University of Lille

École Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement

Full quantum simulations of the interaction between atmospheric molecules and model soot particles.

Thesis presented by

Ramón L. PANADÉS-BARRUETA

the 23rd of October, 2020

in fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

Supervised by

Prof. Dr. Daniel PELÁEZ

Jury:

Céline TOUBIN	Professor	President
Fabien GATTI	Directeur de Recherche	Referee
Oriol VENDRELL	Professor	Referee
Annika BANDE	Professor	Examiner
Shirin FARAJI	Professor	Examiner
Andre S. PEREIRA GOMES	Chargé de Recherche	Examiner

Université de Lille

École Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement

---

Simulations quantiques de l'interaction entre molécules atmosphériques et particules de suies modèle

---

Thèse présentée par

Ramón L. PANADÉS-BARRUETA

le 23 octobre 2020

en vue de l'obtention du titre de

Docteur en Physique

Dirigée par

Prof. Dr. Daniel PELÁEZ

Jury:

Céline TOUBIN	Professeur	Présidente
Fabien GATTI	Directeur de Recherche	Rapporteur
Oriol VENDRELL	Professeur	Rapporteur
Annika BANDE	Professeur	Examinatrice
Shirin FARAJI	Professeur	Examinatrice
Andre S. PEREIRA GOMES	Chargé de Recherche	Examinateur

A mi padre, por todo y por siempre.

UNIVERSITÉ DE LILLE

Résumé

Laboratoire de Physique des Lasers, Atomes et Molécules

Doctorat en Physique

par Ramón L. PANADÉS-BARRUETA

Nous visons à simuler avec des arguments purement quantiques (noyaux et électrons) les processus d'adsorption et de photoréactivité du NO_2 adsorbé sur des particules de suie (modélisées comme de grands hydrocarbures aromatiques polycycliques, HAP) dans les conditions atmosphériques. Une description détaillée de ces processus est nécessaire pour comprendre le comportement différentiel (jour-nuit) de la production de HONO [1, 2], qui est un précurseur du radical hydroxyle (OH) [3]. En particulier, le mécanisme spécifique de l'interconversion entre NO_2 et HONO par la suie n'est pas encore totalement compris. En raison de sa pertinence particulière dans ce contexte, nous avons choisi le système Pyrène-NO_2 [1].

La première étape de cette étude a consisté à déterminer les configurations stables (états de transition et minima) du système Pyrène-NO_2. À cette fin, nous avons utilisé la méthode van der Waals Transition State Search using Chemical Dynamics Simulations (vdW-TSSCDS) [4], la généralisation de l'algorithme TSSCDS [5, 6] récemment développée dans notre groupe. Ainsi, le présent travail représente la première application de vdW-TSSCDS à un grand système (81D). Partant d'un ensemble de géométries d'entrée judicieusement choisies, la méthode susmentionnée permet de caractériser la topographie d'une surface d'énergie potentielle intermoléculaire (SEP), ou en d'autres termes, de déterminer les conformations les plus stables du système, de manière entièrement automatisée et efficace.

Les informations topographiques recueillies ont été utilisées pour obtenir une description globale (fit) du potentiel d'interaction, nécessaire à l'élucidation dynamique de l'interaction intermoléculaire (physisorption), des propriétés spectroscopiques et de la réactivité des espèces adsorbées. Pour atteindre ce dernier objectif, nous avons développé deux méthodologies différentes ainsi que les progiciels correspondants. La première d'entre elles est l'algorithme SRP-MGPF (Specific Reaction Parameter Multigrid POTFIT), qui est implémenté dans le progiciel SRPTucker [7]. Cette méthode calcule des SEPs (intermoléculaires) chimiquement précis par reparamétrage de méthodes semi-empiriques, qui sont ensuite tenseur-décomposées sous forme Tucker à l'aide de MGPF [8].

Ce logiciel a été interfacé avec succès avec la version Heidelberg du paquet MCTDH (Multi-configuration Time-Dependent Hartree) [9, 10]. La seconde méthode permet d'obtenir la SEP directement sous la forme mathématique requise par MCTDH, d'où son nom de Sum-Of-Products Finite-Basis-Representation (SOP-FBR) [11]. La SOP-FBR constitue une approche alternative aux méthododes d'ajustement NN. L'idée la sous-tend est simple : à partir d'une expansion Tucker *low rank* sur la grille, nous remplaçons les fonctions de base basées sur la grille par une expansion en termes de polynômes orthogonaux. Comme dans la méthode précédente, l'intégration avec la MCTDH a été assurée. Les deux méthodes ont été testées avec succès à un certain nombre de problèmes de référence, à savoir : le Hamiltonian Hénon-Heiles [12], la SEP global du H_2O [13], et la SEP d'isomérisation HONO (6D) [14].

