obtained from the `MOLCAS` chain.

## 3.4 Evaluation: Rydberg states

*This section is based on the article at Ref. 118.*

This section focuses on the description of vertical transitions to valence and Rydberg excited states with Single-State NEVPT on formaldehyde and acetone. It aims at completing the detailed description of the excited states of carbonyl. In the presented evaluation a state-average CASSCF has been performed. This is different from the previous analysis, where a single state CASSCF approach was used.

### 3.4.1 Formaldehyde

The vertical spectrum of the formaldehyde molecule was computed at the ground state experimental geometry[102] ($R_{CO}$=1.208 Å, $R_{CH}$=1.116 Å and $\theta_{HCH}$=116.5°). As in the previous evaluation, the molecule belongs to the $C_{2v}$ point group symmetry and lies in the $yz$ plane with the C and O atoms on the $z$ axis. The ANO-1 basis set[51] has been used with two different contraction schemes: the smaller (indicated with ANO(S)) is C,O [$4s3p1d$]/H[$2s1p$] and the larger (ANO(L)) is C,O [$6s5p3d2f$]/H[$4s3p2d$]. These valence basis sets have been augmented with diffuse functions using the procedure described in the previous section, in order to properly describe the diffuse orbitals involved in the Rydberg states. These basis functions have been obtained by contraction of a set of $8s8p8d$ gaussian primitives. Two contraction schemes are considered: [$1s1p1d$], (Ryd(S)) and

[$3s3p3d$], (Ryd(L)). For a direct comparison between NEVPT, CASPT2[93] and Size-Consistent Self-Consistent CI ((SC)$^2$-CI),[117] we used in the calculations the combinations ANO(S)-Ryd(S) and ANO(L)-Ryd(L).

The molecular orbitals are obtained from average CASSCF calculations which involve the lowest states of a given symmetry. The active spaces, together with the number and the nature of the states considered in the average procedure are reported in Tab. 3.4.1 and are taken from Ref. 93.

| # MOs[a] | Symmetry and nature of states | # states[b] |
|---|---|---|
| (0340) | $^1$A$_1$ (GS; $n_y \rightarrow 3p_y$,$3d_{yz}$; $\pi \rightarrow \pi^*$) | 4 |
| (2200) | $^1$B$_1$ ($\sigma \rightarrow \pi^*$) | 1 |
| (0211) | $^1$B$_1$ ($n_y \rightarrow 3d_{xy}$) | 1 |
| (4210) | $^1$B$_2$ ($n_y \rightarrow 3s$, $3p_z$, $3d_{x^2-y^2}$, $3d_{z^2}$) | 4 |
| (0410) | $^1$A$_2$ ($n_y \rightarrow \pi^*$, $3p_x$, $3d_{xz}$) | 3 |

**Table 3.4.1:** Active spaces and number of states used in the average CASSCF calculations for the formaldehyde molecule (in all cases 4 active electrons)

[a] number of molecular orbitals in the active space for the four irreducible representations ($a_1$,$b_1$,$b_2$,$a_2$)

[b] number of states used in the average procedure

The number of orbitals for each irreducible representation was chosen by the authors so that all the states of interest could be correctly described. In some cases the active space was enlarged in order to minimize the effect of the intruder states problem in the CASPT2 calculations. Given that NEVPT is not affected by the intruder states problem, in this evaluation a reduction of the active space should be possible but the comparison of our results with those of Refs. 93 and 117 would be in this case less clear.

The energies of the states are computed in a state specific multireference perturbation scheme. The zero-order description of each state is obtained from a CASCI calculation using the average CASSCF active orbitals. The inactive orbitals have been canonized to diagonalize the state-specific Fock matrixes.

A second-order correction to the energy is computed using the NEVPT2 Strongly Contracted and Partially Contracted variants. All orbitals and electrons are included in the perturbative treatment. The excitation energies are computed with respect to the same ground state energy, which is evaluated as the second order correction to the energy with the reference energy and wavefunction obtained from a state-specific CASSCF calculation with four electrons in two $b_1$ ($\pi$ + $\pi^*$) and two $b_2$ ($n_y$ + virtual) orbitals. This approach for the calculation of the excitation energies differs from the one used in the CASPT2 calculation, where a different ground state energy was used for each irreducible representation.

| Method | 2 $A_1$ ($3p_y$) | 3 $A_1$ ($3d_{yz}$) | 1 $B_2$ ($3s$) | 2 $B_2$ ($3p_z$) | 3 $B_2$ ($3d_{x^2-y^2}$) | 4 $B_2$ ($3d_{z2}$) | 2 $A_2$ ($3p_x$) | 3 $A_2$ ($3d_{xz}$) | 2 $B_1$ ($3d_{xy}$) |
|---|---|---|---|---|---|---|---|---|---|
| CASSCF[a,b] | 8.07 | 9.18 | 7.37 | 8.15 | 9.08 | 9.21 | 8.84 | 9.78 | 9.16 |
| CASSCF[a,c] | 8.04 | 9.12 | 7.29 | 8.08 | 8.99 | 9.12 | 8.81 | 9.72 | 9.12 |
| NEVPT SC[a,b] | 8.27 | 9.42 | 7.28 | 8.11 | 9.13 | 9.30 | 8.33 | 9.34 | 9.26 |
| | (0.077) | (0.076) | (0.092) | (0.090) | (0.083) | (0.082) | (0.080) | (0.079) | (0.079) |
| NEVPT SC[a,c] | 8.39 | 9.56 | 7.32 | 8.16 | 9.17 | 9.37 | 8.46 | 9.48 | 9.39 |
| | (0.084) | (0.083) | (0.090) | (0.090) | (0.087) | (0.086) | (0.099) | (0.097) | (0.086) |
| NEVPT PC[a,b] | 8.20 | 9.34 | 7.28 | 8.12 | 9.14 | 9.31 | 8.33 | 9.34 | 9.27 |
| | (0.084) | (0.084) | (0.095) | (0.093) | (0.085) | (0.084) | (0.082) | (0.081) | (0.081) |
| NEVPT PC[a,c] | 8.31 | 9.49 | 7.33 | 8.17 | 9.17 | 9.38 | 8.45 | 9.48 | 9.39 |
| | (0.092) | (0.090) | (0.092) | (0.092) | (0.089) | (0.088) | (0.103) | (0.100) | (0.087) |
| CASPT2[93] | 8.12 | 9.24 | 7.30 | 8.09 | 9.13 | 9.31 | 8.32 | 9.31 | 9.23 |
| MC/BMP[187] | 7.95 | 9.11 | 6.90 | 7.77 | 8.95 | 9.11 | 8.46 | 8.82 | 9.06 |
| $(SC)^2$ CAS+SD[b117] | 8.14 | 9.26 | 7.17 | 7.96 | 9.00 | 9.19 | 8.30 | 9.28 | 9.12 |
| $(SC)^2$ MR+SD[d117] | 8.27 | 9.31 | 7.12 | 7.95 | 8.96 | 9.18 | 8.36 | 9.34 | 9.36 |
| CCR(3)[b117] | 8.01 | 9.16 | 7.11 | 7.91 | 8.99 | 9.21 | 8.25 | 9.26 | 9.12 |
| CCR(3)[c117] | 8.14 | 9.27 | 7.16 | 7.99 | 9.04 | 9.27 | 8.38 | 9.40 | 9.25 |
| MRD-CI[94] | 8.10 | 9.25 | 7.15 | 8.05 | 9.05 | 9.25 | 8.32 | 9.34 | 9.32 |
| MR-CISD + Q[114] | 8.13 | 9.28 | 7.27 | 8.10 | 9.15 | 9.30 | 8.34 | 9.36 | 9.26 |
| MR-AQCC[114] | 8.24 | 9.38 | 7.21 | 8.03 | 9.09 | 9.24 | 8.46 | 9.49 | 9.37 |
| EOM-CCSD[108] | 7.99 | 10.16 | 6.99 | 7.93 | 9.25 | 9.98 | 8.45 | 10.67 | 9.84 |
| EOM-CCSD[125] | 7.98 | 9.13 | 7.04 | 7.88 | 8.94 | 9.12 | 8.21 | 9.29 | 10.89 |
| Exp.[87] | 7.97 | | 7.11 | 8.14 | 8.88 | | 8.37 | | |
| Exp.[156] | | | | | | | | | 9.22 |
| Exp.[188,189] | | | 7.09 | | | | | | |

**Table 3.4.2:** Vertical excitation energies (eV) for the Rydberg states of the formaldehyde molecule. The numbers in parentheses are the squared norms of the first order corrections to the wave function. The squared norm for the ground state is 0.075 (NEVPT SC[b]), 0.076 (NEVPT PC[b]), 0.091 (NEVPT SC[c]) and 0.097 (NEVPT PC[c]).

[a] This work

[b] ANO basis set with contraction $[4s3p1d/2s1p]$ + $1s1p1d$, ANO(S)+Ryd(S)

[c] ANO basis set with contraction $[6s5p3d2f/4s3p2d]$ + $3s3p3d$, ANO(L)+Ryd(L)

The vertical excitation energies obtained in our calculations are reported in Tab. 3.4.2 for the Rydberg states and in Tab. 3.4.3 for the valence states, together with the previous theoretical and experimental results.

In Tab. 3.4.4 we show the comparison between our results and those of Pitarch-Ruiz *et al*,[117] which can be considered a good reference since they involve the whole single plus double excitations space on top of a CAS at a variational level with a size-consistence correction. We remark that the mean absolute error (MAE) of our results is always small with the worst case being represented by the Strongly Contracted NEVPT2 in the ANO(L) + Ryd(L) basis (0.15 eV). The small errors appearing in Tab. 3.4.4 bear out the reliability of NEVPT2 which can yield results of good accuracy, comparable with much more refined calculations, but at a reduced computational cost.

For the case of the smaller basis (ANO(S) + Ryd(S)) the CASPT2 results are also available. NEVPT2 and CASPT2 appear to be of the same quality, with the former showing in all cases a small squared norm of the wavefunction perturbation correction thus getting over the intruder states problem.

As to the comparison with the experimental data, beyond a satisfactory general agreement with our theoretical results we can make the following observations.

| Method | 4 $A_1$ ($\pi \rightarrow \pi^*$) | 1 $A_2$ ($n_y \rightarrow \pi^*$) | 1 $B_1$ ($\sigma \rightarrow \pi^*$) |
|---|---|---|---|
| CASSCF[a,b] | 10.59 | 5.28 | 9.89 |
| CASSCF[a,c] | 10.47 | 5.27 | 9.82 |
| NEVPT SC[a,b] | 10.09 | 4.04 | 9.53 |
| | (0.087) | (0.105) | (0.081) |
| NEVPT SC[a,c] | 9.97 | 3.93 | 9.37 |
| | (0.095) | (0.113) | (0.091) |
| NEVPT PC[a,b] | 9.94 | 4.03 | 9.45 |
| | (0.100) | (0.109) | (0.088) |
| NEVPT PC[a,c] | 9.80 | 3.91 | 9.28 |
| | (0.112) | (0.118) | (0.097) |
| CASPT2[93] | 9.77 | 3.91 | 9.09 |
| MC/BMP[187] | 10.37 | 3.83 | 13.69 |
| $(SC)^2$ CAS+SD[b117] | 9.89 | 4.15 | 9.35 |
| $(SC)^2$ MR+SD[c117] | 9.74 | 4.04 | 9.33 |
| CCR(3)[b117] | 9.80 | 4.01 | 9.29 |
| CCR(3)[c117] | 9.64 | 3.97 | 9.25 |
| MRD-CI[94] | 9.60 | 4.05 | 9.35 |
| MR-CISD + Q[114] | 9.80 | 4.07 | 9.40 |
| MR-AQCC[114] | 9.84 | 4.04 | 9.37 |
| EOM-CCSD[108] | 9.47 | 3.98 | 9.33 |
| EOM-CCSD[125] | 9.37 | 4.04 | 9.43 |
| Exp.[87] | | 4.07 | |
| Exp.[90] | | 3.79 | |

**Table 3.4.3:** Vertical excitation energies (eV) for the valence states of the formaldehyde molecule. The numbers in parentheses are the squared norms of the first order corrections to the wavefunction. The squared norm for the ground state is 0.075 (NEVPT SC[b]), 0.076 (NEVPT PC[b]), 0.091 (NEVPT SC[c]) and 0.097 (NEVPT PC[c]).

[a] This work
[b] ANO(S)+Ryd(S) basis (see text)
[c] ANO(L)+Ryd(L) basis (see text)

First, in accordance with most theoretical calculations our vertical 2 $^1A_1$ and 1 $^1B_2$ Rydberg transitions appear in inverted order with respect to experimental adiabatic transitions:[87] a more stringent comparison would require the calculation of adiabatic transition with due allowance for the zero point energy correction

Second, the calculation of the $^1A_1$ Rydberg states with the larger basis set introduces one more Rydberg state ($n_y \rightarrow \pi^*$) below the valence $\pi \rightarrow \pi^*$. Such state has been ignored in the average CASSCF since we were interested in transitions involving Rydberg orbitals not exceeding the quantum number $n = 3$

Finally, mixing between Rydberg and valence character may occur in both the $^1A_1$ and $^1B_1$ transitions.[93,94] For a correct treatment of such states a quasi-degenerate treatment is required.[190,191]

| | NEVPT PC[a] | NEVPT SC[a] | CASPT2[b] |
|---|---|---|---|
| 2 $A_1$ | 0.06 | 0.13 | -0.02 |
| 3 $A_1$ | 0.08 | 0.16 | -0.02 |
| 4 $A_1$ | 0.05 | 0.20 | -0.12 |
| 1 $B_2$ | 0.11 | 0.11 | 0.13 |
| 2 $B_2$ | 0.16 | 0.15 | 0.13 |
| 3 $B_2$ | 0.14 | 0.13 | 0.13 |
| 4 $B_2$ | 0.12 | 0.11 | 0.12 |
| 1 $A_2$ | -0.12 | -0.11 | -0.24 |
| 2 $A_2$ | 0.03 | 0.03 | 0.02 |
| 3 $A_2$ | 0.06 | 0.06 | 0.03 |
| 1 $B_1$ | 0.10 | 0.18 | -0.26 |
| 2 $B_1$ | 0.15 | 0.14 | 0.11 |
| MAE[c] | 0.10 | 0.13 | 0.11 |

**Table 3.4.4:** Energy differences (eV) between the perturbation and the $(SC)^2$ CAS+SD results of Ref. 117

[a] This work
[b] Ref. 93
[c] Mean Absolute Error

### 3.4.2 Acetone

The computational strategy used for acetone closely follows the one applied to formaldehyde. The vertical spectrum has been computed at the ground state experimental geometry.[102] The molecule belongs to the $C_{2v}$ point group symmetry with the OCCC skeleton in the $yz$ plane (C and O atoms on the $z$ axis and with an orientation of the $CH_3$ groups that place the two H atoms lying in the $yz$ plane as far as possible). For acetone, we only consider the C,O[$4s3p1d$]/H[$2s1p$] contraction of the ANO basis set. The Rydberg states are described using a set of $8s8p8d$ diffuse functions contracted to $1s1p1d$.

As in formaldehyde, average CASSCF calculations provide the molecular orbitals: the active spaces and the number and the nature of the states considered in the average procedure are reported in Tab. 3.4.5 and are taken from Ref. 124.

With respect to formaldehyde, the active space has been modified adding the two CO $\sigma$ and $\sigma^*$ orbitals and the two CO $\sigma$ electrons, except for the $A_1$ symmetry where three virtual orbitals (one of $b_1$ and two of $b_2$ symmetry) have been removed. In Ref. 124 the CO $\sigma$ and $\sigma^*$ orbitals have been added to the active space to correctly describe the adiabatic electronic transitions for the valence states, in which an elongation of the CO bond is observed.

Given that we present here only results on the vertical transitions, also in the case of acetone a reduction of the active space used in Ref. 124 would have been possible, but we have chosen to maintain the same active space in order to have a meaningful comparison with the CASPT2 data.

| # MOs [a] | Symmetry and nature of states | # states[b] |
|---|---|---|
| (2230) | $^1A_1$ (GS; $n_y \rightarrow 3p_y$,$3d_{yz}$; $\pi \rightarrow \pi^*$) | 4 |
| (2200) | $^1B_1$ ($\sigma \rightarrow \pi^*$) | 1 |
| (2211) | $^1B_1$ ($n_y \rightarrow 3d_{xy}$) | 1 |
| (6210) | $^1B_2$ ($n_y \rightarrow 3s$, $3p_z$, $3d_{x^2-y^2}$, $3d_{z^2}$) | 4 |
| (2410) | $^1A_2$ ($n_y \rightarrow \pi^*$, $3p_x$, $3d_{xz}$) | 3 |

**Table 3.4.5:** Active spaces and number of states used in the average CASSCF calculations for the acetone molecule (always 6 active electrons except in the case of the $^1B_1$ ($\sigma \rightarrow \pi^*$) which is computed with 4 active electrons)

[a] number of molecular orbitals in the active space for the four irreducible representations ($a_1$,$b_1$,$b_2$,$a_2$)

[b] number of states used in the average procedure

The energies of the states are computed following the strategy outlined for formaldehyde and the transition energies are reported in Tab. 3.4.6 for the Rydberg states and in Tab. 3.4.7 for the valence states, together with the results of other theoretical calculations and with some experimental results.

We note that our Partially Contracted NEVPT2 results compare very well with the CASPT2 ones. We also remark that in two of the valence transitions ($4\ ^1A_1$, $\pi \rightarrow \pi^*$ and $1\ ^1A_2$, $n_y \rightarrow \pi^*$) the difference between Strongly and Partially Contracted NEVPT2 appears to be unusually sizable (0.59 eV and 0.20 eV, respectively). We think that this is symptom for the zero-order wavefunction to necessitate significant improvement.

| Method | 2 $A_1$ ($3p_y$) | 3 $A_1$ ($3d_{yz}$) | 1 $B_2$ ($3s$) | 2 $B_2$ ($3p_z$) | 3 $B_2$ ($3d_{x2-y2}$) | 4 $B_2$ ($3d_{z2}$) | 2 $A_2$ ($3p_x$) | 3 $A_2$ ($3d_{xz}$) | 2 $B_1$ ($3d_{xy}$) |
|---|---|---|---|---|---|---|---|---|---|
| CASSCF[a] | 7.91 | 8.46 | 6.02 | 6.75 | 7.30 | 7.39 | 7.29 | 7.99 | 7.38 |
| NEVPT2 SC[a] | 7.40 | 8.03 | 6.75 | 7.67 | 8.25 | 8.37 | 7.48 | 8.24 | 8.36 |
| | (0.182) | (0.179) | (0.159) | (0.155) | (0.154) | (0.153) | (0.168) | (0.166) | (0.154) |
| NEVPT2 PC[a] | 7.27 | 7.91 | 6.71 | 7.64 | 8.22 | 8.34 | 7.39 | 8.17 | 8.35 |
| | (0.195) | (0.192) | (0.166) | (0.161) | (0.160) | (0.158) | (0.177) | (0.175) | (0.158) |
| CASPT2[124] | 7.26 | 7.91 | 6.58 | 7.48 | 8.04 | 8.18 | 7.34 | 8.09 | 8.20 |
| EOM-CCSD[108] | 7.45 | 8.23 | 6.39 | 7.51 | 7.95 | 8.48 | 7.41 | 8.44 | 8.43 |
| EOM-CCSD[125] | 7.41 | 8.02 | 6.42 | 7.39 | 7.82 | 8.10 | 7.31 | 8.04 | 8.11 |
| Exp.[124] | | 7.8 | | | 8.09 | | | | 8.17 |
| Exp.[192] | | | 6.35 | | | | | | |
| Exp.[87] | | | 6.36 | | | | 7.45 | | |
| Exp.[193] | 7.41 | | | 7.45 | | | 7.36 | | |

**Table 3.4.6:** Vertical excitation energies (eV) for the Rydberg states of the acetone molecule

[a] This work. The numbers in parentheses are the squared norms of the first order corrections to the wavefunction. The squared norm for the ground state is 0.164 (NEVPT2 SC) and 0.167 (NEVPT2 PC).

| Method | 4 $A_1$ ($\pi \rightarrow \pi^*$) | 1 $A_2$ ($n_y \rightarrow \pi^*$) | 1 $B_1$ ($\sigma \rightarrow \pi^*$) |
|---|---|---|---|
| CASSCF[a] | 11.60 | 5.57 | 10.38 |
| NEVPT2 SC[a] | 9.60 | 4.42 | 9.29 |
| | (0.204) | (0.189) | (0.182) |
| NEVPT2 PC[a] | 9.01 | 4.22 | 9.23 |
| | (0.293) | (0.207) | (0.190) |
| CASPT2[124] | 9.16 | 4.18 | 9.10 |
| EOM-CC[108] | 9.15 | 4.48 | 9.30 |
| EOM-CC[125] | 8.52 | 4.47 | 8.87 |
| Exp.[90] | | 4.38 | |
| Exp.[87] | | 4.43 | |

**Table 3.4.7:** Vertical excitation energies (eV) for the valence states of the acetone molecule

[a] This work. The numbers in parentheses are the squared norms of the first-order corrections to the wavefunction. The squared norm for the ground state is 0.164 (NEVPT2 SC) and 0.167 (NEVPT2 PC).

### Final remarks

Among the formal requirements satisfied by NEVPT2, the absence of intruder states appears particularly interesting for the application to the calculation of electronically excited states. The results shown in the precedent section for the vertical transitions of formaldehyde and acetone are of good quality and exhibit good agreements with the best calculations so far performed as well as with the existing experimental data.

Our calculations have been carried out starting from rather modest size CASSCF wavefunctions. The computational overhead involved by the two forms of NEVPT2 (Strongly and Partially Contracted) amounts to only a small fraction of the CASSCF calculation for such small active orbital spaces and this favorable situation is not expected to drastically change when passing on to larger molecules, provided that the active space can be kept within manageable dimensions.

No evidence of divergences or misbehaviors in the perturbation summation has been found in the calculation of the Rydberg states, which are particularly prone to exhibiting the appearance of intruder states.

## 3.5 Quasi-Degenerate NEVPT2

Single-State NEVPT2, presented in the previous sections, performs the perturbative treatment on a given state, disregarding possible interaction between multiple states which are nearly degenerate before or after the perturbation. This is due to the definition of a state-specific zero-order Hamiltonian. To address the needs

for quasi-degeneracy problems, frequently encountered when analyzing avoided crossings, the Quasi-Degenerate NEVPT2 (QD-NEVPT) approach has been implemented.[194]

The approach used for NEVPT follows the description made by Lindgren.[195] We define a model space, as generated by a set of $n$ eigenfunctions, in our case of CASSCF type $\left\{\Psi_1^{(0)}, \Psi_2^{(0)}, \ldots, \Psi_n^{(0)}\right\}$ which undergo altogether the perturbative correction. These functions define a space $\mathcal{P}$, and have associated zero-order energies $E_i^{(0)}$. We can also define the complementary space $\mathcal{Q} = 1 - \mathcal{P}$ with the set of functions $\left\{\Psi_{n+1}^{(0)}, \Psi_{n+2}^{(0)}, \ldots\right\}$. Projection operators can be defined for each space

$$\hat{\mathcal{P}} = \sum_{i=1}^{n} \left|\Psi_i^{(0)}\right\rangle \left\langle\Psi_i^{(0)}\right| \tag{3.5.1}$$

$$\hat{\mathcal{Q}} = \sum_{a>n} \left|\Psi_a^{(0)}\right\rangle \left\langle\Psi_a^{(0)}\right| \tag{3.5.2}$$

It can be demonstrated that the following relationships hold

$$\hat{\mathcal{P}}^2 = \hat{\mathcal{P}}^+ = \hat{\mathcal{P}} \tag{3.5.3}$$

$$\hat{\mathcal{P}}\hat{\mathcal{Q}} = \hat{\mathcal{Q}}\hat{\mathcal{P}} = 0 \tag{3.5.4}$$

Defining the true eigenfunctions of the Schrödinger equation as

$$\hat{\mathcal{H}}\Psi_i = E_i\Psi_i \tag{3.5.5}$$

we can now write the projection of the true eigenfunctions inside the spaces defined above

$$\tilde{\Psi}_i = \hat{\mathcal{P}}\Psi_i \qquad \forall 1 \leq i \leq n \tag{3.5.6}$$

$$\tilde{\Psi}_a = \hat{\mathcal{Q}}\Psi_a \qquad \forall a > n \tag{3.5.7}$$

and define a "wave operator" $\hat{\Omega}$ which produces the true eigenfunctions $\Psi_i$ from the projected ones

$$\hat{\Omega}\tilde{\Psi}_i = \Psi_i \tag{3.5.8}$$

but produces zero when applied to the complementary space

$$\hat{\Omega}\tilde{\Psi}_a = 0 \tag{3.5.9}$$

It can be proved that the following relationships hold

$$\hat{\mathcal{P}}\hat{\Omega} = \hat{\mathcal{P}} \tag{3.5.10}$$

$$\hat{\Omega}\hat{\mathcal{P}} = \hat{\Omega} \tag{3.5.11}$$

$$\hat{\Omega}\hat{\mathcal{Q}} = 0 \tag{3.5.12}$$

It is possible to rewrite the Schrödinger equation in the form

$$\hat{\mathcal{H}}\hat{\Omega}\tilde{\Psi}_i = E_i\hat{\Omega}\tilde{\Psi}_i \tag{3.5.13}$$

and applying $\hat{\mathcal{P}}$ on both sides

$$\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\Omega}\tilde{\Psi}_i = E_i\tilde{\Psi}_i \tag{3.5.14}$$

$$\hat{\mathcal{H}}_{\text{eff}}\tilde{\Psi}_i = E_i\tilde{\Psi}_i \tag{3.5.15}$$

where an effective Hamiltonian $\hat{\mathcal{H}}_{\text{eff}} = \hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\Omega}$ has been defined. This Hamiltonian produces the eigenvalues $E_i$ of the true Hamiltonian, but operates in the model space.

The final goal of this technique is to obtain the matrix $\left\langle \Psi_n^{(0)} \middle| \hat{\mathcal{H}}_{\text{eff}} \middle| \Psi_m^{(0)} \right\rangle$ and diagonalize it. This matrix has the dimensionality of the number of states considered in the Quasi-Degenerate scheme. Once diagonalized, it provides the corrected energies and the coefficients mixing the perturbed wavefunctions.

To accomplish this task, further elaborations are needed: making use of Eqn. 3.5.15 by applying on both sides $\hat{\Omega}$ and subtracting 3.5.13 we obtain

$$\left(\hat{\Omega}\hat{\mathcal{H}}_{\text{eff}} - \hat{\mathcal{H}}\hat{\Omega}\right)\tilde{\Psi}_i = 0 \tag{3.5.16}$$

which is valid also for $\tilde{\Psi}_a$, therefore the operator itself is zero

$$\hat{\Omega}\hat{\mathcal{H}}_{\text{eff}} - \hat{\mathcal{H}}\hat{\Omega} = 0 \tag{3.5.17}$$

Replacing the definition of $\hat{\mathcal{H}}_{\text{eff}}$ we obtain the generalized Bloch equation

$$\hat{\Omega}\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\Omega} - \hat{\mathcal{H}}\hat{\Omega} = 0 \tag{3.5.18}$$

Introducing the expression of the Hamiltonian as $\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{V}$ we can now perform a substitution inside the Bloch equation to obtain

$$\left[\hat{\Omega}, \hat{\mathcal{H}}_0\right] = \hat{\mathcal{Q}}\hat{V}\hat{\Omega} - \hat{\mathcal{Q}}\hat{\Omega}\hat{\mathcal{P}}\hat{V}\hat{\Omega} \tag{3.5.19}$$

To proceed further in the development, a perturbative expansion for the wave operator and the effective Hamiltonian is performed

$$\hat{\Omega} = \hat{\mathcal{P}} + \hat{\Omega}^{(1)} + \hat{\Omega}^{(2)} + \dots \tag{3.5.20}$$

and the following relationships can be obtained

$$\left[\hat{\Omega}^{(1)}, \hat{\mathcal{H}}_0\right] = \hat{\mathcal{Q}}\hat{V}\hat{\mathcal{P}} \tag{3.5.21}$$

$$\left[\hat{\Omega}^{(2)}, \hat{\mathcal{H}}_0\right] = \hat{\mathcal{Q}}\hat{V}\hat{\Omega}^{(1)} - \hat{\mathcal{Q}}\hat{\Omega}^{(1)}\hat{\mathcal{P}}\hat{V}\hat{\mathcal{P}} \tag{3.5.22}$$

$$\vdots$$

Applying the commutator in 3.5.21 to $|\Psi_i\rangle$ and keeping into account that $\hat{\mathcal{Q}}\hat{\Omega}^{(1)} = \hat{\Omega}^{(1)}$ we obtain

$$\left(E_i^{(0)} - \hat{\mathcal{Q}}\hat{\mathcal{H}}_0\hat{\mathcal{Q}}\right)\hat{\Omega}^{(1)}\left|\Psi_i\right\rangle = \hat{\mathcal{Q}}\hat{V}\left|\Psi_i\right\rangle \tag{3.5.23}$$

which can be rearranged to produce

$$\hat{\Omega}^{(1)}\left|\Psi_i^{(0)}\right\rangle = \left(E_i^{(0)} - \hat{\mathcal{Q}}\hat{\mathcal{H}}_0\hat{\mathcal{Q}}\right)^{-1}\hat{\mathcal{Q}}\hat{V}\left|\Psi_i^{(0)}\right\rangle \tag{3.5.24}$$

and knowing that

$$\hat{\mathcal{Q}}\hat{\mathcal{H}}_0\hat{\mathcal{Q}} = \sum_a \left|\Psi_a^{(0)}\right\rangle E_a^{(0)}\left\langle\Psi_a^{(0)}\right| \tag{3.5.25}$$

we obtain

$$\hat{\Omega}^{(1)}\left|\Psi_i^{(0)}\right\rangle = \sum_{a>n}\left|\Psi_a^{(0)}\right\rangle\frac{\left\langle\Psi_a^{(0)}\left|\hat{V}\right|\Psi_i^{(0)}\right\rangle}{E_i^{(0)} - E_a^{(0)}} \tag{3.5.26}$$

We can now apply the perturbative expansion also to $\hat{\mathcal{H}}_{\text{eff}}$

$$\hat{\mathcal{H}}_{\text{eff}} = \hat{\mathcal{H}}_{\text{eff}}^{(0)} + \hat{\mathcal{H}}_{\text{eff}}^{(1)} + \hat{\mathcal{H}}_{\text{eff}}^{(2)} + \ldots \tag{3.5.27}$$

to obtain the following terms

$$\hat{\mathcal{H}}_{\text{eff}}^{(0)} = \hat{\mathcal{P}}\hat{\mathcal{H}}^{(0)}\hat{\mathcal{P}} \tag{3.5.28}$$

$$\hat{\mathcal{H}}_{\text{eff}}^{(1)} = \hat{\mathcal{P}}\hat{V}\hat{\mathcal{P}} = 0 \tag{3.5.29}$$

$$\hat{\mathcal{H}}_{\text{eff}}^{(2)} = \hat{\mathcal{P}}\hat{V}\hat{\Omega}^{(1)} \tag{3.5.30}$$

$$\vdots \tag{3.5.31}$$

hence, the $\mathbf{H}_{\text{eff}}$ matrix up to the second-order can be obtained from the formulation of the $\hat{\mathcal{H}}_{\text{eff}}^{(2)}$ operator.

When applying the above formulations to the NEVPT case, we have to consider that the model space is generated by the CAS wavefunctions $\left\{\Psi_m^{(0)}\right\}$, and the zero-order Hamiltonian is defined through its spectral decomposition

$$\hat{\mathcal{H}}_0(m) = \left|\Psi_m^{(0)}\right\rangle E_m^{(0)}\left\langle\Psi_m^{(0)}\right| \tag{3.5.32}$$

$$+ \sum_{m'\neq m}\left|\Psi_{m'}^{(0)}\right\rangle E_{m'}^{(0)}\left\langle\Psi_{m'}^{(0)}\right| \tag{3.5.33}$$

$$+ \sum_{l,k,\mu}\left|\Psi_{l,\mu}^k(m)\right\rangle E_{l,\mu}^k(m)\left\langle\Psi_{l,\mu}^k(m)\right| \tag{3.5.34}$$

where $\Psi_{l,\mu}^k(m)$ are the perturbers generated by the application of excitation operators to the state $m$. As a consequence, the zero-order Hamiltonian is state dependen. A partition of the Hamiltonian in a state-specific way is needed using the multipartitioning approach.[196] For each state $m$ we define a different partitioning scheme of the Hamiltonian

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0(m) + \hat{V}(m) \tag{3.5.35}$$

It must be noted however that both spaces $\mathcal{P}$ and $\mathcal{Q}$ (and consequently the projection operators) are not state-specific, since $\mathcal{P}$ is defined by the CAS space, and $\mathcal{Q} = 1 - \mathcal{P}$. We obtain for $\hat{\Omega}^{(1)}$

$$\hat{\Omega}^{(1)} \left|\Psi_m^{(0)}\right\rangle = \sum_{k,l,\mu} \frac{\left|\Psi_{l,\mu}^k(m)\right\rangle \left\langle \Psi_{l,\mu}^k(m) \left|\hat{\mathcal{H}}\right| \Psi_m^{(0)}\right\rangle}{E_m^{(0)} - E_{l,\mu}^k(m)} \tag{3.5.36}$$

and the final $\hat{\mathcal{H}}_{\text{eff}}$ matrix element is therefore

$$\left\langle \Psi_n^{(0)} \left|\hat{\mathcal{H}}_{\text{eff}}\right| \Psi_m^{(0)}\right\rangle = \delta_{nm} E_m^{(0)} + \sum_{l,k,\mu} \frac{\left\langle \Psi_n^{(0)} \left|\hat{\mathcal{H}}\right| \Psi_{l,\mu}^k(m)\right\rangle \left\langle \Psi_{l,\mu}^k(m) \left|\hat{\mathcal{H}}\right| \Psi_m^{(0)}\right\rangle}{E_m^{(0)} - E_{l,\mu}^k(m)} \tag{3.5.37}$$

It can be noted that the diagonal elements of the obtained matrix are the single state contributions. In a Single-State approach, the non-diagonal elements of the effective matrix are assumed as zero. The Quasi-Degenerate approach provides these non-diagonal elements, whose role is to mix the perturbed wavefunctions.

The $\mathbf{H}_{\text{eff}}$ matrix in general is not hermitian, but a hermitian variant can be obtained:[197]

$$\mathbf{H}'_{\text{eff}} = \mathbf{S}^{-\frac{1}{2}} \mathbf{H}_{\text{eff}} \mathbf{S}^{\frac{1}{2}} \tag{3.5.38}$$

with $S_{ij} = \left\langle \tilde{\Psi}_i \middle| \tilde{\Psi}_j \right\rangle$.

## 3.6 Evaluation: QDPT on formaldehyde

*This section is based on the article at Ref. 198*

A Quasi-Degenerate NEVPT2 evaluation has been conducted over the formaldehyde molecule. Four electronic states of $A_1$ symmetry have been considered in the analysis, namely the ground state ($1A_1$), the $n_y \rightarrow 3p_y$ Rydberg state ($2A_1$), the $n_y \rightarrow 3d_{yz}$ Rydberg state ($3A_1$) and the $\pi \rightarrow \pi^*$ valence state ($4A_1$).

The energy curves for these states have been studied as a function of the CO distance. The rationale behind this study is to analyze the qualitative behavior of the avoided crossing of the potential energy curves arising from these states. A simple CASSCF wavefunction does not keep fully into account the dynamic correlation needed for the description of the ground and $\pi \rightarrow \pi^*$ state. These states require higher correlative corrections than the ones needed by the Rydberg states, due to the high average distance of the promoted electron from the molecular framework in the Rydberg case. Also, the large mixing occurring between Rydberg and valence states invokes the need for a Quasi-Degenerate treatment.

A state averaged CASSCF procedure with an active space of six electrons in seven orbitals has been performed. The interested active orbitals are the $\sigma$, $\sigma^*$,

the $\pi$, $\pi^*$, the $n_y$ and the $3p_y$ and $3d_{yz}$ Rydberg orbitals. To describe the potential curve, CO bond distances from 2.00 bohr to 3.40 bohr have been considered, with a step of 0.1 bohr. Occasionally, the step length has been reduced to address particular features of the curves. Other geometrical parameters have been kept fixed at the values obtained by a HF/MP2 complete optimization with the same basis set.

The basis set is the ANO-1, augmented with a Rydberg set of functions built as detailed in Sec. 3.3. The contraction scheme produced a set $[1s1p1d]$ from a decontracted set of diffuse functions $[8s8p8d]$, centered on the charge centroid evaluated at the ground state geometry. The position of the centroid is unchanged during the CO elongation, given its marginal influence on the final results.

The procedure is briefly presented here: for each molecular geometry an averaged CASSCF evaluation has been performed, producing a set of averaged molecular orbitals. For each state, the average orbitals are brought in canonical form in the inactive set.