À l'aide de toutes les méthodes mentionnées ci-dessus, nous avons abordé le calcul de la SEP global du système Pyrène-NO_2. Des routines de transformation de coordonnées appropriées ont été développées afin de faire correspondre les coordonnées cartésiennes aux coordonnées internes. Dans le domaine de la physisorption, les preuves recueillies avec la vdW-TSSCDS ont suggéré que la géométrie de la molécule de NO_2 n'est presque pas perturbée dans les points stationnaires par rapport à la molécule isolée. Ce fait a permis de la traiter comme un monomère rigide (6D). Les SEP seront utilisés pour obtenir l'état électronique de base (GS) et l'énergie du point zéro (ZPE) correspondante du système avec MCTDH. Le ZPE peut offrir une estimation précise de l'énergie d'adsorption de la molécule de NO_2 sur le Pyrène. En outre, le spectre d'absorption électronique du système sera obtenu en calculant la somme (pondérée par la distribution GS) des excitations verticales individuelles de chaque point stationnaire.

UNIVERSITÉ DE LILLE

Abstract

Laboratoire de Physique des Lasers, Atomes et Molécules

Ph.D. in Physics

by Ramón L. PANADÉS-BARRUETA

This work aims at simulating full quantum mechanically (nuclei and electrons) the processes of adsorption and photoreactivity of NO_2 adsorbed on soot particles (modeled as large Polycyclic Aromatic Hydrocarbons, PAHs) in atmospheric conditions. A detailed description of these processes is necessary to understand the differential day-nighttime behavior of the production of HONO [1, 2], which is a precursor of the hydroxyl radical (OH) [3]. In particular, the specific mechanism of the soot-mediated interconversion between NO_2 and HONO is to date not fully understood. Due to its particular relevance in this context, we have chosen the Pyrene-NO_2 as our model system [1].

The first stage in this study has consisted of the determination of the stable configurations (transition states and minima) of the Pyrene-NO_2 system. To this end, we have used the recently developed van der Waals Transition State Search using Chemical Dynamics Simulations (vdW-TSSCDS) method [4], the generalization of the TSSCDS algorithm [5, 6] developed in our group. In this way, the present work represents the first application of vdW-TSSCDS to a large system (81D). Starting from a set of judiciously chosen input geometries, the aforementioned method permits the characterization of the topography of an intermolecular Potential Energy Surface (PES), or in other words the determination of the most stable conformations of the system, in a fully automated and efficient manner.

The gathered topographical information has been used to obtain a global description (fit) of the interaction potential, necessary for the dynamical elucidation of the intermolecular interaction (physisorption), spectroscopic properties, and reactivity of the adsorbed species. To achieve this last goal, we have developed two different methodologies together with the corresponding software packages. The first one of them is the Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF) algorithm, which is implemented in the SRPTucker package [7]. This method computes chemically accurate (intermolecular) PESs through the reparametrization of semiempirical methods, which are subsequently tensor decomposed into Tucker form using MGPF [8]. This software has been successfully interfaced with the Heidelberg version of the Multi-Configuration Time-Dependent

Hartree (MCTDH) package [9, 10]. The second method allows for obtaining the PES directly in the mathematical form required by MCTDH, thence its name Sum-Of-Products Finite-Basis-Representation (SOP-FBR) [11]. SOP-FBR constitutes an alternative approach to Neural Networks (NN)-fitting methods. The idea behind it is simple: from the basis of a low-rank Tucker expansion on the grid, we replace the grid-based basis functions by an expansion in terms of orthogonal polynomials. As in the previous method, smooth integration with MCTDH has been ensured. Both methods have been successfully benchmarked with a number of reference problems, namely: the Hénon-Heiles Hamiltonian [12], a global H_2O PES [13], and the HONO isomerization PES (6D) [14].