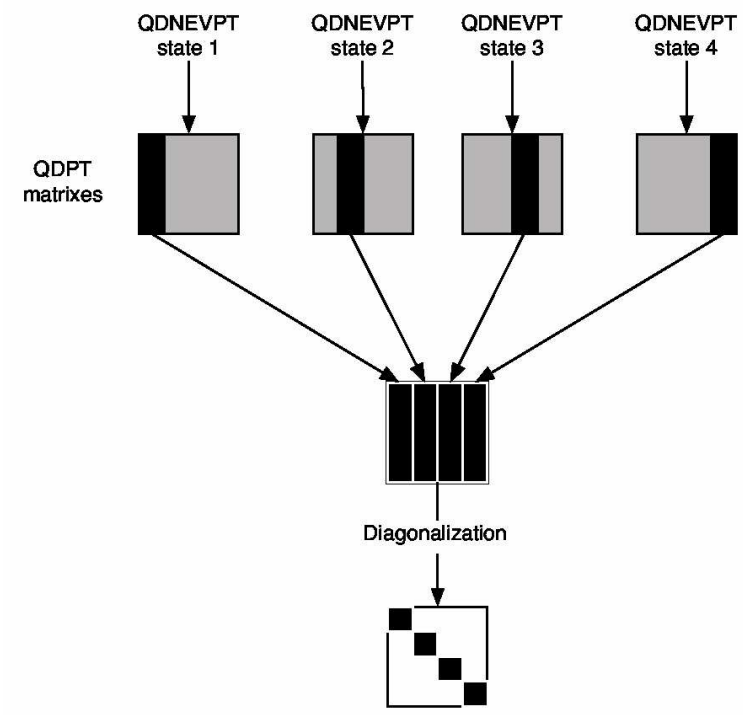


**Figure 3.6.1:** The computational scheme for the evaluation of the Quasi Degenerate treatment. Four different QDPT matrixes are obtained by the procedure, each one referring to a particular state. A single column is extracted from each matrix and gathered in a final matrix which is diagonalized.

This procedure provides four orbital sets, which are used in the Quasi-Degenerate evaluation. Four matrixes are obtained (refer to Fig. 3.6.1) each one referring to

a particular state. To build the final effective Hamiltonian matrix to be diagonalized, the following steps must be applied:

1. decompose the four matrixes columnwise
2. extract the column number $n$ from the matrix referring to the $n$-th electronic state
3. build a final matrix, made of the obtained columns
4. diagonalize it to obtain the final energies and coefficients

The curves obtained at CASSCF level are represented in Fig. 3.6.2. The main feature is the behavior of the $\pi \rightarrow \pi^*$ $4A_1$ state with respect to $2A_1$ and $3A_1$ Rydberg states. The Rydberg states are nearly parallel to the ground state. This is expected since the electron is promoted from a non-bonding orbital to a distant one, introducing very little destabilization in the electronic asset that governs the molecular geometry.

On the other hand, the valence excited state introduces a serious variation in this asset, moving from a double bond situation to a nearly single bond situation. This imposes a steep elongation of the bond, and the potential energy curve for this state crosses the Rydberg ones. Due to interaction between these states, an avoided crossing situation occurs, producing the resulting diagram.

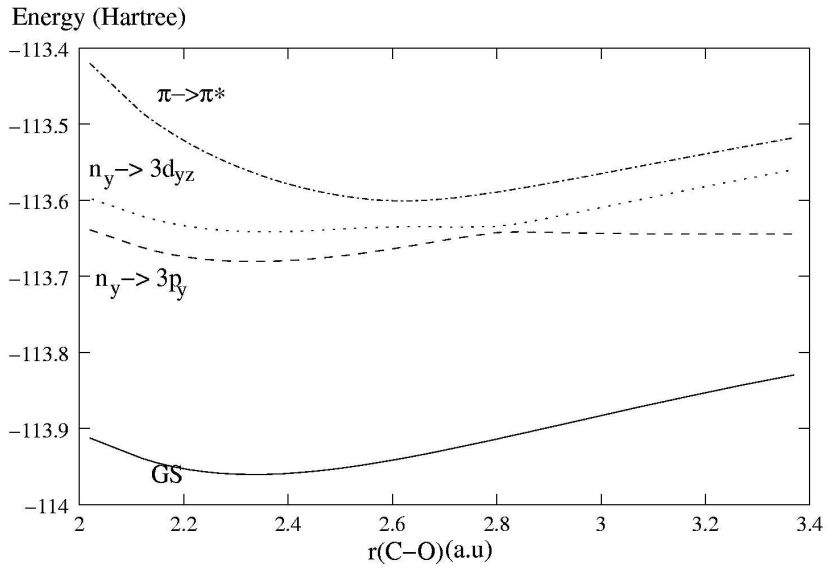


**Figure 3.6.2:** CASSCF energy curves for the ground state, $n \rightarrow$ Rydberg and $\pi \rightarrow \pi^*$ of formaldehyde with respect to the CO bond lenght.

Applying the perturbation treatment, it is expected that the correlation for valence states will be higher than for Rydberg states. As a direct consequence the

avoided crossings move at shorter distances, see Fig. 3.6.3, solid line. The same plot also reports the results produced by a Single-State NEVPT2 treatment (dotted lines). These curves have no physical meaning and behave in incorrect ways. The presence of non-diagonal elements in the effective Hamiltonian matrix fixes the invalid behavior, allowing the wavefunction to interact at NEVPT2 level.

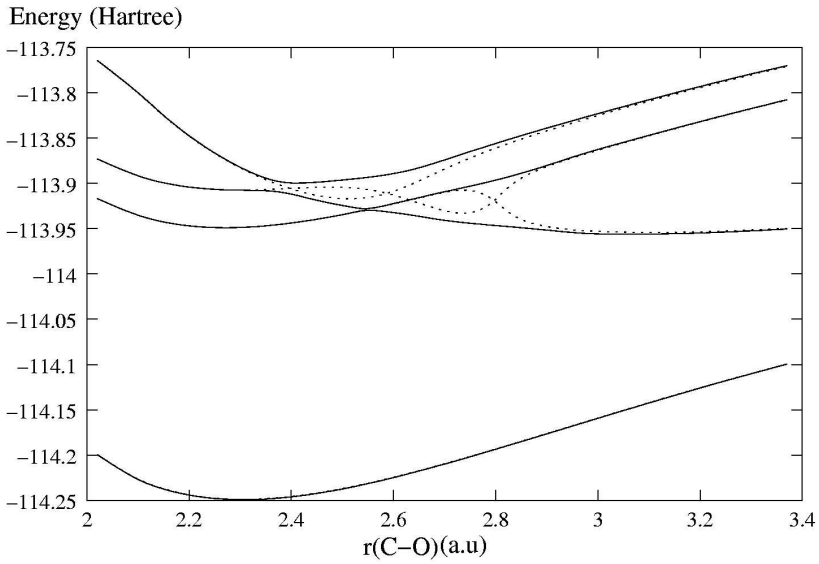


**Figure 3.6.3:** Energy curves for the ground state, $n \rightarrow$ Ryd and $\pi \rightarrow \pi^*$ of formaldehyde with respect to the CO bond lenght. The solid line represents the QD-NEVPT evaluation. The dotted lines report the Single-State evaluation.

A final consideration must however be raised working with a more accurate point grid. The plot in Fig. 3.6.4 represents the $2A_1$ and $3A_1$ Rydberg states at QD-NEVPT level of theory, using a finer grid (step 0.01 bohr) between 2.7 bohr and 2.9 bohr. An incorrect behavior is observed for both curves. This behavior is not appreciable in the Ground State and $\pi \rightarrow \pi^*$ curves. For comparison, the avoided crossing at CASSCF level is reported between the QDPT curves. The CASSCF curves have been transposed in energy to fit correctly into the plot.

It can be noted that the position of the incorrect behavior matches the position of the avoided crossings at CASSCF level. This is probably due to a wrong estimation of the coupling between the curves. Tests performed with larger CAS spaces lead to a reduction of the incorrect behavior, which expresses itself at shorter distances, following again the avoided crossing. A possible explanation of the problem lies in a bad description of the zero-order situation, and the inability for the NEVPT2 treatment to fully recover the coupling. This brings a "memory-effect" on the QDPT curves. A non trivial solution to the problem could be looked for in the definition of an iterative procedure, where the QD-NEVPT treatment is performed on the refined wavefunctions obtained from the previous step.

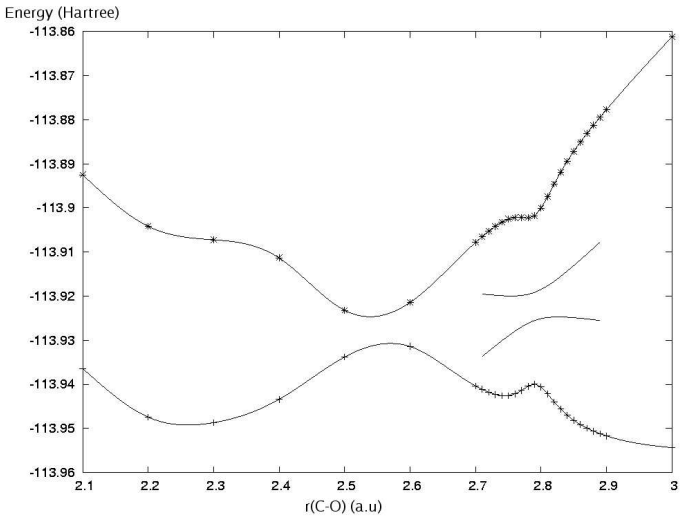


**Figure 3.6.4:** The incorrect behavior of the $2A_1$ (cross symbol) and $3A_1$ (star symbol) Rydberg states at QD-NEVPT level, with respect to the elongation of the CO bond. A grid step of 0.01 bohr between 2.7 and 2.9 bohr has been used. Between the curves the plot also depicts the CASSCF avoided crossing taking place at the same bond length. The avoided crossing curves have been transposed to better fit into the plot.

## 3.7 Non Canonical NEVPT

All the cases so far examined of NEVPT were performed assuming canonical orbitals, since the situation is greatly simplified using these orbitals. Let us consider the inactive part of the Dyall's Hamiltonian in generalized form

$$\hat{\mathcal{H}}_i^D = \sum_{ij} f_{ij} E_{ij} + \sum_{rs} f_{rs} E_{rs} \tag{3.7.1}$$

where the $f_{xy}$ elements are relative to the Fock matrix on a CASSCF wavefunction $\Psi_0$

$$f_{xy} = -\left\langle \Psi_0 \left| a_x^+ \left[ \hat{\mathcal{H}}, a_y \right] \right| \Psi_0 \right\rangle + \left\langle \Psi_0 \left| a_x \left[ \hat{\mathcal{H}}, a_y^+ \right] \right| \Psi_0 \right\rangle \tag{3.7.2}$$

Diagonalization of $\hat{\mathcal{H}}^D$ is simplified if we assume that canonical orbitals are used and the inactive part is cast as the well-known diagonal case

$$\hat{\mathcal{H}}_i^D = \sum_{i} \epsilon_i E_{ii} + \sum_{r} \epsilon_r E_{rr} \tag{3.7.3}$$

The simplification arises by the fact that the perturber energies are defined by a simple quantity. For example, we recall the case for the $S^0$ class where the energy of the perturber $E_{ri}E_{sj}\Psi_m^{(0)}$ is $E_m^{(0)} + \epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j$. This obviously cannot be obtained quickly if non-canonical orbitals are used.

A solution resides in a different evolution of the first-order perturbation equation:

$$\left(\hat{\mathcal{H}}^{(0)} - E_m^{(0)}\right) \Psi_m^{(1)} = \left(E_m^{(1)} - \hat{V}\right) \Psi_m^{(0)} \tag{3.7.4}$$

Instead of expressing the $\Psi_m^{(1)}$ as a linear combination of the zero-order wavefunctions

$$\Psi_m^{(1)} = \sum_k c_k \Psi_k^{(0)} \tag{3.7.5}$$

we develop using generic functions $\Phi_l$

$$\Psi_m^{(1)} = \sum_l c_l \Phi_l \tag{3.7.6}$$

A system of linear equations must be solved

$$\sum_l c_l \left[ \left\langle \Phi_k \left| \hat{\mathcal{H}}_0 \right| \Phi_l \right\rangle - E_m^{(0)} \left\langle \Phi_k | \Phi_l \right\rangle \right] = - \left\langle \Phi_k \left| \hat{V} \right| \Psi_m^{(0)} \right\rangle \tag{3.7.7}$$

$$E_m^{(2)} = \left\langle \Psi_m^{(0)} \left| \hat{V} \right| \Psi_m^{(1)} \right\rangle = \sum_l c_l \left\langle \Psi_m^{(0)} \left| \hat{V} \right| \Phi_l \right\rangle \tag{3.7.8}$$

using the internally contracted $E_{wx}E_{yz}\Psi_m^{(0)}$ as $\Phi_l$.

## 3.8 NEVPT and Localization

A first successful integration of localization and NEVPT2 has been deployed to show the feasibility of this approach. An acetone molecule and two water molecules coordinating the CO oxygen have been optimized at 6-311G*/MP2 level. The molecular system has $C_{2v}$ geometrical symmetry, and the $CH_3$ groups have been rotated to feature on-plane hydrogens far from one another. The symmetry has been considered only during geometry optimization, while for the rest of the evaluation the $C_1$ symmetry group has been used to work with non-symmetric localized orbitals.

A minimal atomic basis set ANO-1 with $2s1p$ contraction for heavy atoms, and $1s$ for hydrogen has been used. From this basis, a set of localized orbitals is obtained with the procedure presented in Sec. 2.3. These orbitals are optimized by means of a Super-CI based optimization chain, converging to a localized CASSCF solution. An active space comprising six electrons in five active orbitals ($\sigma$, $\sigma^*$, $\pi$, $\pi^*$ of the acetone carbonyl group, and the $n_y$ lone pair of the carbonyl oxygen) has been defined. The active orbitals are visible in Fig. 3.8.1

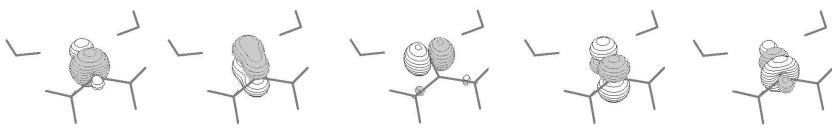

**Figure 3.8.1:** The active orbitals for the evaluation on Acetone + 2 $H_2O$. From left to right, the $\sigma$, $\pi$, $n_y$, $\pi^*$, $\sigma^*$.

These orbitals are used in a Single-State NEVPT2 and a Non-Canonical procedure, leading to the perturbative results presented in the left side of Tab. 3.8.1.

| | Non-canonized | | Canonized | |
|---|---|---|---|---|
| | SS-NEVPT | NC-NEVPT | SS-NEVPT | NC-NEVPT |
| | | Ground State | | |
| (0) | -0.1562219292 | -0.1610160883 | -0.1610160655 | -0.1610160655 |
| (+1) | -0.0060289124 | -0.0063728566 | -0.0063728572 | -0.0063728572 |
| (-1) | -0.0083210019 | -0.0084680922 | -0.0084680918 | -0.0084680918 |
| (+2) | -0.0021959738 | -0.0022262691 | -0.0022262691 | -0.0022262691 |
| (-2) | -0.0012134175 | -0.0011964952 | -0.0011964953 | -0.0011964954 |
| (+1)' | -0.0011298902 | -0.0010747735 | -0.0010747734 | -0.0010747734 |
| (-1)' | -0.0081324857 | -0.0080043857 | -0.0080043841 | -0.0080043842 |
| (0)' | -0.0095463813 | -0.0099518340 | -0.0099518346 | -0.0099518346 |
| Total | -0.1927899918 | -0.1983107946 | -0.1983107710 | -0.1983107711 |
| | | $^1(n \rightarrow \pi^*)$ | | |
| (0) | -0.1569058671 | -0.1616113655 | -0.1616113419 | -0.1616113419 |
| (+1) | -0.0089370064 | -0.0093753612 | -0.0093753625 | -0.0093753624 |
| (-1) | -0.0059930216 | -0.0061061218 | -0.0061061209 | -0.0061061209 |
| (+2) | -0.0009248766 | -0.0009503478 | -0.0009503479 | -0.0009503479 |
| (-2) | -0.0002921853 | -0.0002913366 | -0.0002913366 | -0.0002913366 |
| (+1)' | -0.0043652467 | -0.0046995715 | -0.0046995705 | -0.0046995704 |
| (-1)' | -0.0025446880 | -0.0025151544 | -0.0025151554 | -0.0025151554 |
| (0)' | -0.0117850231 | -0.0123542659 | -0.0123542653 | -0.0123542652 |
| Total | -0.1917479149 | -0.1979035247 | -0.1979035010 | -0.1979035008 |

**Table 3.8.1:** Comparison between the obtained contributions for the non-canonical technique. The "Non-canonized" columns show the values obtained with non-canonical localized orbitals using Partially Contracted Single-State NEVPT (SS-NEVPT) and Non-Canonical NEVPT (NC-NEVPT). The "Canonized" columns report the same evaluations, where a preliminar canonization of the localized orbitals has been performed. Both SS-NEVPT and NC-NEVPT confirm the values obtained in the Non-canonized NC-NEVPT evaluation.

As can be seen, a non-negligible difference is present between the obtained values. Standard single-state NEVPT2 supposes the non-diagonal elements of the Fock matrix as zero. The `dypcNC` program provides the highest non-diagonal element of the core and virtual Fock matrixes

```
Maximum of off-diagonal core Fock matrix is   0.238695436106739
Maximum of off-diagonal virtual Fock matrix is   8.252258688390988E-002
```

which can be appreciated as non zero. This results in an error of the single state treatment.

The Non-Canonical evaluation keeps into account the non diagonality of the Fock matrix, producing the correct value. The correctness of this result is confirmed by canonizing the optimized orbitals right before the perturbative evalua-

tion. In this case, the localization is lost, but the `dypcNC` confirms the diagonality of the Fock matrix

```
Maximum of off-diagonal core Fock matrix is   6.924101073191773E-008
Maximum of off-diagonal virtual Fock matrix is   4.398960914047209E-008
```

and the obtained results (right side of Tab. 3.8.1) confirm the value obtained with the first evaluation.

A plot of the different nature of the orbitals used in the evaluation can be seen in Fig. 3.8.2. The upper part of the figure shows an extract of the localized set of non canonical orbitals, while the lower part plots an extract of the orbitals after canonization. The delocalization of the orbitals is clearly appreciable.

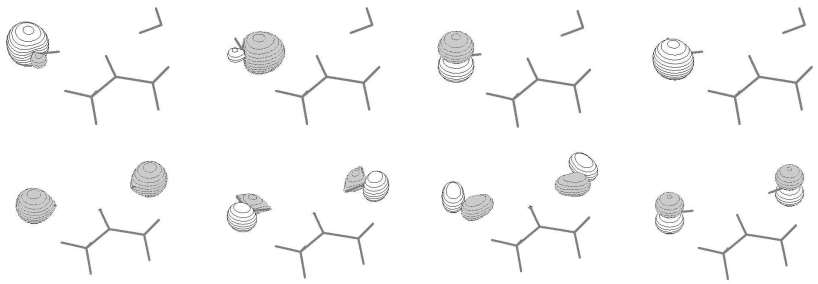

**Figure 3.8.2:** A sample plot of non canonical localized orbitals (upper part of the figure) and canonical delocalized orbitals (lower part) for water orbitals.

Evaluating the energy difference provides consistency of the results. For the reported test, no high difference in the relative energies can be found, as can be seen in Tab. 3.8.2.

| | Ground | Excited | Diff. |
|---|---|---|---|
| CASSCF | -343.573883 | -343.411249 | 4.42 |
| SS-NEVPT | -343.766673 | -343.602997 | 4.45 |
| NC-NEVPT | -343.772194 | -343.609153 | 4.43 |

**Table 3.8.2:** Absolute energies and their differences for the acetone+2 $H_2O$ system. From the comparison of the relative excitation energies, no high differencies between the methods can be found in the analyzed system.

This must be further investigated, given the small basis set used. However, the target of this investigation was to demonstrate the consistency of the NC-NEVPT results.

# Resumé chapitre 4
# Integration de reseau

La puissance computationnelle et l'espace disque ont connu une incroyable augmentation depuis quelques ans. La vitesse d'interconnexion réseau est aussi augmentée, et cet incrément ouvre des nouvelles possibilités pour obtenir puissance computationnelle et espace disque utilisant plusieurs ordinateurs peu employé, à créer une méta-plateforme commune: le réseau à grid. Avec une réseau à grid, il est possible obtenir

- ressources d'infrastructure physique, comme espace disque, temp de calcul, espace RAM, agrégation et archiviation des données, utilisation condivise d'instrumentation.
- ressources logiques, comme expérimentation, analyse des données, modélisation et simulation.
- ressources humaines, comme condivision des compétences et communication

Le résultat est un système hétérogène distribué, qui donne une structure virtuelle à l'utilisateur. L'interaction avec les services est standardisée et uniforme, sans problèmes d'architecture de CPU ou de localisation géographique de l'utilisateur.

Le réseau à grid le plus connu est probablement le web. Le WWW est un réseau pour le stockage des données. Ces données sont consultées avec une interface standard, le navigateur web, avec une protocole à bas niveau, l'HTTP. Cette abstraction est suffisante pour lire et écrire fichiers ou exécuter un script.

En ce moment, trois types d'applications peuvent être décrit pour le réseau

- *Applications à communication réduite*: où le problème est décomposé en morceaux presque indépendant, et le calcul est effectué sur chaque morceau. Seti@Home[199] ou le calculs météorologiques sont des exemples.
- *Applications à pas*: où la procédure est décomposé en pas séquentiels, chacun effectué dans l'environnement de le reseau à grid. Cette approche est utilisée pour partager l'accés à des instruments et l'élaboration des résultats, la distribution du coût d'administration et l'acces aux logiciels locaux.
- *Accés à ressources*: pour partager ressources comme bas des données, instrumentation, espace de stockage.

Pour les processus de chimie quantique, de l'expérience a été faite avec la parallélisation des algorithmes avec PVM et MPI, mais les évaluations *ab ini-*

*tio* sont principalement séquentiels: normalement, on a l'évaluation des integrales, l'optimisation d'un fonction d'onde, et l'application d'autres techniques. L'architecture de référence est donc l'Applications à pas.

Le reseau Metachem est un type de reseau développé dans le cadre d'une action COST de la communauté européenne. L'objectif est l'intégration de logiciels de chimie quantique mis à disposition par plusieurs laboratoires européens.

Pendant ce travail de thèse, deux problèmes pour l'intégration ont été étudié et résolus. Le premier problème est le développement d'un format de fichier commun pour échanger information binaire. Le formats de fichier dejà développés ont été analisés, et leur défects pour l'exécution en grid ont été evalués. Une reseau à grid nécessite de considérations additionnelles: le format ne doit pas être platform-dependent, parce que le reseau peut être construit avec ordinateurs ayant des architectures différentes. Le format idéal pour ce travail est HDF5, mais malheureusement ce format est complexe à utiliser. Autres considérations sont la simplicité d'utilise de la bibliothèque d'acces, exactitude des messages d'erreur pendant l'acces et extensibilité du format.

La recherche effectuée a conduit au développement d'un modèle des données qui a été implémenté dans la bibliothèque Q5Cost. Cette bibliothèque donner l'acces à des entitées chimiques et simplifie l'utilisation à des programmateurs avec une expérience chimique.

Q5Cost a été utilisée pour l'integration des différents logiciels, et ouvre des nouvelles possibilités pour le partage d'information quantistique et l'integration des méthodes *ab initio*. La bibliothèque a aussi une testsuite pour contrôler l'interface et chercher les problèmes dans l'implementation. Le résultat est une bibliothèque très solide et extensible.

Le problèmes principaux qui restent à resoudre dans Q5Cost sont: le stockage des integrals n'est pas encore efficient en terms d'espace occupé, et l'implementation du modèle des données n'est pas complete, mais l'implémentation courante est suffisante pour le premiers développements et implementations.

Le deuxième problème est le management de la séquence d'opérations à effectuer et la conversion de ces instructions dans un format local pour chaque logiciel. Cette conversion doit être effectuée par un logiciel wrapper à l'entrée et à la sortie, et donc est fortement liée au logiciel impliqué. La séquence des opérations sera probablement décrite avec le format de fichier XML, et donc une bibliothèque de lecture et écriture de ce format avec le langage Fortran, F90xml, a été développée. Cette bibliothèque interface libgdome2 (bibliothèque pour le langage C) avec Fortran.

Des expériences préliminaires de communication des données ont été effectuées, et les résultats sont très encourageants.

# Chapter 4

# Grid integration

Computational power is cheap. Storage is also cheap and networking as well. A well known empirical law for the increase of processing power is the Moore's law,[200] which dictates the doubling of computational speed every 18 months. Similar laws have been envised for networking (doubling every 9 months) and storage capacity (doubling every 12 months). For scientific needs, computational power and storage often are the limiting factors for new developments and insights.

The dramatic increase of networking performances opens new possibilities to obtain computational power and storage from unused or underutilized resources. This deploys the fundation of a new paradigm for distributed computing: grid metasystems. Using grids, it is possible to pool

- *Fabric resources*, like disk space, computational power, RAM storage, data aggregation and archiviation, instruments
- *Logical resources*, like experimentation, data analysis, modeling, simulation
- *Human resources*, like competence sharing, administrative burden, communication.

The result is a distributed, heterogeneous metasystem, which provides a virtual structure to the end user. Interaction with services is standardized and uniform, regardless of differences like CPU architecture or geographical location. In this panorama, a grid provides virtually infinite resources ready to use to produce a better science.

The most famous example of a grid is probably the Web. The well-known WWW is a very simple storage grid where documents and informations are geographically distributed. Users can access these informations through a well-known interface (a web browser) which makes use of a low-level uniform protocol (HTTP[201]) in order to access resources. The abstraction provided by the Apache

webserver[202] translates the HTTP requests into fabric level access, like reading or writing a given file, or running a script.

Consortiums like IETF,[203] W3C[204] and OASIS[205] ratified standards in order to define guidelines from low-level issues to high-level protocols and interactions: TCP/IP,[206,207] HTTP, XML,[208] SOAP,[209] WSDL,[210] UDDI,[211] RDF,[212] WSFL[213] are some examples of these standards. Toolkits, frameworks and utilities have been developed, both commercial and opensource to exploit the grid paradigm: Legion,[214] Globus,[215,216] and Unicore[217] are some examples.

At the moment, a restricted set of grid applications can be devised:

- *Low communication applications*: also known as "embarassingly parallel" applications. For these applications, the problem is decomposed in small almost independent pieces (usually called *Working Units*), and the computation is carried over each working unit. This category is well represented by Seti@Home[199] or weather analysis

- *Staged applications*: the procedure is decomposed in serialized steps, each one performed on a given site on the grid. This approach is useful to manage human factors, like competence sharing and administrative burden, license problems, use of private code, security reasons

- *Resource access*: a resource from a service provider is made accessible by request. Classic examples could be databases, instruments, storage farms.

For quantum chemical processes some experience has been made, mainly with parallelization of various algorithms (using PVM,[218] MPI[219] or OpenMP[220]). However, quantum chemistry evaluations are mainly serialized in nature: for *ab initio* calculations, a step involving the evaluation of integrals is followed by other treatments, which in turn provide information for others and so on. This calls for a "Staged application" architecture. This architecture is also enforced by administrative issues, competence sharing and license problems, and finally a current intrinsic complexity in the creation of multiplatform codes.

## 4.1 Metachem grid

The Metachem grid is a project for producing "a meta-laboratory for code integration in *ab initio* methods". The project, still in development, has been carried on under the D23 COST action,[221] with the objective to develop a common grid infrastructure where computational chemists can access several *ab initio* codes produced by different laboratories through a simple and user-friendly interface.

Most of the interested parties developed quantum chemistry codes for Post-SCF evaluations, mainly for internal research. These techniques can be integrated

in a common infrastructure, but with the requisite to leave each code on the platform it was originally designed for, under the responsibility of its owner for maintenance and production. This will distribute the administrative burden and avoid binary code duplication.

Finally, a set of standard quantum chemistry codes (`MOLCAS`, `COLUMBUS` or `DALTON`) must be integrated into the grid to provide commonly used entities, like integrals, overlap, CASSCF orbitals and wavefunctions, and so on.

Various considerations, ranging from technical factors to license restrictions prevent direct changes to existing codes. A more conservative strategy has been considered, at least in the initial deployment of the project (see Fig. 4.1.1): input and output wrappers convert files and workflow logic between the local environment of a given program and the shared grid environment, and viceversa. Input wrappers convert information provided by the grid to low-level input such as Fortran namelists and proprietary binary files. Output wrappers parse the products of the program and create the standard information which will return to the grid.

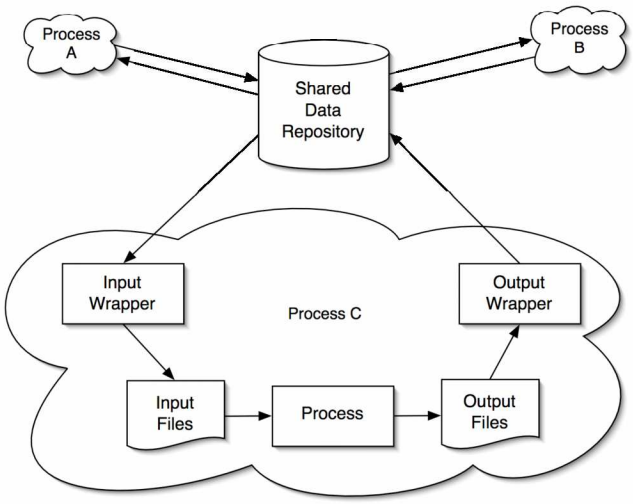


**Figure 4.1.1:** A scheme for the data handling inside a single calculation environment. Data are made available by the data repository, adapted to the local input environment of the process. The result is converted back to the standard format and returned to the repository.

This strategy imposes the development and maintenance of the wrappers by the interested parties. They are supposed to track the changes of their own codes and adapt the wrappers accordingly. For historical reasons, the reference language for the development of these wrappers is Fortran.

The shared environment must also agree on a common file format. Due to the nature of the project, this common file format must be platform independent,

code independent, and easy to debug. Two different kinds of information have been identified in quantum chemistry calculations:

- *small data*, mainly ASCII encoded, like atom labels, geometry, symmetry, basis sets, but also workflow specific information
- *large data*, normally binary encoded, like integrals and expansion coefficients

For the first kind of data, many initiatives are active nowadays. The more extensive is the work of Murray-Rust et al.[222–226] with the development of CML (Chemistry Markup Language) under the eScience project in UK. The choice of XML allows both human readability and machine parsing.

Unfortunately, CML does not provide direct support for quantum chemical entities. After a brief analysis of the problem, an experimental *XML-schema* for quantum chemistry entities, QCML (Quantum Chemistry Markup Language) has been deployed. The parsing of this information in Fortran is possible through a library, named F77/F90xml, specifically developed for the project.

For the second kind of data (large binary data) XML is not a good technology, mainly due to its verbosity. For several reasons, HDF5[227] is considered the best technology for large binary data.

The next step, to be done in the future, is the definition and setup of user defined workflows based on heterogeneous codes, located on different platforms, communicating through the common formats. The actual technology has yet to be chosen, but the choice seems to be oriented towards UNICORE,[217,228,229] together with a common language for describing both workflow and intrinsic differences between work paths.

The final infrastructure is expected to satisfy both grid requirements (fault tolerance, reliability) and human interface requirements (access via web interface or specialized clients). An access point to grid services will be provided by the CINECA supercomputing center. Its role is to coordinate the federated system, where each node provides one or more services to the grid.

## 4.2 Q5Cost library

In the deployment of a common suite of codes, a standard file format is necessary. A first attempt in file standardization has been made in Toulouse with the COST format. This format has been also implemented in most of the Ferrara suite.

The format is made of four files, each one with a different extension:

- `.Mono`, a binary file containing the atomic basis set overlap and the one-electron integrals on the molecular orbital basis set

- `.ijcl`, a binary file containing the two-electron integrals on the molecular orbital basis set
- `.Info`, a cleartext file containing namelists describing various metainformation, like the number of active orbitals, the nuclear repulsion energy, the geometry and so on
- a file called `INPORB`, containing the molecular orbitals. The format is the same provided by the `MOLCAS` program, enriched with additional information needed by the Toulouse suite of programs.

The COST files are written using the standard sequential access provided by `READ` and `WRITE` statements in Fortran. No intention-revealing interfaces are provided for accessing the data. This implementation, although simple to manage, is limited in a lot of points:

- the standard sequential access provided by Fortran imposes to agree on data size, data ordering, and chunking. It does not allow random access, making difficult to update already written information, or to posticipate the writing operation of some entity. It complicates reorganization or expansion of the stored information set
- files are not platform independent. A file written on an Intel x86 machine is unreadable by an IBM SP4 supercomputer, and viceversa. This limits the output sharing
- it is difficult to access the stored information in a clear representation. This is a major issue when debugging computational chemistry codes. At the moment the access is through ad-hoc visualizers or hexadecimal editors
- the file is difficult to split, merge, compare, or describe with metainformation
- the API for accessing the file format is limited in features, allowing no basic consistency checks or a clean interface. Documentation is written as code comments. This makes difficult to direcly access the information.

Q5Cost is a library devoted to solve these problems. Based on the platform-independent file format HDF5, Q5Cost allows the user to produce binary files through a well documented interface. The library also takes care of providing useful information during debug, explicitly stating inconsistent data or wrong usage of the routines.

### HDF5 file format

As defined on the website,[227] "HDF5 is a general purpose library and file format for storing scientific data". HDF5 targets the data management needs of scientists and engineers working in high performance, data intensive computing environments. This is achieved by providing an efficient library and file format, specifically tuned to read and write binary data efficiently.

A HDF5 file appears to the user as a single file containing a structured graph of entities. This structure is conceptually very similar to a UNIX filesystem, and is defined by means of three entities: *Groups*, similar to directories, *Datasets*, similar to files and *Attributes*, similar to metainformation such as permissions and owner. Data are stored into Datasets and Attributes, and can be organized with Groups. The HDF5 library provides an interface to handle these entities.

The most striking feature of HDF5 is the platform independence of the files. A file written on a 32 bit Little Endian architecture (like, for example, an Intel x86 machine) can be transferred and read transparently on a 64 bit Big Endian machine (like, for example, the IBM SP4) or a 64 bit Little Endian machine (like an IA64 machine). This is of critical importance in grid environments, where heterogeneity is the rule, rather than the exception.

Another important feature of HDF5 is the suite of utilities, which allows the user to find differences between two HDF5 files, obtain a textual dump of Groups, Datasets and Attributes layout (resembling the "ls" program in UNIX environment) or a more detailed view presenting also the numerical data. This simplifies the access to this information, which is critical in the debug phase.

The HDF5 library is made of a directly accessible C layer. Alternative APIs are available for C++, Fortran 90, and Java. The library is actively developed, is free and released as open source software.

### Q5Cost file format and library

The Q5Cost library is implemented on top of the HDF5 library. It provides read and write access to HDF5 files with a specifically designed high-level interface for quantum chemistry developers.

The rationale is to provide a Fortran interface based on well known chemical entities, rather than groups or datasets like in the HDF5 interface. HDF5 takes care of low-level management of the file, and Q5Cost provides the high-level API to store and retrieve chemical entities.

An extensible data model has been rationalized thanks to the knowledge obtained by the involved researchers. Some firm points have been discovered.

The first point is that many different types of simple and small data must be handled. Some examples are nuclear energy, molecular orbital labels, molecular

symmetry and so on. We will refer to these data as *metadata*, in order to distinguish them from the real large information on the chemical system, the integral values.

Metadata represents well-known chemical entities and belong to three generic data classes: *Scalars*, *Vectors* and *Matrixes*. For example, the nuclear repulsion energy is a floating point scalar, molecular orbitals can be represented as a (N,M) floating point matrix, the associated orbital energies is a floating point vector, the molecular orbital labels is a vector of strings and so on. The library should provide an interface for accessing these data, both as generic or specialized entities.

A second point is the fact that in quantum chemistry large sparse matrixes with an arbitrary number of indexes (rank-$n$ arrays) are very common data structures. These data usually scale aggressively with the system size, and they are normally accessed with a chunked sequential approach. This is the case of entities like two-electron integrals, or atomic orbitals overlap, but also much other more application-specific information, like the four-particles density matrix.

The sparse nature of these matrixes encourages to store only non-zero elements, each one associated to $n$ indexes in the case of a rank-$n$ array. This representation of the data, although not always efficient in term of space occupation, is well known by the interested parties, easy to debug and already integrated in the current codebase both for memory representation and file storage. The library structure allows future expansions of the low-level storage format, in order to reduce the storage needs, but in the first development phase this issue has not been investigated, and the current format is useful for the direct debug of the file.

These large data arrays share common features:

- they usually are integrals, whose evaluation involves one operator and a given number of functions. These functions are referred by the indexes of the matrix. For example, two-electron integrals on the molecular orbital basis are stored as a rank-4 array with indexes referring to the molecular orbitals. In the case of atomic basis set overlap integrals, the indexes refer to the atomic basis set

- the rank of the array depends on the physical meaning of the integrals. Atomic basis set overlap is described by two indexes, and can be stored as a rank-2 array. Two-electron integrals have four indexes imposing a rank-4 array and the four-particles density matrix has eight indexes, requiring a rank-8 array

- additional information is needed to describe the involved operator (for example its symmetry) and the entities the indexes refer to.

All this information can be classified under a generalized *Property* concept, which is described by the matrix rank and the definition of the involved operator and functions. Integrals are stored by means of a dedicated abstraction, named *CompactMatrix*, enriched with metadata to define the Property, like its name, rank, symmetry and real/complex nature.