With the aid of all the above mentioned methods, we have tackled the computation of the global PES of the Pyrene-NO_2 system. Suitable coordinate transformation routines have been developed to map the Cartesian coordinates to internal coordinates. In the physisorption domain, the evidence collected with vdW-TSSCDS has suggested that the geometry of the NO_2 molecule is almost not perturbed in the stationary points with respect to the isolated molecule. This fact has enabled its treatment in a rigid monomer fashion (6D). The PESs will be used to obtain the electronic ground state (GS) and corresponding Zero-Point Energy (ZPE) of the system with MCTDH. The ZPE can offer an accurate estimate of the adsorption energy of the NO_2 molecule over the Pyrene. Additionally, the electronic absorption spectrum of the system will be obtained by computing the sum (weighted by the GS distribution) of the individual vertical excitations of each stationary point.

Acknowledgements

A mi Santa Madre Yemayá junto a todos los Orishas, a mis abuelos y bisabuelos paternos. ¡Gracias por la fuerza!

I would like to start by profoundly thanking my *Doktorvater* Daniel Peláez Ruiz for giving me the opportunity of developing this interesting project. His vast experience and valuable advice have been fundamental in my growth as a scientist. During these three years we have had countless discussions about Physics, Chemistry, Programming and *what a molecule is*, which have become cherished memories that I would like to transmit someday. Thanks as well for the constant exchange of music, books, politics, and life experiences in general. *¡Gracias mi amigo!*

I would like to acknowledge all the organizations that financially supported this thesis: the University of Lille, the Région Hauts-de-France and the German Academic Exchange Service (DAAD). I also thank the support of the Labex CaPPA and CLIMIBIO projects, and the Centre National de la Recherche Scientifique (CNRS).

I am extremely grateful to Maurice Monnerville for the long scientific discussions, for his helpful guidance in difficult situations, for his music, and for letting me be part of his family. I deeply thank Alejandro Rivero Santamaría, who was the person that introduced me to the field of Molecular Dynamics. I learned a lot from him in my License days at InSTEC (and I still do), both as a student and as a human being. Without his support, I would not be here today.

I thank all my colleagues at the PCMT group of the PhLAM laboratory for their help, especially Céline Toubin, Valérie Vallet, Denis Duflot, Andre S. Pereira, Florent Réal, Stéphane Briquez and Hanna Oher. I am very grateful to all my colleagues at the TC department of the University of Heidelberg, particularly Oriol Vendrel, Hans-Dieter Meyer, Jochen Schirmer, Horst Köppel, Lorenz Cederbaum, Sudip Sasmal, Markus Schröder, Kirill Gokhberg, and Alex I. Kuleff. I thank Emilio Martínez-Nuñez, James Stewart and Sabine Kopec for very helpful scientific discussions.

I would like to express my gratitude to the educators that assisted my formation in its different stages: the professors of the Master 2 Atmospheric Environment, in particular Thérèse Huet, Philippe Goloub and Philippe Dubuisson; the professors of the InSTEC, especially Alfo Batista, Juan Estévez, Helen C. Peñate and Evelin Gonzáles; the teachers of the *Preseleción Nacional de Química*, particularly Rolando Alfonso and Luis Guerra; the teachers of the IPVCE Eusebio Olivera, especially Agustín Plasencia, Félix Macías and Félix R. Gonzáles; and finally, my Math High School teacher Nelson Zuñiga.

Last but not least, I will refer to the foundations of any individual: the family. I have always counted with the strong support of my father Pupi, my brothers Leo and Misa, my mother Elsa, my aunt Ana Martha, my uncles Ede, Mauri and Pedri, my *padrinos* María and Leonel, and the rest of the *tropa* from Trinidad, Sancti Spíritus and Güines. But that is not all about it: 8326 km away from Cuba I have found the rest. I thank God for bringing the radiance of Cora into my life. She is my lover and confidant, who gives me strength to always look forward, and the confirmation of the powerful message at the end of "Into the Wild". Without her constant motivation and encouragement, this manuscript would not be finished today. I deeply thank Ines, Benno, Carl, Karin, Klaus, Hannelore, *und der ganzen Familie*, for welcoming me as part of their home. Finally, I thank all my friends (wherever they scatter in the world) especially Jimmy and Dumitro, and all the other persons that one way or another helped me reach this day.