Some properties are well-known chemical entities, and chemists are used to refer to them by name. A specific library access has been provided to handle these entities, easing the need to specify well-known information about them.

A further step is to recognize containment relationships and simple consistency patterns between the involved entities. A graphical representation can be seen in Fig. 4.2.1.

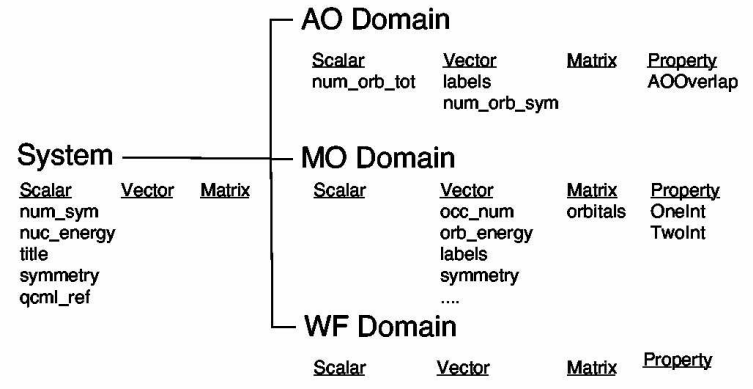


**Figure 4.2.1:** A graphical representation of the containment relationship between the System and the Domains. Specific metainformation is also presented, classified under the appropriate environment.

The first high-level logical container is the *System* object. This object represent a molecular system as described by its structural informations, such as symmetry and atomic coordinates, and also contains all the metadata that are invariant at this level, like for example nuclear repulsion energy, or a title. Multiple Systems make possible to handle different molecular geometries.

A System can contain several *Domains*. The role of a Domain is to group together Properties whose indexes conceptually refer to the same entity. Three Domains have been recognized as fundamental: *AO* for Atomic Orbital, *MO* for Molecular Orbitals and *WF* for Wavefunction. Each Domain can contain an arbitrary number of Properties, and additional metadata represented as specialization of Scalar, Vector and Matrix entities.

The AO Domain holds Properties whose indexes refer to the atomic basis set funtions: overlap, one-electron and two-electron integrals on the atomic basis set

are some examples, in addition to the generic Property. The associated metadata hold information about the atomic orbitals, like their number, labels and symmetry.

The MO Domain holds Properties referring to molecular orbitals: one-electron and two-electron integrals on the MO basis set are examples, in addition to the generic Property. The descriptive metadata refer to the molecular orbitals, their number, labels and symmetry, the AO basis they were derived from, orbital energies, classification and occupation numbers.

The WF Domain holds Properties referring to the electronic states. At the moment a complete definition of this Domain is not available and not subjected to major research and development, given its non-critical nature for the first deployment of the library.

Each Domain can be defined under different occurrences, by means of a textual identifier (*tag*) chosen by the user, with a default value if no tag is provided. The aim is to provide storage of multiple entries for each Domain, like in case of multiple molecular orbitals in the MO Domain, or multiple basis sets in the AO Domain.

### Library structure

The library is written in Fortran 95 and is made of different modules, each one providing different facilities. The most important modules are

- *Q5Cost*: defines the high-level API, providing subroutines designed to be accessible from the final programmers of the laboratories involved in the project
- *Q5Core*: provides a wrapping facility for HDF5 routines, in order to perform additional useful services like reference counting and debugging. Also provides simplified routines to perform frequently used low-level tasks
- *Q5Error*: provides facilities for high-level debugging of the library and client codes. This module implements a ring buffer for error messages, different logging levels, generic reference counting for catching leaks of HDF5 references, and a subroutine call stack trace.

Subroutines for each module have been namespaced with an appropriate prefix, and names have been chosen to provide an explicit and intention-revealing interface to the entities described in the previous section.

Although Fortran 95 does not allow object oriented (OO) programming, some OO concepts have been used in the development of the library, keeping into account the procedural programming background of the final developers. The state

is preserved into the HDF5 file, and subroutines refer to the file directly through the HDF5 identifier, a more easy concept for Fortran programmers used to unit descriptors.

### The Q5Cost module

This module is the main reference for the final users. It provides subroutines to read and write HDF5 files in the Q5Cost format with a high-level abstraction. The users can deal with high-level concepts without worrying about low-level implementation details. If a finer access is needed to the underlying HDF5 file, the Q5Core module provides this access in a simpler way with respect to raw HDF5 routines.

The Q5Cost module is made up of routines belonging to the following classes:

- *File*: create, open, close the file and write/read root metainformation, like creation time, access time and file version
- *System*: create or check the existence of the System and set/get its metainformation
- *AO*: create or check the existence of the AO Domain and set/get its metainformation
- *AOOverlap*: create, read and write data for the atomic basis set overlap Property
- *MO*: create or check the existence of the MO Domain and set/get its metainformation
- *MOOneInt*: create, read and write data for the one-electron integrals in molecular orbitals basis
- *MOTwoInt*: create, read and write data for the two-electron integrals in molecular orbitals basis
- *WF*: create or check the existence of the WF Domain and set/get its metainformation. Still experimental and mostly unimplemented.

Additional routines are provided for generic access to the Property class, allowing to manage generic property data. Subroutines for AOOverlap, MOOneInt and MOTwoInt delegate calls to these Property routines, automatically passing well-known parameters about the involved property.

Q5Cost module routines provide a context-based access to chemical entities. This access is transformed into a path-based access, creating an appropriate layout

for HDF5 groups, datasets and attributes, and writing the user-provided data on the file. Some data are automatically written by the library, like the creation or access time and the Q5Cost library version.

One important aspect of this format is that the user is not forced to produce information for all quantities. The user can store only those quantities that are actually available. Constraint checks are however mandatory in order to assure basic file consistency. For example, no MO Domain can be created if a System and an AO Domain have not been provided, in order to guarantee the presence of fundamental data, like the number of symmetry species and the number of basis functions for each symmetry species.

### The Q5Core module

The Q5Core module is a low-level module designed to provide wrapping facilities between the Q5Cost library and the HDF5 library. At the moment it is focused at providing additional debug information, reference counting for HDF5 objects, additional low-level API to simplify common tasks.

This module also provides path-based creation of Scalar, Vector and Matrix entities (in contrast with the context-based approach of the Q5Cost module, which focuses on chemical concepts rather than HDF5 paths) and also routines to ease handling of the Property index/value data, relative to the CompactMatrix class. Most of the Property routines accessing index/value data delegate to Compact-Matrix subroutines.

Q5Core module could also be useful to guarantee transparency toward a low-level format migration from HDF5 to other storage formats. This could allow the Q5Cost module to remain unchanged and unaware of the final low-level format, but in the current deployment this was not a priority. Final users are not expected to access Q5Core module routines, except for highly experimental tasks.

### The Q5Error module

The Q5Error module provides subroutines to monitor the behavior of the library and the client code. A ring buffer is provided to keep track of error messages generated by the library. A verbosity level can be set from totally silent to highly verbose, where each subroutine call and return are reported into the buffer. Also, a stack for backtracing has been implemented to keep track of the call tree. The tree is printed out when an error occurs and error reporting is requested.

Different specific error codes have been devised to report anomalous behavior to the client code or to the library itself. Error codes are defined as numeric parameters and report situations ranging from invalid parameters to non-existence

of some information into the file. The presence of an error condition is reported to the client code through the last parameter of each subroutine.

#### Additional details: testsuite and documentation

A dedicated module for testsuite deployment and a set of tests have been implemented to stress the library through a high number of well-known critical situations, evaluating the correctness of the response. At the moment, more than 300 tests are implemented, covering the most common usage patterns. Reference counting is checked to prevent leaks of HDF5 references. The testsuite provides an effective watchdog for debugging, refactoring, and bugfixing.

Library documentation is embedded into the Fortran code as comments, using a custom markup system. A simple parser, written in the Python[230] programming language, extracts the documentation producing HTML files.

## 4.3 Q5Cost examples

Various programs have been implemented using Q5Cost: `q5dump`, `cas2hdf`, `hdf2cost` and, of course, the testsuite presented in the previous section.

`q5dump` is a program designed and implemented by Antonio Monari. It allows access to the information inside a Q5Cost file in a clean and explicit way. It resembles the `h5dump` utility from the HDF5 library, but `q5dump` presents information in chemical speech, rather than low-level HDF5 concepts like groups and datasets.

A typical output of the program is reported

```
**********************************
*          Q5Costdump            *
*                                *
*     a tool for analysis of     *
*           Q5Cost files         *
*                                *
**********************************

Creation time 2005/03/30  15.28.05

SYSTEM ATTRIBUTES
Title: LiHs33
Order of the symmetry group  4
Nuclear Repulsion (Core Energy)  0.995024875621900007

Groups present  2
ao 1
tag-default
```

```
mo 1
tag-default
-----------------------------------------------------

Properties of AO group <default>

Number of Orbitals 33
Orbital in Symm. Classes   17   7   7   2
AO is empty
-----------------------------------------------------

Properties of MO group <default>
 AO REF: <default>

Number of Orbitals 33
Orbital in Symm. Classes   17   7   7   2

dipolex is present
dipolez is present
oneint is present
twoint is present
orbitals is empty

-----------------------------------------------------
```

which can be compared against the output produced by `h5ls`

```
/ao                      Group
/ao/tag-default          Group
/mo                      Group
/mo/tag-default          Group
/mo/tag-default/Property-dipolex Group
/mo/tag-default/Property-dipolex/CompactMatrix-data Group
/mo/tag-default/Property-dipolex/CompactMatrix-data/index Dataset {2, 171/Inf}
/mo/tag-default/Property-dipolex/CompactMatrix-data/value Dataset {171/Inf}
/mo/tag-default/Property-dipolez Group
/mo/tag-default/Property-dipolez/CompactMatrix-data Group
/mo/tag-default/Property-dipolez/CompactMatrix-data/index Dataset {2, 181/Inf}
/mo/tag-default/Property-dipolez/CompactMatrix-data/value Dataset {181/Inf}
/mo/tag-default/Property-oneint Group
/mo/tag-default/Property-oneint/CompactMatrix-data Group
/mo/tag-default/Property-oneint/CompactMatrix-data/index Dataset {2, 182/Inf}
/mo/tag-default/Property-oneint/CompactMatrix-data/value Dataset {182/Inf}
/mo/tag-default/Property-twoint Group
/mo/tag-default/Property-twoint/CompactMatrix-data Group
/mo/tag-default/Property-twoint/CompactMatrix-data/index Dataset {4, 39790/Inf}
/mo/tag-default/Property-twoint/CompactMatrix-data/value Dataset {39790/Inf}
/mo/tag-default/orbitals Dataset {0/Inf, 0/Inf}
```

`cas2hdf` and `hdf2cost` are converter utilities, designed to stress a first deployment of the library in a production environment. In Fig. 4.3.1, a representation of the final intended procedure is depicted.

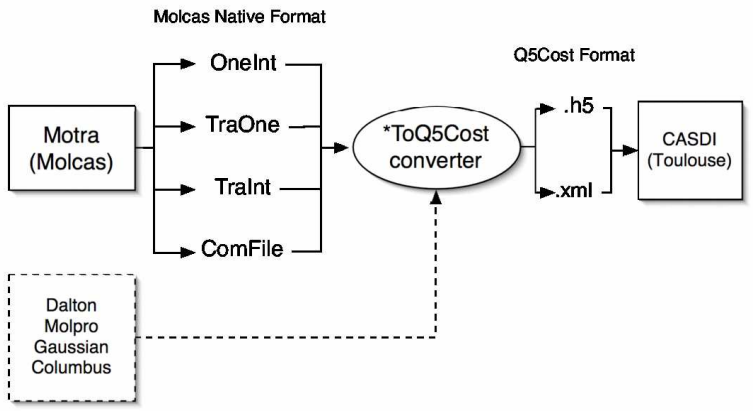


**Figure 4.3.1:** One of the possible final code layouts, involving direct changes in the code. An integral producer, such as `MOLCAS`, provides proprietary files. These files are fed to a converter, which creates the new Q5Cost format. The program `CASDI` from Toulouse directly reads these files. This infrastructure at the moment is not possible, because the current `CASDI` implementation reads the old COST format.

Output integrals are produced by a commercial package, `MOLCAS` in the example, but converters could be developed for other programs, like `DALTON`, `COLUMBUS` and so on. These integrals are fed into a converter to produce the Q5Cost file format, made of the HDF5 file and an XML (QCML) file providing additional information. The Q5Cost format will be read by CASDI, a program from the Toulouse chain of programs. This approach assumes that `CASDI` is directly interfaced with the Q5Cost format.

A more conservative way to handle this interface is represented in Fig. 4.3.2.

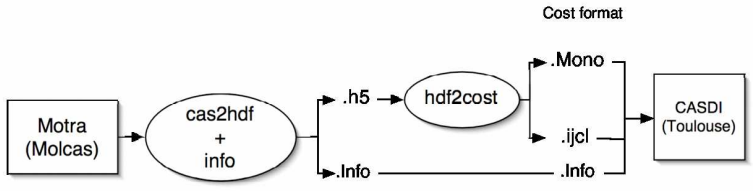


**Figure 4.3.2:** A conservative solution. The `cas2hdf` program produces the Q5Cost file, which is converted to the old COST format with `hdf2cost`. The `.Info` file is used as a temporary replacement of the XML file, and is provided by the `info` program, directly interfaced with the `MOLCAS` suite. This solution does not need changes in the CASDI code, and is therefore preferred in the initial deployment.

The `cas2hdf` program directly accesses the `MOLCAS` files, producing the Q5Cost file. The `.Info` file is used as a temporary replacement of the XML file, and it is provided by the `info` program. The resulting Q5Cost file is read by `hdf2cost` and converted to the old COST format, which can be directly read by CASDI. This solution does not need changes in the `CASDI` code, and is therefore preferred in the initial deployment.

The deployment of these tools made possible to obtain a successful interface between `MOLCAS` and `CASDI`. Some successful experiment has been also performed with `DALTON`.

An experiment of direct integration has been performed with the `Full-CI`[231,232] program developed in the Professor Bendazzoli research group at Bologna University. The `Full-CI` code is now able to read Q5Cost files, and can be interfaced with both `MOLCAS` and `DALTON`.

## 4.4 F77/F90xml library

The Extensible Markup Language (XML) file format plays a central role in the modern internet infrastructure. XML is derived by the more general format Standard Generalized Markup Language (SGML), and is visually very similar to HTML (the file format for web pages) although HTML itself is not a XML document.

The official definition provided by the W3 Consortium[204] states that the "*Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere*".

The main advantages of this format are:

- internet standard, as accepted by the W3 Consortium
- platform independent
- clean and terse human readable textual format
- very practical and extensible
- can be checked for syntactical and semantical correctness
- allows the description of data and their meaning
- can be commented inline
- already existent libraries for accessing the document

An example of an XML document is given below

```
<?xml version="1.0"?>
<book ISBN="0140390839">
 <!-- a comment -->
 <author>Mark Twain</author>
 <title>Tom Sawyer</title>
 <chapter>
  <title>First Chapter</title>
  This is the content of the first chapter.
 </chapter>
 <emptypage />
 <chapter>
  <title>Second Chapter</title>
  This is the content of the second chapter.
 </chapter>
</book>
```

An XML document holds *markup* and *content*. The most common type of markup is the *Element*, delimited by angle brackets and normally present in pairs of begin tag (for example "`author`") and end tag (the same tag name, but prepended by a slash "/" symbol). These elements wrap content and other markups in a nested parent/child relationship, defining a tree-like structure. If an Element has no content, an alternative abbreviated syntax can be used to represent the begin/end as a single tag. The "`emptypage`" Element is an example of such syntax.

Another kind of markup is the *Attribute*. Attributes are name/value pairs defined in the context of an Element. An example of this markup is the "`ISBN`" entry for the "`book`" Element.

Finally, a comment markup can be seen in the provided example. Other kinds of markups, such as processing instruction, entity references or CDATA sections, will not be analyzed in this thesis.

Two constraints define the correctness of an XML document:

- *well-formedness*
- *validity*

An XML document is well-formed if it complies with the XML specification. Using square brakets instead of angular brakets is an example of non well-formedness. Improper nesting of elements such as

```
 <author>Mark<Title>Twain</author>Tom Sawyer</title>
```

is another example. XML Parsers must report an error when a non well-formed document is provided.

Validity is related to well-formed documents that comply with a logical meaning, related to the chosen data model. For example, the document

```
<?xml version="1.0"?>
<title>
 <book>
  Tom Sawyer
  <author ISBN="0140390839">
   <chapter>
    This is the content of the first chapter.
   </chapter>
  </author>
 </book>
</title>
```

is a well-formed document, but has poor or no meaning in the analyzed data model. Validity depends on the context and is defined by Document Type Definitions, or XML-Schemas, which will not be detailed in this thesis. Pure parsing cannot check for validity unless DTD or schemas are provided.

Accessing XML documents can be performed through two standardized interfaces:

- *SAX*, Simple API for XML
- *DOM*, Document Object Model

SAX is an event-driven parser. While the document is read (from a file, or from a network socket) the parser calls specific callback routines (handlers) to process the current Element (start or stop markups), or Attribute, or content (TextNode). Each handler is normally defined by the developer of the client code, and expresses the contextual behavior related in finding a certain Element, or TextNode.

SAX has some important advantages: first, parsing can be performed on a partial document. This is very important if the document is very large, or if the parsing must begin as soon as the first data are available. This can produce a more pleasant experience of responsiveness to the user if, for example, the document is downloaded from a slow network connection.

Second, the SAX parsing has a very small memory footprint. The document does not need to be held in memory. Nodes can be parsed and discarded after the

event has been dispatched to the appropriate handler. This is very important if the document is very large.

SAX has also disadvantages: is a readonly parser (it can only read documents, but cannot build and write an XML tree), is stateless, is purely sequential, and finally is rather difficult to use.

The other available parser is DOM. DOM is an interface to an Object Oriented description of the XML document. The main advantage of DOM is represented by the ability to handle the document in ways not provided by SAX: DOM interface directly exposes the tree data abstraction in terms of Nodes (*flat interface*) or specialization of the Node concept, like Element, Attribute, Document, TextNode, and so on (*specialized interface*). The document is parsed as a whole and a tree representation is built in memory. This representation can be parsed, read and altered with a random access approach, and finally written back to disk or to a network connection. DOM interface is also simpler to use than SAX.

The main drawback of DOM is the need to hold a complete representation of the document in memory. For this reason, DOM parsing is generally not applicable on partial or very large documents.

Two interface levels are available, and a third is almost ratified. DOM Level 1 provides the basic handling of the tree. DOM Level 2 and 3 provide additional advanced features like namespacing and events. Only a restricted set of Level 2 features are needed for the project.

### XML and Fortran

Calculations in scientific environments such as physics, chemistry, astronomy and so on are still bound to the Fortran language. Other high-level languages such as C, C++, Java, Python are of difficult deployment. Techological problems (compatibility, library availability, efficiency) human factors and historical reasons (learn new abstractions and patterns, rewrite the legacy code) prevent a switch to other languages.

Of course, this is a strong limiting factor in keeping the pace with the rapid evolution of network and computational resources. Using these new technologies often requires dedicated middleware, specifically designed to simplify access by wrapping low-level functionality with a high-level interface. XML handling is an example of the reported situation: an XML document is a plain text file, but its complex, extensible structure goes beyond its pure textual nature. SAX and DOM parsers are high-level frontends to XML document management, taking care of low-level issues and presenting a new abstraction to the programmer.

The lack of a Fortran library to handle XML can be a problem when the deployment of new standards is needed and the handling of these standards must

be performed by Fortran programmers, like in the case of the AbiGrid project. Currently available libraries for XML access in Fortran are "xml"[233] and "xml-fortran",[234] not DOM compliant and written in pure Fortran, and "xmlf90"[235] which supports SAX and DOM Level 1 and is written in the F subset of Fortran.

Among the problems of these solutions are:

- non DOM Level 2 compliant (no namespaces support)
- rather difficult to use and extend
- read only
- string handling is very rigid
- bound to F90 compilers

The first five points are indeed critical. Deploying advanced features like DOM or XPath (a standardized method to access Nodes using a syntax similar to a filesystem path) is very difficult with a pure Fortran approach. Fortran has limits in strings, file handling and object management.

The latter point is more controversial. Many commercial Fortran 90 compilers exists. Some of them are available at no cost, but this policy could change in the future. With the release of free Fortran 90 compilers like g95[236] and gfortran,[237] the latter point is no longer an issue, but was a prerequisite in the initial development of F77/F90xml.

## 4.5 F77/F90xml implementation

The F77/F90xml library[238] is a C library designed to provide a Fortran interface to libgdome2,[239] an opensource library (also developed using C language) part of the GNOME[240] project. Most of the design is implemented to give access to a DOM level 2 interface,[241] with the exclusion of events, which are not needed in the target environment. XPath support is available, although no testing has been performed and therefore must be considered experimental.

At the time of writing, F77/F90xml library depends on libgdome2-0.8.0, glib-1.2.10,[242] libxml2-2.5.11.[243] Building of the library also depends on Python 2.3[230] or above. The library has been successfully compiled on Intel/AMD Linux platform, with gcc and Intel Fortran Compiler, and IBM SP4 with the xlf compiler suite. It is released under the terms of the LGPL.[244]

F77/F90xml has been designed with Fortran 77 backward compatibility in mind. The library provides two interfaces: a Fortran 90/95 interface, similar to

DOM Level 2 in order to reduce the need for specialized documentation, and a Fortran 77 interface, using specialized subroutines named *multiplexers*.

The Fortran 77 interface is particularly complex and error prone. An experimental preprocessor has been deployed to translate pseudo-F90 code into F77 code, but Fortran 77 is very old and the choice for new code should be Fortran 95. Support for this language was an initial request from the interested parties, and provides full freedom even to those projects that still work in a pure Fortran 77 environment for their maintenance.

When the library development started, no stable and free Fortran 90/95 compiler existed, and most of the potentially interested parties had the requisite to compile the code using the GNU F77 compiler. This situation changed with the release of g95 and gfortran, however the Fortran 77 interface is directly obtained from the implementation, simplifying rather than making more difficult the development of the library.

In the following, we will refer to the Fortran 90 interface with the term "F90xml", and with "F77xml" to the specific Fortran 77 interface. Finally, the term "F77/F90xml" refers to the library as a whole.

### The Fortran 90/95 interface

The Fortran 90/95 interface provides a clean and simple access. All gdome2 functions are mapped to Fortran subroutines, collected together into a `MODULE`. A simple code example is provided

```
INTEGER :: first, last, elem, err

Comment <... get elem by some other call ...>

CALL f90xml_el_firstChild(first,elem,err)
CALL f90xml_el_lastChild(last,elem,err)
```

which can be compared with the equivalent C code

```
GdomeElement *elem;
GdomeNode *first, *last;
GdomeException exc;

/* <... get elem by some other call ...> */

first = gdome_el_firstChild(elem, exc);
last = gdome_el_lastChild(elem, exc);
```

This comparison presents the main featured differences of the F90xml usage with respect to gdome2. Apart of the routine names, where the standard prefix "`gdome_`" is replaced with "`f90xml_`", two major differences exist.

The first difference is bound to the data reference handling. The C interface provided by gdome2 is structured in Object Oriented style. Routines handle pointers to structures, dynamically allocated inside the library.

To prevent issues relative to the handling of memory pointers between C and Fortran, the F77/F90xml library provides a simple mapping between an integer value and a memory pointer (for example to `GdomeNode`, `GdomeDOMString`, `GdomeDomImplementation`, and so on). Fortran client code always handle these integers, hereafter named *codes*, univocally identifying a particular gdome pointer. The F77/F90xml library internally implements a cache to store these pointers and the corresponding code, providing an internal facility to obtain codes from memory pointers and viceversa. The current implementation of the library uses a simple linked lists to store the mapping, and this correspondence is kept until the gdome2 object is completely deallocated. Substitution of the linked list with a more efficient hash table can be implemented transparently.

The second major difference is the arguments layout: in F90xml, the first argument of the subroutine is the gdome2 returned value, therefore the Fortran 90 interface declares this argument as `INTENT(OUT)`. The subsequent arguments are the requested gdome2 arguments, marked as `INTENT(IN)` in the Fortran 90 interface, with the exception of the last one, holding the returning error condition and therefore marked as `INTENT(OUT)`.

If the gdome2 function returns `void` (no value), the corresponding F90xml subroutine accepts an integer as a first argument. On return, this value is set to the standard numeric parameter `NullCode`, which evaluates to zero.

### String handling

For most of the time, strings in F77/F90xml are handled as codes representing `DOMString` objects: when an element name is requested, the F77/F90xml subroutine returns a code referencing a `DOMString` object. Setting an element name or TextNode data requires passing a `DOMString` code. A set of routines has been provided to simplify the handling of these objects:

- `f90xml_str_mkref`: creates a new `DOMString` from a Fortran `CHARACTER` string. Accepts the Fortran string and returns a code referencing the newly created `DOMString` object.

- `f90xml_str_length`: accepts the code of the `DOMString` and returns the length of the string. This routine can be useful to know in advance how many bytes are needed to read from the `DOMString`, and act accordingly. A classical example could be a `DOMString` containing 300 bytes and the Fortran code has a `CHARACTER(LEN=100)` available.

- `f90xml_str_toFortran`: extracts Fortran character data from an existing `DOMString` object. This routine accepts the code of the `DOMString`, a `CHARACTER(LEN=*)` variable, an `INTEGER` offset and returns a `LOGICAL` value. The `DOMString` data will be extracted starting at the position provided by the zero-based offset. No more than the length of the `CHARACTER` variable will be extracted from the `DOMString`. The returned `LOGICAL` value is set to `.TRUE.` if the `DOMString` has been extracted up to the last character, otherwise `.FALSE.`. Using the returned logical value and working with appropriate offsets, it is possible to read long strings in a chunked fashion, regardless of the effective dimensions of the `DOMString` and `CHARACTER` variable.
- `f90xml_str_print`: Prints the `DOMString` to standard output. Returns `void`.
- `f90xml_str_equal`: Performs a comparison between a `DOMString` referenced by its code and a Fortran `CHARACTER` string. Returns `.TRUE.` if the strings are equal, otherwise `.FALSE.`.
- `f90xml_str_unref`: deletes the `DOMString`. Returns `void`.

### Errors

The F77/F90xml library returns the error status *via* the last `INTEGER` argument. The returned value depends on the kind of error, and a list of `PARAMETER`s account for all the available situations. The `ERR_NO_ERROR` value, which evaluates zero, is returned if no error occurs. The library checks for various error conditions, such as

- a passed code is not of the expected type, after inquiry into the cache (e.g. passing the code corresponding to `DOMString` to `f90xml_el_firstChild`)
- a code does not reference to any cached object
- a `NullCode` is passed to a routine unable to handle it
- internal errors of the gdome2 library

### Library architecture and Fortran 77 interface

Standard Fortran 77 expects names limited to 6 characters, although at our knowledge no recent Fortran 77 compiler imposes this strict limit. Deploying the complete DOM interface with a one-to-one mapping in such limited namespace would result in name collisions and routine names with no meaning.

To face this issue, a few multipurpose C functions have been created, named *multiplexers*. The role of each multiplexer is to create a many-to-one correspondence between a subset of the gdome2 routines accepting the same number and type of arguments and the single generic multiplexer function. Multiplexers give access to the complete interface with a reduced namespace footprint.

Each C multiplexer is directly mapped by the library to a F77 `SUBROUTINE`, in a one-to-one relationship. Code between the Fortran frontend and the corresponding C multiplexer is needed to interface these two languages and their different standards for string handling, memory management, routine name mangling and parameters.

The multiplexers are the core of the library. When a multiplexer is called, it dispatches (demultiplexes) the call to the appropriate function of the subset it describes. In turn, this function performs the actual call to the gdome2 routine. To select which routine to call, a string containing the name of the routine is passed as an argument. Internally, this information is used to invoke the correct function. Fig. 4.5.1 represents a high-level view of a single multiplexer.

Each multiplexer routine and its Fortran interface have a standardized name, bound to the number and type of arguments it accepts.

A schematic classification of the routines has been devised to describe with a short notation the number and type of accepted parameters, and the returned value. Routines with the same type signature are handled by the same multiplexer. The core implementation resorts on the equality of signatures to collect pointers to functions.

In the following, a textual representation of the signature will be written in the compact form `(return|arguments)`. For each kind of argument, a letter has been assigned: "`c`" for code, "`b`" for boolean (logical), "`e`" for error, "`s`" for string, "`u`" for unsigned integer. Also, the "`v`" is used with subroutines that return no value (are `void` in C syntax).

As an example, all the gdome2 functions given below

```
GdomeNode* gdome_el_firstChild(GdomeElement *self, GdomeException *exc);

GdomeElement* gdome_doc_documentElement(GdomeDocument *self, GdomeException *exc);

GdomeNode* gdome_el_lastChild(GdomeElement *self, GdomeException *exc);
```

accept a code and an error and return a code. The signature of these functions therefore is `(c|ce)` and they are handled by the same multiplexer. The associated Fortran entry point is the subroutine `xp3t1`, where "`x`" is a standard prefix, "`p3`" (literally "parameters 3") denotes the number of arguments involved in the signature (two codes and the error) and "`t1`" (literally, "type 1") is bound to the

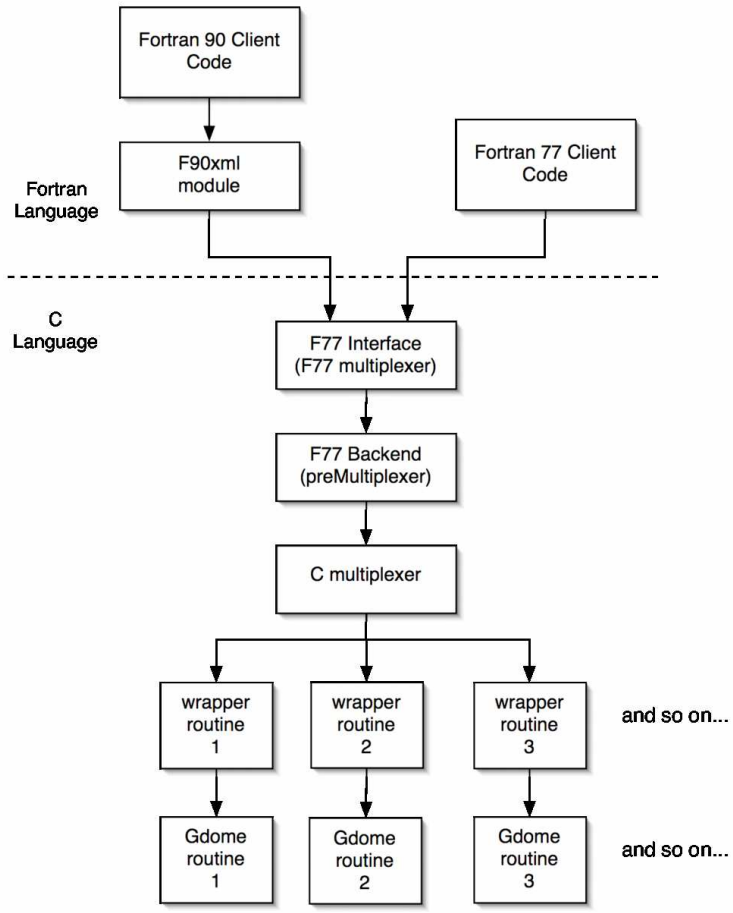


**Figure 4.5.1:** A graphical scheme representing the F77/F90xml library structure. Each multiplexer groups a high number of gdome2 subroutines.

type of the arguments. The type is needed to distinguish routines that accept 3 parameters of a different type. For example, the following routine

```
GdomeBoolean gdome_el_hasChildNodes(GdomeElement *self, GdomeException *exc);
```

has signature `(b|ce)` and is handled by a multiplexer with the same `p` value but different `t` value: `xp3t4`

An exception is represented by routines returning `void`, `(v|ce)`. These routines are handled as `(c|ce)` routines, accepting a dummy return argument which is set to the special value `NullCode`. In other terms, both `(c|ce)` and `(v|ce)` routines are managed by the same multiplexer, the `xp3t1` multiplexer. It must be pointed out that no matching scheme exists to obtain the type number from the letter sequence in the signature.

The `xp3t1` multiplexer by itself maps and gives access to approximatively 240 gdome2 functions. The complete DOM interface (more than 450 functions) is

mapped by 21 multiplexers. As already said, the gdome routine is chosen by the means of a `CHARACTER` argument which is passed to the multiplexer. By convention, this argument is passed between the return value and the first `INTENT(IN)` argument, therefore always in the second position.

As a final result, the Fortran 77 interface to the multiplexer is represented in the example below

```
      CHARACTER*128 fnName
      INTEGER first, last , elem, err

Comment <... get elem by some other call ...>

      fnName='el_firstChild'
      CALL xp3t1(first,fnName,elem,err)

      fnName='el_lastChild'
      CALL xp3t1(last,fnName,elem,err)
```

As can be seen from the example, the function name is case sensitive and is the exact copy of the gdome2 function, stripped of the "`gdome_`" prefix. The Fortran 90 module provides long named subroutines performing the mapping between each subroutine and the corresponding multiplexer.

The library architecture here presented was chosen for two reasons:

- reduce the development cost, providing automatized creation of the library
- keep potential Fortran 77 compatibility.

A large part of the library is developed in XML. An XML file contains all the informations to create the binding routines called from the C multiplexers. The file is parsed by a Python script which collects the needed informations, and deploys the C and Fortran code.

The logic involved is to define templates for the binding routines, and to replace ad-hoc placeholder keywords with appropriate entries for the specific routine. When a particular specific implementation cannot be generated by a template, the format allows to override the template mechanism and to pass the specific implementation.

We can note that the Fortran 77 interface is a direct subproduct of the internal implementation of the library, therefore is convenient to deploy it even if not used by client code.

# Conclusions

This work surveyed two innovative techniques dedicated to the *ab initio* study of molecular systems.

The localization technique makes use of an optimization procedure that preserves the locality of a starting guess. An extension of this technique, named Freeze-and-Cut, allows the reduction of the computational cost by performing a neglection of those parts of the molecular system not involved in the phenomenon under study. The localized molecular orbitals are frozen at a lower level of theory (SCF in this case) to keep into account their effect. The remaining orbitals are projected onto a subset of the original basis set, allowing a reduced AO/MO two-electron integrals transformation.

The method proved its efficiency on two molecular system: the aminoacid zwitterion (7Z)-13 ammoniotridec-7-enoate, specifically designed to test the response of the method to charge interactions, and the highly conjugated $C_{13}$ polyenal. A third test on the acetone plus 6 $H_2O$ was attempted to evaluate excited states. In this case, a higher dimensionality of the basis set is needed to produce effective results, but this attempt reported a sensivity of the Freeze-and-Cut method to diffuse basis sets.

A possible solution to this problem is the use of non-orthogonal orbitals between the frozen and non-frozen sets, and although it seems to produce stable results, the current procedure have to be reconsidered with the assumption of non-orthogonality. Further developments are needed to face this improvement.

The second technique studied during this Ph.D. is the n-Electron Valence state Perturbation Theory (NEVPT). NEVPT is a multireference perturbation theory applicable to CAS wavefunctions. It is based on a two-electron Hamiltonian, the Dyall Hamiltonian, which provides a straightforward and efficient development for both theoretical and computational implementation. The perturber wavefunctions are multireference in nature, and belongs to eight excitation classes, depending on the promotion scheme of the electrons.

NEVPT is available in two accuracy levels, depending on the degree of contraction of the perturber space: the Strongly Contracted NEVPT uses a monodimensional space for each class, while the Partially Contracted uses a space of higher dimensionality but smaller than the complete space. The Strongly Contracted approach produces results of good accuracy when compared to the Partially Contracted, and a difference between the two approaches has been noted as a symptom of a poor zero-order wavefunction.

NEVPT is available as a Single-State approach, where the perturbation is applied to a specified electronic state, and also as a Quasi-Degenerate approach,

where the perturbation effects are applied simultaneously to a set of states. The Quasi-Degenerate approach keeps into account the interactions between wavefunctions after the perturbative corrections, allowing mixing of the states.

Finally, a variation of NEVPT named Non-Canonical has been deployed to obtain invariability for intraclass orbital rotation for core and virtual orbitals. This approach removes the need to perform the evaluation on canonical orbitals, allowing the direct application of NEVPT on a set of localized orbitals.

The various aspects of NEVPT has been applied to a large set of cases, all involving the carbonyl system. Every performed evaluation highlighted the strengths of this method, in particular its computational efficiency, its accuracy and the absence of intruder states. In the Quasi-Degenerate NEVPT further developments can be applied to open the possibility toward an iterative treatment. An attempt has been successfully made to deploy an integrated Localized+NEVPT Non-Canonical evaluation. This will open the possibility to perform highly accurate evaluations on large molecular systems.

Finally, a collaboration with the CINECA supercomputing center lead to accurate insights in the problems related to the deployment of a grid infrastructure for quantum chemistry evaluations. Two problems have been recognized as rather central in the first phase of the project: the need of a library for parsing XML with the Fortran language, and a common file format for large binary data.

The first problem has been addressed with the development of the F90xml library. This library provides a DOM compliant interface to the Fortran environment, acting as a wrapper of the gdome2 library.

The second problem involved an accurate analysis of previous common formats in the field, keeping into account a very large set of new requisites, such as platform independence, ease of use, debugging accuracy, and extensibility. This research lead to the deployment of a data model which has been implemented in the high-level library Q5Cost. This library proved its effectiveness allowing various codes to share informations, thus opening new possibilities in the integration of methods and the development of a grid metasystem. The library also comes with a proficient testsuite, aimed at stressing the interface for unexpected error conditions. The result is a very accurate, extensible and solid library.