Contents

List of Figures

List of Tables

Abbreviations

BBFS	**B**ond **B**reakage-**F**ormation **S**earch
BFGS	**B**royden **F**letcher **G**oldfarb **S**hanno algorithm
CC	**C**oupled-**C**luster
DFT	**D**ensity **F**untional **T**heory
DOF	**D**egrees **O**f **F**reedom
DVR	**D**iscrete **V**ariable **R**epresentation
ERI	**E**electron **R**epulsion **I**ntegral
FBR	**F**inite **B**ase **R**epresentation
FCI	**F**ull **C**onfiguration **I**nteraction
GTO	**G**aussian-**T**ype **O**rbitals
HL	**H**igh **L**evel
HOOI	**H**igher-**O**order **O**rthogonal **I**teration
HOSVD	**H**igher-**O**order **S**ingular **V**alue **D**ecomposition
IRC	**I**ntrinsic **R**eaction **C**oordinate
LL	**L**ow **L**evel
MBPT	**M**any **B**ody **P**erturbation **T**heory
MCPF	**M**onte **C**arlo **P**OT**F**IT
MCSCF	**M**ulti-**C**onfiguration **S**elf-**C**onsistent **F**ield
MCTDH	**M**ulti-**C**onfiguration **T**ime-**D**ependent **H**artree
MGPF	**M**ulti-**G**rid **P**OT**F**IT
MLSL	**M**ulti-**L**evel **S**ingle **L**inkage
MNDO	**M**odified **N**eglect **D**iatomic **O**verlap
NN	**N**eural **N**etworks
PAH	**P**olycyclic **A**aromatic **H**ydrocarbon
PES	**P**otential **E**nergy **S**urface

PM7	**P**arametric **M**ethod 7
RHF	**R**estricted **H**artree-**F**ock
ROHF	**R**estricted **O**pen-shell **H**artree-**F**ock
SD	**S**later **D**eterminant
SOP-FBR	**S**um **O**f **P**roducts **F**inite **B**ase **R**epresentation
SOP	**S**um **O**f **P**roducts
SPF	**S**ingle **P**article **F**unction
SPP	**S**ingle **P**article **P**otential
SRP-MGPF	**S**pecific **R**eaction **P**arameters **M**ulti-**G**rid **P**OT**F**IT
SRP	**S**pecific **R**eaction **P**arameters
STO	**S**later-**T**ype **O**rbitals
TDSE	**T**ime-**D**ependent **S**chrödinger **E**quation
TISE	**T**ime-**I**ndependent **S**chrödinger **E**quation
TT	**T**ensor-**T**rain
UHF	**U**nrestricted **H**artree-**F**ock
vdW-TSSCDS	**v**an **d**er **W**aals **T**ransition **S**tate **S**earch using **C**hemical **D**ynamics **S**imulations
vMCG	**v**ariational **M**ulti-**C**onfigurational **G**aussian

Chapter 1

Introduction

> *Ainsi se formèrent ces immenses couches de charbon qu'une consommation excessive doit pourtant épuiser en moins de trois siècles si les peuples industrieux n'y prennent garde.*
>
> Jules Verne, *Voyage au centre de la Terre*

The present work aims at developing the tools for the full quantum simulation of the processes of adsorption and photoreactivity of the NO_2 molecule interacting with soot particles in atmospheric conditions. The considered problem is extremely challenging from both the experimental and theoretical points of view [15–17], and approximations are mandatory in order to obtain the desired observables. Consequently, we have modeled the soot particles as large Polycyclic Aromatic Hydrocarbons (PAHs). The previous might look as a somehow drastic simplification, but it is justified up to some extent by the physicochemical properties of soot (*vide infra*). On the other hand, the theoretical value of our novel methods is elevated, as it paves the way to further refinements and establishes a work methodology for treating complex systems. In the following pages, the motivations and relevance of the study will be discussed in detail.

1.1 The atmospherical problem

Soot particles are major pollutants with a huge impact in environmental conditions and human health [18–20]. Despite its short lifetime in the atmosphere ($\sim$ 1 week [21]), soot has been recognized as a major contributor to radiative forcing, second only to CO_2 in greenhouse effect [22]. Although it must be mentioned that recent studies seem to challenge this previous claim [23]. Soot particles have also been recently linked to a lowering in cognitive development in children on regions of elevated urban pollution [24].

In addition to this, their chemical composition confers them with a strong carcinogenic character [25]. In fact, the first report of *occupational cancer* in history (the Chimney sweeps' carcinoma) was filed in 1775 by Percivall Pott and is strongly related to soot exposure [26, 27]. Long term inhalation of soot particles has been also associated with a large number of respiratory and heart diseases [28].