The main drawbacks of this library can be found in the followings: the envisioned data model at the moment is not completely implemented in the interface, and the data storage for integrals can be made more efficient in terms of space occupation. In any case, the current implementation is optimal for a first-time experimentation in a new panorama for the computational quantum chemistry, and the detailed informations about the indexes of the stored integrals can be removed, providing files of reduced size.

# Conclusions

Ce travail a regardé deux nouvelles techniques pour l'étude *ab initio* des systèmes moléculaires. La technique de localisation utilise une procedure d'optimisation qui preserve la localité d'un groupe d'orbitales de guess localisée. Une extension de cette technique, la technique Freeze-and-Cut, permet la réduction du coût computationnel après élimination de la partie du système moléculaire pas intéressée au phénomène étudié. Les orbitales localisées sont gelées à un niveau plus bas de théorie (dans notre cas, une évaluation SCF) et leur effect est preservé. Pour réduire le poids computationnel pour la transformation des integrales bielectroniques, les autres orbitales sont projectées sur un sous-ensemble de la bas atomique original. Cette méthode a été essayé sur deux systèmes moléculaires: le (7Z)-13 ammoniotridec-7-enoate, un aminoacide construit pour évaluer la réponse de la méthode al'interaction de charge, et le polyenal C13. Une troisième évaluation a été effectué sur l'etat excité $n \rightarrow \pi^*$ de la molecule d'acetone avec 6 molecules d'eau. Dans ce cas, la dimensionalité plus grande de la bas atomique a démontré une sensibilité de la méthode Freeze-and-Cut à des bases diffuses. Une solution peut être l'utilise de orbitales non-ortogonaux entre le groupe gelé et non-gelé. Cette solution donne des résultats plus corrects, mais la procedure courant doit être adaptée avec l'assumption de non orthogonalité, et donc autres développements theoretiques sont nécessaires.

La deuxiéme technique etudiée pendant ce travail de These est la théorie perturbative "n-Electron Valence state Perturbation Theory" (NEVPT). La NEVPT est une théorie perturbative multiréférence qui peut être appliquée à des fonction d'onde de type CAS. Le développement theoretique utilise un Hamiltonien bielectronique, l'Hamiltonien de Dyall, qui donne une théorie propre et efficiente. Les fonctions perturbatives sont de nature multiréférencielle, et peuvent être classifiée dans huit classes d'excitation, selon le schéma de promotion des electrons. La NEVPT existe dans deux versions, la Strongly Contracted et la Partially contracted, selon le niveau de contraction de l'espace perturbatif. La Strongly Contracted NEVPT utilise un espace monodimensionnel pour chaque classe, et la Partially Contracted utilise un espace à dimensionalité plus haute, mais plus petite que l'espace complet des determinants excités. L'approche Strongly Contracted donne des résultats de bonne qualité si comparée avec la Partially Contracted. Une différence entre eux est normalement un symptôme d'une mauvaise déscription à l'ordre-zero.

La NEVPT est disponible comme single état, oú la perturbation est appliquée à un particulier état électronique. La NEVPT Quasi Degenerate applique l'effet perturbatif sur un ensemble d'états à la fois. Cette approche considère les inter-

actions entre les fonctions d'onde après la perturbation, et permit le mélange de ce fonctions. Enfin, une variation de la NEVPT appellée Non Canonical a été développée pour obtenir l'invariabilité des résultats pour le mélange des orbitales de core et virtuelles. Cette approche élimine la nécessité de travailler avec des orbitales canoniques, et permit l'utilisation directe avec un ensemble d'orbitales localisées.

Les différents types de NEVPT ont été utilisés sur un ensemble de systémes concernant le group carbonyl. Toutes les évaluations effectuées donnent des résultats intéressants, et sont aussi interéssantes les propriétés de la NEVPT, comme l'efficacité computationnelle et l'absence d'états intrus. Les résultats de la théorie QD-NEVPT peuvent être ameliorées avec une procedure iterative, qui peut être une bonne solution à des problémes dans les résultats obtenus. L'integration de la théorie NEVPT avec l'approche localisée a été effectuée avec la NEVPT Non Canonique. Cette travail sera important pour effectuer des calculs sur systémes moléculaires de grande taille.

Une collaboration avec le centre de supercalcul CINECA a été aussi effectuée dans le cadre d'un projet international entre différent instituts de recherche en chimie quantique. L'objectif est l'integration des différentes méthodologies *ab initio* dans une réseau d'ordinateurs. Deux problèmes sont prioritaires dans la phase initial du projet: la nécessité d'une bibliotheque pour la lecture du format de fichier XML avec le langage Fortran, et le développement d'un format de fichier binaire commun pour le stockage de donnees de grand taille. Pour le premier problème, la bibliothèque F90xml a été développée. Cette bibliothèque interface libgdome2 (bibliothèque pour le langage C) avec Fortran.

Pour le deuxième problème, on a analisé le format de fichier dejà développé et evalué leur défects pour l'exécution en grid, qui nécessite de considérations additionnelles, comme indépendance de plateforme, simplicité d'utilise de la bibliothèque d'acces, exactitude des messages d'erreur et extensibilité. La recherche effectuée a conduit au développement d'un modèle des données qui a été implémenté dans la bibliothèque Q5Cost. Cette bibliothèque a été utilisée pour l'integration des différents logiciels, et ouvre des nouvelles possibilités pour le partage d'information quantistique et l'integration des méthodes *ab initio*. La bibliothèque a aussi une testsuite pour contrôler l'interface et chercher les problèmes dans l'implementation. Le résultat est une bibliothèque très solide et extensible. Le problèmes principaux qui restent à resoudre dans Q5Cost sont: le stockage des integrals n'est pas encore efficient en terms d'espace occupé, et l'implementation du modèle des données n'est pas complete, mais l'implémentation courante est suffisante pour le premiers développements et implementations.

# Bibliography


[1] Attila Szabo and Neil S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill (1989).

[2] R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press (1989).

[3] T. Koopmans, *Physica* **1**, pp. 104 (1933).

[4] L. Brillouin, *Actualités scientifiques et Industrielles* **71** (1933).

[5] L. Brillouin, *Actualités scientifiques et Industrielles* **159** (1934).

[6] B. Levy and G. Berthier, *Int. J. Quant. Chem.* **2**, pp. 307 (1968).

[7] S. Borini, D. Maynau, and S. Evangelisti, *J. Comp. Chem.* **26**(10), pp. 1042–1051 (2005).

[8] M. J. Field, P. A. Bash, and M. J. Karplus, *J. Comput. Chem.* **11**, pp. 700 (1990).

[9] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, pp. 19357 (1996).

[10] A. Dreuw and M. Head-Gordon, *J. Am. Chem. Soc.* **126**, pp. 4007 (2004).

[11] R. Poteau, I. Ortega, F. Alary, A. R. Solis, J. C. Barthelat, and J. P. Daudey, *J. Phys. Chem. A* **105**, pp. 198 (2001).

[12] R. Poteau, F. Alary, H. A. El Makarim, J. L. Heully, J. C. Barthelat, and J. P. Daudey, *J. Phys. Chem. A* **105**, pp. 206 (2001).

[13] H. Nakatsuji and K. Hirao, *J. Chem. Phys.* **68**, pp. 2053 (1978).

[14] H. Nakatsuji, *Chem. Phys. Lett.* **59**, pp. 362 (1978).

[15] H. Nakatsuji, *Chem. Phys. Lett.* **67**, pp. 329 (1979).

[16] H. Nakatsuji, *Chem. Phys. Lett.* **67**, pp. 334 (1979).

[17] K. Toyota, J. Hasegawa, and H. Nakatsuji, *Chem. Phys. Lett.* **250**, pp. 437 (1996).

[18] O. Chavet, R. Daudel, S. Dineri, and J. P. Malrieu, *Localization and Delocalization in Quantum Chemistry*, D. Reidel Ed. Dordrecht-Holland (1975).

[19] S. Goedecker, *Rev. Mod. Phys.* **71**, pp. 1085–1123 (1999).

[20] P. Y. Ayala and G. E. Scuseria, *J. Chem. Phys.* **110**, pp. 3660–3671 (1999).

[21] P. Pulay and S. Saebø, *Geometrical Derivatives of Energy Surfaces and Molecular Properties*, P. Jorgensen and J. Simons Reidel, Dordrecht (1986).

[22] G. Rauhut, P. Pulay, and H. J. Werner, *J. Comput. Chem.* **19**, pp. 1241–1254 (1998).

[23] G. Hetzer, P. Pulay, and H. J. Werner, *Chem. Phys. Lett.* **290**, pp. 143–149 (1998).

[24] P. Pulay and S. Saebø, *Theor. Chim. Acta* **69**(5-6), pp. 357–368 (1986).

[25] S. Saebø and P. Pulay, *J. Chem. Phys.* **86**(2), pp. 914–922 (1987).

[26] P. Reinhardt, J. P. Malrieu, À. Povill, and J. Rubio, *Int. J. Quantum Chem.* **70**, pp. 167–180 (1998).

[27] C. Hampel and H. J. Werner, *J. Chem. Phys.* **104**(16), pp. 6286–6297 (1996).

[28] G. E. Scuseria and P. Y. Ayala, *J. Chem. Phys.* **111**(18), pp. 8330–8343 (1999).

[29] M. Shütz, *J. Chem. Phys.* **113**(22), pp. 9986–10001 (2000).

[30] M. Shütz and H. J. Werner, *J. Chem. Phys.* **114**(2), pp. 661–681 (2001).

[31] C. Edminston and K. Ruedenberg, *Rev. Mod. Phys.* **34**, pp. 457 (1963).

[32] S. F. Boys, *Rev. Mod. Phys.* **32**, pp. 2 (1960).

[33] J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, pp. 4916 (1989).

[34] C. Angeli, G. Del Re, and M. Persico, *Chem. Phys. Lett.* **233**, pp. 102 (1995).

[35] W. A. Adams, *J. Chem. Phys.* **34**, pp. 89 (1961).

[36] P. W. Anderson, *Phys. Rev. Lett.* **4**, pp. 17 (1969).

[37] T. L. Gilbert, *Molecular Orbitals in Chemistry, Physics and Biology*, Academic, New York (1964).

[38] W. J. Hunt, P. J. Hay, and W. A. Goddard III, *J. Chem. Phys.* **57**(2), pp. 738–748 (1972).

[39] F. W. Bobrowicz and W. A. Goddard III, *Methods of Electronic Structure Theory*, Plenum: New York (1977).

[40] J. V. Pitarch-Ruiz, S. Evangelisti, and D. Maynau, *Int. J. Quantum Chem.* **101**(3), pp. 325–333 (2005).

[41] J. V. Pitarch-Ruiz, S. Evangelisti, and D. Maynau, *Int. J. Quantum Chem.* **97**, pp. 688–699 (2004).

[42] J. V. Pitarch-Ruiz, S. Evangelisti, and D. Maynau, *Chem. Phys. Lett.* **372**, pp. 22–27 (2003).

[43] C. Angeli, C. J. Calzado, R. Cimiraglia, S. Evangelisti, N. Guihéry, T. Leininger, J. P. Malrieu, D. Maynau, J. V. Pitarch-Ruiz, and M. Sparta, *Mol. Phys.* **101**(9), pp. 1389–1398 (2003).

[44] J. V. Pitarch Ruiz, S. Evangelisti, and D. Maynau, *J. Mol. Struct. (THEOCHEM)* **681**(1-3), pp. 203–207 (2004).

[45] N. Suaud, G. M. Pastor, S. Evangelisti, and D. Maynau, *Chem. Phys. Lett.* **378**, pp. 503–508 (2003).

[46] C. J. Calzado, S. Evangelisti, and D. Maynau, *J. Phys. Chem. A* **107**, pp. 7581–7588 (2003).

[47] C. J. Calzado, S. Evangelisti, and D. Maynau, *J. Mol. Stuct (THEOCHEM)* **621**, pp. 51–58 (2003).

[48] N. Guihéry, J. P. Malrieu, S. Evangelisti, and D. Maynau, *Chem. Phys. Lett.* **349**, pp. 555–561 (2001).

[49] D. Maynau, S. Evangelisti, N. Guihéry, C. J. Calzado, and J. P. Malrieu, *J. Chem. Phys.* **116**(23), pp. 10060–10068 (2002).

[50] C. Angeli, S. Evangelisti, R. Cimiraglia, and D. Maynau, *J. Chem. Phys.* **117**(23), pp. 10525–10533 (2002).

[51] P. O. Widmark, P. A. Malmqvist, and B. O. Roos, *Theor. Chim. Acta* **77**(5), pp. 291 (1990).

[52] K. G. Dyall, *J. Chem. Phys.* **102**, pp. 4909 (1995).

[53] K. Andersson and M. Barysz and A. Bernhardsson and M. R. A. Blomberg and D. L. Cooper and M. P. Fülscher and C. de Graaf and B. A. Hess and G. Karlström and R. Lindh and P.-Å. Malmqvist and T. Nakajima and P. Neogrády and J. Olsen and B. O. Roos and B. Schimmelpfennig and M. Schütz and L. Seijo and L. Serrano-Andrés and P. E. M. Siegbahn and J. Stålring and T. Thorsteinsson and V. Veryazov and and P.-O. Widmark, "MOLCAS Version 5.4" (2002), http://www.teokem.lu.se/molcas.

[54] "http://www.gnu.org, http://www.linux.org".

[55] G.Schaftenaar and J.H. Noordik, "Molden: a pre- and post-processing program for molecular and electronic structures", *J. Comput.-Aided Mol. Design* **14**, pp. 123 (2000), http://www.caos.kun.nl/~schaft/molden/molden.html.

[56] P. Kok and E. J. J. Groenen, *J. Am. Chem. Soc.* **118**, pp. 7790 (1996).

[57] M. Klessinger and E. Gunkel, *Tetrahedron* **34**, pp. 3591 (1978).

[58] B. M. Brandow, *Rev. Mod. Phys.* **39**, pp. 771 (1967).

[59] B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, pp. 5745 (1973).

[60] C. Møller and M. S. Plesset, *Phys. Rev.* **46**, pp. 618 (1934).

[61] J. P. Malrieu, Ph. Durand, and J. P. Daudey, *J. Phys. A* **18**, pp. 809 (1985).

[62] R. Cimiraglia and M. Persico, *J. Comp. Chem.* **8**, pp. 39 (1987).

[63] R. Cimiraglia, *Int. J. Quantum Chem.* **60**, pp. 167 (1996).

[64] C. Angeli, R. Cimiraglia, M. Persico, and A. Toniolo, *Theor. Chem. Acc.* **98**, pp. 57 (1997).

[65] C. Angeli and M. Persico, *Theor. Chem. Acc.* **98**, pp. 117 (1997).

[66] C. Angeli and R. Cimiraglia, *Theor. Chem. Acc.* **105**, pp. 259 (2001).

[67] P. S. Epstein, *Phys. Rev.* **28**, pp. 695 (1926).

[68] R. K. Nesbet, *Proc. Roy. Soc. London* **230**, pp. 312 (1955).

[69] R. Cimiraglia, *J. Chem. Phys.* **83**, pp. 1746 (1985).

[70] K. Wolinski and P. Pulay, *J. Chem. Phys.* **90**, pp. 3647 (1997).

[71] K. Hirao, *Chem. Phys. Lett.* **190**, pp. 374 (1992).

[72] R. J. Harrison, *J. Chem. Phys.* **94**, pp. 5021 (1991).

[73] R. B. Murphy and R. P. Messmer, *Chem. Phys. Lett.* **183**, pp. 443 (1991).

[74] P. M. Kozlowski and E. R. Davidson, *Chem. Phys. Lett.* **222**, pp. 615 (1994).

[75] K. Andersson, P. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, *J. Phys. Chem.* **94**, pp. 5483 (1990).

[76] K. Andersson, P. Malmqvist, and B. O. Roos, *J. Chem. Phys.* **96**, pp. 1218 (1992).

[77] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, *J. Chem. Phys.* **114**(23), pp. 10252 (2001).

[78] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *J. Chem. Phys.* **117**(20), pp. 9138–9153 (2002).

[79] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *Chem. Phys. Lett.* **350**(3-4), pp. 297 (2001).

[80] C. Angeli, R. Cimiraglia, and J. P. Malrieu, *Chem. Phys. Lett.* **317**, pp. 472 (2000).

[81] C. Angeli, S. Borini, L. Ferrighi, and R. Cimiraglia, *J. Chem. Phys.* **122**, pp. 114304–114313 (2005).

[82] C. Angeli, S. Borini, L. Ferrighi, and R. Cimiraglia, *J. Mol. Struct. (THEOCHEM)* **718**(1-3), pp. 55–69 (2005).

[83] E. W.-G. Diau and C. Kötting and A. H. Zewail, *Chem. Phys. Chem.* **2**(5), pp. 273 (2001).

[84] W. A. Noyes Jr., *Photochemistry and Reaction Kinetics*, P. G. Ashmore, F. S. Dainton and T. M. Sugden, Cambridge University Press (1967).

[85] Y. Haas, *Photochem. Photobiol. Sci.* **3**(1), pp. 6 (2004).

[86] C. B. Moore and J. C. Weisshaar, *Annu. Rev. Phys. Chem.* **34**, pp. 525 (1983).

[87] M. B. Robin, *Higher Excited States of Polyatomics Molecules, Volume 3*, Academic New York (1985).

[88] A. D. Baker, C. Baker, C. R. Brundle, and D. W. Turner, *Int. J. Mass Spectrom. Ion Phys.* **1**, pp. 285 (1968).

[89] C. R. Brundle, M. B. Robin, N. A. Kuebler, and H. Basch, *J. Am. Chem. Soc.* **94**, pp. 1451 (1972).

[90] K. N. Walzl, C. F. Koerting, and A. Kuppermann, *J. Chem. Phys.* **87**, pp. 3796 (1987).

[91] D. E. Reisner, R. W. Field, J. L. Kinsey, and H.-L. Dai, *J. Chem. Phys.* **80**(12), pp. 5968 (1984).

[92] K. B. Wiberg, Y. Thiel, L. Goodman, and J. Leszczynski, *J. Phys. Chem.* **99**(38), pp. 13850 (1995).

[93] M. Merchan and B. O. Roos, *Theor. Chem. Acc.* **92**, pp. 227 (1995).

[94] M. R. J. Hachey, P. J. Bruna, and F. Grein, *J. Phys. Chem.* **99**(20), pp. 8050–8057 (1995).

[95] M. R. J. Hachey, P. J. Bruna, and F. Grein, *J. Mol. Spectrosc.* **176**(2), pp. 375 (1996).