Soot (a term that is not always exchangeable with black carbon) is a carbon rich material that condenses from the vapor phase of incomplete combustion processes [3, 29]. These processes are not yet completely understood, but the actual consensus is that they start with the thermal cracking of the fuels (in internal combustion engines for example), followed by the formation of free radicals, which in turn generate the first PAHs that will engage a complicated nucleation and surface growth process (see Reference [30] for a detailed description). From a physicochemical perspective, soot particles are complex clusters of nanospheres with diameters typically lower than 100 nm, formed by concentric layers of graphene-like materials [29].

Morphological studies with high resolution Transmission Electron Microscopy (TEM) confirm that these nanospheres are poorly organized and have an onion-like form [31]. Inside the soot particles, the carbon atoms belonging to the same layer are closer to each other than in graphite, and there is strong evidence that hydrogen is part of their composition [32], which can justify the presence of aromatic compounds (like PAHs) in their structure. This last fact has a great influence in the chemical inhomogeneities of the particles, as aromatic rings can break the periodicity of graphene and create defects on the material. This lack of morphological constraint is responsible for the spherical folding of the different layers on the particle [29]. The large specific surface area of fractal structure of soot particles turns them into exceptionally efficient substrates for heterogeneous reactions of adsorbed atmospheric molecules [15, 17, 33].

There are dozens of molecular species whose concentration in the atmosphere is constantly monitored due to their environmental impact [34]. Between them, the strongest oxidants are O_3, NO_3 and OH, largely due to their abundance and elevated chemical potentials [35]. Interestingly, the three of them can be formed by (photo)chemical reactions involving the NO_2 molecule [1, 35]. The study of the aforementioned molecule has a great relevance due to both its theoretical complexity and environmental role. The principal sources of NO_2 are fossil fuel combustion, biomass burning and bacterial activity [36]. Systematic exposure to NO_2 can lead to significant damage in the respiratory system [37]. This free radical is also a major contributor to one of the most dire natural phenomena: the acid rain [36].

Due to the above facts, significant efforts have been devoted to the study of soot as substrate for heterogeneous reactions with NO_2 [15–17], particularly in the process of formation of HONO after the first evidence of this possibility [38]. Since then, numerous experimental studies have shown that different types of soot can act as substrate for the NO_2 interconversion in HONO [1]. Furthermore, it has been also shown that NO_2 induces modifications (*ageing*) of surface PAHs leading to the formation of nitro-PAHs, compounds which possess a much higher potential of carcinogenicity and mutagenicity than their non-nitro PAH precursors [39].

The study of the systems here presented is one of the major research topics of several experimental groups at the University of Lille. As a result of of this, combined efforts have materialized into large scale projects as the Laboratory of Excellence CaPPA[1] (Chemical and Physical Properties of the Atmosphere) and CLIMIBIO[2], whose wide scope includes the description of the heterogeneous chemistry of soot particles and PAHs possibly interacting with atmospheric molecules. Examples of recent publications on the field come from the groups of Pascale Desgroux [40], Denis Petitprez [41] and Cristian Focsa [42]. The interested reader is invited to consult the official sites of the projects for a detailed publication list.

1.2 The physicochemical problem

The electronic structure of NO_2 has been described from the experimental and theoretical points of view [43–47, and references therein]. The molecule has conical intersections between the PESs of the two lowest electronic states, X^2A_1 and A^2B_2 [48], which in turns produces strong vibronic interaction between the dense set of highly excited vibrational levels of X^2A_1 and the low lying vibrational levels of A^2B_2. A global *ab initio* PES for the ground and first excited state was obtained by Schinke *et al.* at the MRCI level, which was used to perform nuclear quantum dynamical calculations [49, 50]. In addition, some calculations has been carried out using both nonadiabatic wave packet dynamics [51] and the MCTDH method [52].

Fully quantum mechanical simulations of photophysical and photochemical processes involving NO_2 adsorbed on soot (modeled as large PAHs) in atmospheric conditions is a very challenging task. First, there are large discrepancies in the estimated uptake coefficients, thus revealing a lack of understanding of the adsorption process of NO_2 on soot [1]. Second, the series of chemical transformations through which this reaction proceeds is not known yet [1]. In this respect, particularly relevant is the differential night-daytime behavior. Whereas in the dark, the formation of HONO proceeds during a certain amount of time after which the soot surface is deactivated, in the presence

---

[1] `http://www.labex-cappa.fr/en`

[2] `http://climibio.univ-lille.fr/`

of UV-A light this process is dramatically enhanced, and the reactivity of the soot is regenerated so that the process can go on for hours [2]. This experimental fact further highlights the relevance of soot as privileged substrate in this atmospheric process [2]. Subsequent studies have pointed towards this direction by specifically investigating the nature of the different photoproducts (e.g. nitro-compounds) [53] as well as the specific steps in the process [1]. Briefly, it has been concluded that in dark conditions the major product is HONO, produced through a redox reaction (first order kinetics). In contrast, in the presence of (simulated) sunlight, reactivity of the soot substrate towards NO_2 takes place, in addition to the previous redox reaction. As a consequence, newly formed nitro-species were shown to lead to further production of HONO [1].

A detailed molecular description of these processes will help us disentangling the observed differential day- and nighttime behavior of this system concerning the production of nitrous acid (HONO) [1, 2]. This problem has been object of sustained research since its first unambiguous detection in the atmosphere [54]. HONO has been shown to play a major role in modulating the composition of the troposphere since it is precursor of the hydroxyl radical (OH), a primary atmospheric oxidant [3]. The concentration of HONO in the troposphere depends on the balance between its cycles of formation and destruction. The HONO generated during the night is stored and subsequently photochemically decomposed into NO and OH in the early morning. With respect to its formation, it is well-established that HONO can be released either from soil [55] or be formed by heterogeneous reaction of NO_2 on different types of surfaces [56], such as inorganic aerosol particles (e.g. mineral dust, TiO_2 nanoparticles) or organic ones (humic substances, different types of soot). However, after four decades of research, the exact mechanisms for these complex processes are still unclear [57, 58].

Gas-phase atmospheric reactions can be theoretically described to a remarkable degree of precision, however atmospheric heterogeneous (photo)chemical processes are much less understood. In this regard, it is openly recognized a lack of theoretical methods and simulations to which could compare to the vast amount of available experimental results [59]. In this work we aim to fill this gap in the context of the soot-mediated heterogeneous generation of HONO from NO_2. Additionally, the interaction of Pyrene and NO_2 might lead to the formation of 1-Nitropyrene, which is usually a by-product of combustion and the predominant nitrated PAH emitted in a diesel engine[60]. We shall tackle the study of the processes of NO_2 adsorbed on soot particles with particular focus on the differential behavior in night-time (formation) and daytime (destruction) conditions. The surface of the soot particle will be modeled with the Pyrene molecule ($C_{16}H_{10}$), which is the smallest peri-fused PAH and a good starting point for a periodical surface.

1.3 Roadmap for the study of the Pyrene-NO_2 system

As we have mentioned before, our first approximation to describe the interaction between the NO_2 molecule and a soot particle consisted in considering only the surface chemistry. This implied to select a suitable PAH molecule to represent the surface of the soot, which in our case was the Pyrene molecule. That being said, the present work had as main goals:

- The determination of the adsorption sites and transition states connecting them
- The development of a global representation of the interaction potential in the mathematical form necessary to study the nuclear quantum dynamics of the system
- The determination of the probability distribution function (ongoing research)
- The determination of the adsorption energy of the system and the electronic absorption spectrum (ongoing research)

Following the proposed roadmap, the modeling of the Pyrene-NO_2 system started with the determination of the stable configurations (transition states and minima) of the PES. This was achieved using the van der Waals Transition State Search using Chemical Dynamics Simulations (vdW-TSSCDS) method [4], which is the generalization of the TSSCDS algorithm [5, 6] recently developed in our group. It should be noted that the present work represents the first application of vdW-TSSCDS to a system of large dimensionality (81D). Starting from a set of judiciously chosen input geometries, the aforementioned method permits a fully automatic and efficient topographical characterization of the PES, or in other words, the determination of the most stable conformations of the system.