[96] T. Helgaker, H.J.Aa. Jensen, P. Jørgensen, J. Olsen, K. Ruud, H. Ågren, A.A. Auer, K.L. Bak, V. Bakken, O. Christiansen, S. Coriani, P. Dahle, E.K. Dalskov, T. Enevoldsen, B. Fernandez, C. Hättig, K. Hald, A. Halkier, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, R. Kobayashi, H. Koch, K.V. Mikkelsen, P. Norman, M.J. Packer, T.B. Pedersen, T.A. Ruden, A. Sanchez, T. Saue, S.P.A. Sauer, B. Schimmelpfennig, K. O. Sylvester-Hvid, P.R. Taylor, and O. Vahtras, "Dalton, a molecular electronic structure program, Release 1.2 (2001)", http://www.kjemi.uio.no/software/dalton.

[97] M. Dallos, T. Muller, H. Lischka, and R. Shepard, *J. Chem. Phys.* **114**(2), pp. 746 (2001).

[98] C. Angeli, C. J. Calzado, R. Cimiraglia, S. Evangelisti, and D. Maynau, *Mol. Phys.* **101**(13), pp. 1937–1944 (2003).

[99] T. Leininger, C. Angeli, S. Evangelisti, R. Cimiraglia, and D. Maynau, *Chem. Phys. Lett.* **371**, pp. 49–55 (2003).

[100] Takagi K. and Oka T., *J. Phys. Soc. Jpn.* **18**, pp. 1174 (1963).

[101] M. D. Harmony, V. W. Laurie, R. L. Kuczkowski, R. H. Schwendeman, D. A. Ramsay, F. J. Lovas, W. J. Lafferty, and A. G. Maki, *J. Phys. Chem. Ref. Data* **8**, pp. 619 (1979).

[102] R. Nelson and L. Pierce, *J. Mol. Spectrosc.* **18**, pp. 344 (1965).

[103] D. C. Moule and A. D. Walsh, *Chem. Rev. (Washington, D.C.)* **75**(1), pp. 67 (1975).

[104] D. J. Clouthier and D. A. Ramsay, *Annu. Rev. Phys. Chem.* **34**, pp. 31 (1983).

[105] J. L. Whitten and M. Hackmeyer, *J. Chem. Phys.* **51**, pp. 5584 (1969).

[106] R. J. Buenker and S. D. Peyerimhoff, *J. Chem. Phys.* **53**, pp. 1368 (1970).

[107] E. R. Davidson and L. E. McMurchie, *Excited States*, E. C. Lim - Academic New York - Vol. 5 (1982).

[108] S. R. Gwaltney and R. J. Bartlett, *Chem. Phys. Lett.* **241**, pp. 26–32 (1995).

[109] S. R. Gwaltney, M. Nooijen, and R. J. Bartlett, *Chem. Phys. Lett.* **248**, pp. 189 (1996).

[110] K. B. Wiberg, R. E. Stratmann, and M. J. Frisch, *Chem. Phys. Lett.* **297**, pp. 60 (1998).

[111] C. Adamo, G. E. Scuseria, and V. Barone, *J. Chem. Phys.* **111**, pp. 2889 (1999).

[112] A. I. Krylov, C. D. Sherrill, and M. Head-Gordon, *J. Chem. Phys.* **113**, pp. 6509 (2000).

[113] P. J. Bruna, M. R. J. Hachey, and F. Grein, *J. Phys. Chem.* **99**, pp. 16576 (1995).

[114] T. Muller and H. Lischka, *Theor. Chem. Acc.* **106**, pp. 369 (2001).

[115] V. N. Staroverov and E. R. Davidson, *Chem. Phys. Lett.* **296**, pp. 435 (1998).

[116] H. Lischka, M. Dallos, and R. Shepard, *Mol. Phys.* **100**, pp. 1647 (2002).

[117] J. Pitarch-Ruiz, J. Sanchez-Marin, A. S. De Meras, and D. Maynau, *Mol. Phys.* **101**, pp. 483 (2003).

[118] C. Angeli, S. Borini, and R. Cimiraglia, *Theor. Chem. Acc.* **111**(2-6), pp. 352–357 (2004).

[119] O. Setokuchi, M. Sato, S. Matuzawa, and Y. Simizu, *J. Mol. Struct. (THEOCHEM)* **365**, pp. 29 (1996).

[120] O. Setokuchi and Y. Simizu, *J. Mol. Struct. (THEOCHEM)* **401**(1-2), pp. 29 (1997).

[121] S. R. Gwaltney and R. J. Bartlett, *J. Chem. Phys.* **110**(1), pp. 62 (1999).

[122] D. W. Liao, A. M. Mebel, M. Hayashi, Y. J. Shiu, Y. T. Chen, and S. H. Lin, *J. Chem. Phys.* **111**(1), pp. 205–215 (1999).

[123] A. B. Rocha and C. E. Bielschowsky, *Chem. Phys. Lett.* **337**(4-6), pp. 331 (2001).

[124] M. Merchan, B. O. Roos, R. McDiarmid, and X. Xing, *J. Chem. Phys.* **104**, pp. 1791 (1996).

[125] K. Wiberg, A. de Oliveira, and G. Trucks, *J. Phys. Chem. A* **106**, pp. 4192 (2002).

[126] O. Setokuchi, S. Matuzawa, and Y. Simizu, *Chem. Phys. Lett.* **284**(1-2), pp. 19 (1998).

[127] D. Liu, W.-H. Fang, and X.-Y. Fu, *Chem. Phys. Lett.* **325**(1-3), pp. 86 (2000).

[128] E. W.-G. Diau and C. Kötting and T. I. Solling and A. H. Zewail, *Chem. Phys. Chem.* **3**, pp. 57 (2002).

[129] C. M. Hadad, J. B. Foresman, and K. B. Wiberg, *J. Phys. Chem.* **97**(17), pp. 4293–4312 (1993).

[130] A. Niño, C. Munoz-Caro, and D. C. Moule, *J. Phys. Chem.* **98**(6), pp. 1519 (1994).

[131] C. Munoz-Caro, A. Niño, and D. C. Moule, *J. Mol. Struct. (Theochem)* **315**, pp. 9 (1994).

[132] M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, *Chem. Phys. Lett.* **219**, pp. 21 (1994).

[133] I. Frank, J. Hutter, D. Marx, and M. Parrinello, *J. Chem. Phys.* **108**(10), pp. 4060 (1998).

[134] R. H. Staley, L. B. Harding, W. A. Goddard, and J. L. Beauchamp, *Chem. Phys. Lett.* **36**, pp. 589 (1975).

[135] E. H. Van Veen, W. L. Van Dijk, and H. H. Brongersma, *Chem. Phys.* **16**, pp. 337 (1976).

[136] D. J. Tozer, R. D. Amos, N. C. Handy, B. O. Roos, and L. Serrano-Andres, *Mol. Phys.* **97**, pp. 859 (1999).

[137] F. Aquilante, M. Cossi, O. Crescenzi, G. Scalmani, and V. Barone, *Mol. Phys.* **101**, pp. 1945 (2003).

[138] H. Sakurai and S. Kato, *J. Mol. Struct (THEOCHEM)* **461**, pp. 145 (1999).

[139] M. Lepage, M. Michaud, and L. Sanche, *J. Chem. Phys.* **112**(15), pp. 6707 (2000).

[140] K. B. Wiberg, C. M. Hadad, J. B. Foresman, and W. A. Chupka, *J. Phys. Chem.* **96**, pp. 10756 (1992).

[141] W. M. St. John III, R. C. Estler, and J. P. Doering, *J. Chem. Phys.* **61**, pp. 763 (1974).

[142] V. A. Bataev, V. I. Pupyshev, and I. A. Godunov, *J. Mol. Struct.* **480**, pp. 263 (1999).

[143] T. Nakajima and S. Kato, *J. Chem. Phys.* **105**(14), pp. 5927 (1996).

[144] K. K. Baeck, *J. Chem. Phys.* **112**(1), pp. 1 (2000).

[145] J. E. Del Bene, S. R. Gwaltney, and R. J. Bartlett, *J. Phys. Chem. A* **102**(26), pp. 5124 (1998).

[146] M. Kreglewski, *J. Mol. Struct.* **60**, pp. 105 (1980).

[147] P. Jensen and P. R. Bunker, *J. Mol. Spectrosc.* **94**(1), pp. 114 (1982).

[148] Y. Yamaguchi, S. S. Wesolowski, T. J. Van Huis, and H. F. Schaefer III, *J. Chem. Phys.* **108**(13), pp. 5281 (1998).

[149] J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.* **96**(1), pp. 135 (1992).

[150] D. J. Clouthier and D. C. Moule, *Top. Curr. Chem.* **150**, pp. 167 (1989).

[151] K. K. Innes, *J. Mol. Spectrosc.* **99**(2), pp. 294 (1983).

[152] V. A. Job, V. Sethuraman, and K. K. Innes, *J. Mol. Spectrosc.* **30**, pp. 365 (1969).

[153] H. Liu, E. C. Lim, C. Munoz-Caro, A. Niño, R. H. Judge, and D. C. Moule, *J. Chem. Phys.* **105**(7), pp. 2547 (1996).

[154] L. M. Hubbard, D. F. Bocian, and R. R. Birge, *J. Am. Chem. Soc.* **103**(12), pp. 3313 (1981).

[155] M. Hachey, P. J. Bruna, and F. Grein, *J. Chem. Soc. Faraday Trans.* **90**(5), pp. 683 (1994).

[156] S. Taylor, D. G. Wilden, and J. Comer, *Chem. Phys.* **70**(3), pp. 291 (1982).

[157] A. D. Walsh, *J. Chem. Soc.* **1953**, pp. 2306 (1953).

[158] W. D. Allen and H. F. Schaefer III, *J. Chem. Phys.* **87**(12), pp. 7076 (1987).

[159] M. R. J. Hachey and F. Grein, *Chem. Phys. Lett.* **256**(1-2), pp. 179 (1996).

[160] M. Noble and E. K. C. Lee, *J. Chem. Phys.* **81**(4), pp. 1632 (1984).

[161] H. Liu, C. Lim, R. H. Judge, and D. C. Moule, *J. Chem. Phys.* **102**(11), pp. 4315 (1995).

[162] M. Baba, I. Hanazaki, and U. Nagashima, *J. Chem. Phys.* **82**, pp. 3938 (1985).

[163] L. D. Waits, R. J. Horwitz, and J. A. Guest, *Chem. Phys.* **155**(1), pp. 149 (1991).

[164] S. W. North, D. A. Blank, J. D. Gezelter, C. A. Longfellow, and Y. T. Lee, *J. Chem. Phys.* **102**(11), pp. 4447 (1995).

[165] R. F. Borkman and D. R. Kearns, *J. Chem. Phys.* **44**, pp. 945 (1966).

[166] W. O. George, B. F. Jones, Rh. Lewis, and J. M. Price, *Journal of Molecular Structure* **550-551**, pp. 281–296 (2000).

[167] J. L. Duncan and P. D. Mallinson, *Chem. Phys. Lett.* **23**, pp. 597 (1973).

[168] T. Nakanaga, S. Kondo, and S. Saeki, *J. Chem. Phys.* **76**(8), pp. 3860 (1982).

[169] H.T. Kim and S. L. Anderson, *J. Chem. Phys.* **114**(7), pp. 3018 (2001).

[170] H. Hollenstein and H. H. Gunthard, *Spectrochim. Acta* **27A**, pp. 2027 (1971).

[171] H. Gu, T. Kundu, and L. Goodman, *J. Chem. Phys.* **97**(2), pp. 7194 (1993).

[172] H. Hollenstein and H. H. Gunthard, *J. Mol. Spectrosc.* **84**(2), pp. 457 (1980).

[173] J. L. Hardwick and S. M. Till, *J. Chem. Phys.* **70**(5), pp. 2340 (1979).

[174] J. M. F. van Dijk, M. J. H. Kemper, J. H. M. Kerp, and H. M. Buck, *J. Chem. Phys.* **69**(6), pp. 2453 (1979).

[175] J. F. Stanton, J. Gauss, N. Ishikawa, and M. Head-Gordon, *J. Chem. Phys.* **103**, pp. 4160 (1995).

[176] V. T. Jones and J. B. Coon, *J. Mol. Spectrosc.* **31**, pp. 137 (1969).

[177] J. C. D. Brand and C. G. Stevens, *J. Chem. Phys.* **58**(8), pp. 3331 (1973).

[178] G. W. Robinson and V. E. DiGiorgio, *Can. J. Phys.* **36**, pp. 31 (1958).

[179] J. S. Yadav and J. D. Goddard, *J. Chem. Phys.* **84**(5), pp. 2682 (1986).

[180] M. Baba, U. Nagashima, and I. Hanazaki, *Chem. Phys.* **93**(3), pp. 425 (1985).

[181] M. Noble, E. C. Apel, and E. K. C. Lee, *J. Chem. Phys.* **78**(5), pp. 2219 (1983).

[182] N. N. Yakovlev and I. A. Godunov, *Can. J. Chem.* **70**(3), pp. 931 (1992).

[183] H. Zuckermann, Y. Haas, M. Drabbels, J. Hainze, L. Meerts, J. Reuss, and J. van Bladel, *Chem. Phys.* **163**(2), pp. 193 (1992).

[184] B. O. Roos, M. P. Fülsher, P.-A. Malmqvist, M. Merchán, and L. Serrano Andrés, *Quantum Mechanical Electronic Structure Calculation with Chemical Accuracy*, Langhoff S. R., Kluwer Academic Publishers, the Netherlands (1995).

[185] K. Kaufmann, W. Baumeister, and M. Jungen, *J. Phys. B* **22**, pp. 2223 (1989).

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

[187] O. Parisel and Y. Ellinger, *Chem. Phys.* **205**, pp. 323 (1996).

[188] P. Brint, J. P. Connerade, C. Mayhew, and K. Sommer, *J. Chem. Soc. Faraday Trans.* **281**, pp. 1643 (1985).

[189] M. Suto, X. Wang, and L. C. Lee, *J. Chem. Phys.* **85**, pp. 4228 (1986).

[190] P. Durand and J.-P. Malrieu, *Adv. Chem. Phys.* **67**, pp. 1 (1987).

[191] J. Finley, P.-A. Malmqvist, B. O. Roos, and L. Serrano-Andrés, *Chem. Phys. Lett.* **288**, pp. 299 (1998).

[192] J. G. Philis and L. Goodman, *J. Chem. Phys.* **98**, pp. 3795 (1993).

[193] R. McDiarmid and A. Sabljić, *J. Chem. Phys.* **89**, pp. 6086 (1988).

[194] C. Angeli, S. Borini, M. Cestari, and R. Cimiraglia, *J. Chem. Phys.* **121**(9), pp. 4043–4049 (2004).

[195] I. Lindgren, *J. Phys. B* **7**(18), pp. 2441–2470 (1974).

[196] A. Zaitsevskii and J. P. Malrieu, *Chem. Phys. Lett.* **233**(5-6), pp. 597–604 (1995).

[197] J. des Cloizeaux, *Nucl. Phys.* **20**, pp. 321 (1960).

[198] C. Angeli, S. Borini, A. Cavallini, M. Cestari, R. Cimiraglia, L. Ferrighi, and M. Sparta, *Int. J. Quantum Chem.* **106**, pp. 686 (2006).

[199] "http://setiathome.ssl.berkeley.edu".

[200] "http://www.intel.com/research/silicon/mooreslaw.htm".

[201] "http://www.ietf.org/rfc/rfc2616.txt".

[202] "http://www.apache.org".

[203] "http://www.ietf.org".

[204] "http://www.w3.org".

[205] "http://www.oasis-open.org".

[206] "http://www.ietf.org/rfc/rfc0793.txt".

[207] "http://www.ietf.org/rfc/rfc0791.txt".

[208] "http://www.w3.org/xml".

[209] "http://www.w3.org/tr/soap".

[210] "http://www.w3.org/tr/wsdl".

[211] "http://www.uddi.org".

[212] "http://www.w3.org/rdf".

[213] "http://www-3.ibm.com/software/solutions/webservices/pdf/wsfl.pdf".

[214] "http://www.cs.virginia.edu/~legion".

[215] "http://www.globus.org".

[216] "http://www.ogsadai.org.uk".

[217] "http://unicore.sourceforge.net".

[218] "http://www.csm.ornl.gov/pvm/pvm_home.html".

[219] "http://www-unix.mcs.anl.gov/mpi".

[220] "http://www.openmp.org".

[221] "http://cost.cordis.lu/src/action_info/ActionD23.cfm".

[222] P. Murray-Rust and H. S. Rzepa, *J. Chem. Inf. Comp. Sci.* **39**, pp. 928 (1999).

[223] P. Murray-Rust and H. S. Rzepa, *J. Chem. Inf. Comp. Sci.* **41**, pp. 1113 (2001).

[224] G. Gkoutos, P. Murray-Rust, and H. S. Rzepa, *J. Chem. Inf. Comp. Sci.* **41**, pp. 1124 (2001).

[225] P. Murray-Rust, H. S. Rzepa, M. Wright, and S. Zara, *ChemComm* , pp. 1471–1472 (2000).

[226] P. Murray-Rust, H. S. Rzepa, and M. Wright, *New J. Chem.* , pp. 618–634 (2001).

[227] "http://hdf.ncsa.uiuc.edu/hdf5".

[228] A. Streit, D. Erwin, Th. Lippert, D. Mallmann, R. Menday, M. Rambadt, M. Riedel, M. Romberg, B. Schuller, and Ph. Wieder, *To be published* (2005).

[229] D. Erwin, editor, *UNICORE plus final report – uniform interface to computing resources*, Forschungszentrum Julich (2003), ISBN 3-00-011592-7.

[230] "http://python.org".

[231] L. Yinghu, S.X. J., Z. Tao, J.R. Dahn, S. Porchet, L. Kiwi-Minsker, R. Doepper, A. Renken, S. Evangelisti, G.L. Bendazzoli, R. Ansaloni, F. Duri, and E. Rossi, *Chem. Phys. Lett.* **252**(10), pp. 437 (1996).

[232] G.L. Bendazzoli and S. Evangelisti, *Chem. Phys. Lett.* **252**(10), pp. 437 (1996).

[233] Mart Rentmeester, "http://nn-online.org/code/xml".

[234] Arjen Markus, "http://xml-fortran.sourceforge.net".

[235] Alberto Garcia and Jon Wakelin, "http://lcdx00.wm.lc.ehu.es/~wdpgaara/xml/index.html".

[236] "http://g95.sourceforge.net".

[237] "http://gcc.gnu.org/fortran".

[238] Stefano Borini, "http://icbe61.unife.it/f77xml".

[239] "http://gdome2.cs.unibo.it".

[240] "http://www.gnome.org".

[241] "http://www.w3.org/TR/2000/REC-DOM-Level-2-Core-20001113".

[242] "http://www.gtk.org".

[243] "http://xmlsoft.org".

[244] "http://www.gnu.org/copyleft/lesser.html".