The topographical information obtained with vdW-TSSCDS was used to obtain a global description (fit) of the interaction potential, necessary for the dynamical elucidation of the intermolecular interaction (physisorption), spectroscopic properties and reactivity of the adsorbed species. To achieve this last goal, we have developed two different methodologies together with the corresponding software packages. The first one of them is the Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF) algorithm, which is implemented in the SRPTucker package [7]. This method computes chemically accurate (intermolecular) PESs through reparametrization of semiempirical methods, which are subsequently tensor decomposed into Tucker form using MGPF [8]. This software was successfully interfaced with the Heidelberg version of the Multi-configuration Time-Dependent Hartree (MCTDH) package [9, 10]. The second method allows for obtaining the PES directly in the mathematical form required by MCTDH, thence its name

Sum-Of-Products Finite-Basis-Representation (SOP-FBR) [11]. SOP-FBR constitutes an alternative approach to NN-fitting methods. The idea behind it is simple: from the basis of a low-rank Tucker expansion on the grid, we replace the grid-based basis functions by an expansion in terms of orthogonal polynomials. As in the previous method, a smooth integration with MCTDH has been ensured. Both methods were successfully benchmarked with a number of reference problems, namely: the Hénon-Heiles Hamiltonian [12], a global H_2O PES [13], and the HONO isomerization PES (6D) [14].

With the aid of all the above mentioned methods, we tackled the computation of the global PES of the Pyrene-NO_2 system. Suitable coordinate transformation routines were developed in order to map the Cartesian coordinates to internal coordinates (in the appropriate reference frame, see Section 4.2). In the physisorption domain, the evidence collected with vdW-TSSCDS suggested that the geometry of the NO_2 molecule is almost not perturbed in the stationary points with respect to the isolated molecule. This fact has enabled its treatment in a rigid monomer fashion (6D). The PESs will be used to obtain the electronic ground state (GS) and corresponding Zero-Point Energy (ZPE) of the system with MCTDH. The ZPE can offer an accurate estimate of the adsorption energy of the NO_2 molecule over the Pyrene. Additionally, the electronic absorption spectrum of the system will be obtained by computing the sum (weighted by the GS distribution) of the individual vertical excitations of each stationary point. This last step is still ongoing research.

This manuscript is structured as follows: first we have established the context and relevance of the study in the present Introduction; in Chapter 2 we present a detailed description of the methodological framework of the thesis; in Chapter 3 we introduce the two new methods (SRP-MGPF, SOP-FBR) and corresponding publications developed by us in order to obtain PESs in an automated fashion; in Chapter 4 we discuss the application of vdW-TSSCDS and SRP-MGPF to the Pyrene-NO_2 system; finally, we summarize the results of the study in the Conclusions 5 chapter. In addition, three appendixes (A, B and C) describing some of the software developed in this work are included.

Chapter 2

Methodological framework

> *-Et pourtant, ajouta Pencroff, qui montra une certaine difficulté à se résigner, le monde est bien savant ! Quel gros livre, monsieur Cyrus, on ferait avec tout ce qu'on sait ! - Et quel plus gros livre encore avec tout ce qu'on ne sait pas, répondit Cyrus Smith.*
>
> Jules Verne, *L'Ile mystérieuse*

Quantum Mechanics has had a fundamental role in our understanding of natural phenomena since its advent in the early twentieth century. Starting from the pioneering works of Max Planck and Albert Einstein on black body radiation and the photoelectric effect, the theory grew thanks to the contributions of Niels Bohr, Louis de Broglie, Erwin Schrödinger, Werner Heisenberg, Paul A. M. Dirac, Douglas Hartree, Richard P. Feynman, John von Neumann and many others. In less than a century, Quantum Mechanics became the source of a significant part of the technological development that shapes modern life [61].

The accurate description of processes occurring at a microscopic scale is only possible using Quantum Mechanics. In concordance, the study of the properties of molecular systems implies the solution of the equations arising from a particular formulation of this theory. In this sense, two main historical pictures were established by Schrödinger and Heisenberg respectively [62]. These two formulations differ on how they treat time evolution: the first one considers a time-dependent state function and static operators, whereas the second one considers a stationary state function and evolving operators. Both formulations were proven to be unitarily equivalent in the Hilbert space by Schrödinger [63], but the rigorous demonstration is attributed to Stone and von Neumann [64, 65]. Additionally, a third formulation that generalizes the action principle of classical mechanics (Lagrangian based instead of Hamiltonian), known as the Path Integral formulation, was later introduced by Richard P. Feynman. [66].

The application of the Quantum Mechanics formalisms to the study of chemical systems is known as Quantum Chemistry[1]. This task implies to obtain a solution to the *many-body problem*, which is in most of the cases impossible (or very unpractical) in an analytic manner. Computational Chemistry is the branch of Quantum Chemistry that uses computer simulations to obtain numerical solutions to these problems.

The following pages are intended to present the theoretical foundations of the Quantum Chemistry methods used to describe molecular systems, with particular emphasis on the ones employed in this work. Fundamental aspects of Electronic Structure, Nuclear Quantum Dynamics and Tensor Decomposition Algorithms will be discussed in a detailed way. Since the Schrödinger picture (Dirac picture if one considers relativistic effects) has been widely adopted in the Quantum Chemistry field, we will assume this formulation in further discussions unless otherwise stated. Due to its succinctness and elegance, all the linear algebra treatments will be presented in Dirac's Bra–ket notation [67]. This chapter is based on the excellent material from Cohen-Tannoudji *et al.* [62], Szabo-Ostlund [68], Jensen [69] and Gatti *et al.* [70], as well as the multiple references included in the text.

2.1 The adiabatic and Born-Oppenheimer approximations

The time evolution of any molecular system is ruled by the Time-Dependent Schrödinger Equation (TDSE):

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi \tag{2.1}$$

where $\hat{H}$ is the molecular Hamiltonian and (Ψ) is the *wave function* of the system, which is hereafter understood according to the *Copenhagen interpretation*. The latter postulates that the description of nature at a microscopic level is given by probability functions that do not refer to underlying microscopic realities, but rather to the macroscopic object of experience [71]. In simple words, the Copenhagen interpretation considers that the wave function contains all the information about a system before an observation, without any other "hidden variable", and that every point of the configuration space $(x_1, \ldots, x_n)$ has an associated probability density proportional to the square of the wave function $|\Psi(x_1, \ldots, x_n)|^2$. The interpretation itself has some paradoxical consequences, like the Schrödinger's cat *Gedankenexperiment* [72].

Two different (though closely related) initial assumptions can be made in order to solve the TDSE[2], i.e. the adiabatic and Born-Oppenheimer approximations. These terms are often confused and used indistinctly due to discrepancies in literature. In this work, we

---

[1] Although some authors relate the term with Electronic Structure calculations only.

[2] One can arrive to similar conclusions starting from the Time Independent Schrödinger Equation and using a slightly different process, see for instance Section 3.1 of reference [69].

have followed the analysis and conventions proposed by Worth and Cederbaum [73]. We will start the treatment by defining the non-relativistic (time-independent) molecular Hamiltonian ($\hat{H}$) as:

$$\hat{H}(\mathbf{R},\mathbf{r}) = \hat{T}_n(\mathbf{R}) + \hat{T}_e(\mathbf{r}) + \hat{U}(\mathbf{R},\mathbf{r}) \tag{2.2}$$

In the above expression, $\hat{T}_n$ and $\hat{T}_e$ are the kinetic energy operators of the nuclei and electrons respectively, which in atomic units[3] are given by:

$$\begin{aligned} \hat{T}_n &= -\frac{1}{2}\sum_\alpha \frac{1}{M_\alpha}\boldsymbol{\nabla}^2_\alpha \\ \hat{T}_e &= -\frac{1}{2}\sum_i \boldsymbol{\nabla}^2_i \end{aligned} \tag{2.3}$$

where the index i runs for the electrons and α runs for the nuclei, and M_α represents the ratio of the mass of nucleus α with respect to the mass of the electron. The potential energy function $\hat{U}$ depends on the coordinates of the nuclei ($\mathbf{R}$) and electrons ($\mathbf{r}$), and is has three main contributions:

$$\hat{U}(\mathbf{R},\mathbf{r}) = -\sum_i\sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_i\sum_{j>i}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_\alpha\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \tag{2.4}$$

where Z_α represents the atomic number of nucleus α. The electronic Hamiltonian can be obtained as a result of neglecting the kinetic energy of the nuclei:

$$\hat{H}_e(\mathbf{r};\mathbf{R}) = \hat{T}_e(\mathbf{r}) + \hat{U}(\mathbf{r};\mathbf{R}) \tag{2.5}$$

It is important to notice that the previous operator depends parametrically on the nuclei coordinates $\mathbf{R}$. The spectral decomposition of this Hamiltonian can be found by solving the electronic Time-Independent Schrödinger Equation[4]:

$$\hat{H}_e\Phi_i(\mathbf{r};\mathbf{R}) = E_i(\mathbf{R})\Phi_i(\mathbf{r};\mathbf{R}) \tag{2.6}$$

Since the Hamiltonian operator is Hermitian (self-adjoint), the eigenvectors $\Phi_i(\mathbf{r};\mathbf{R})$ form a complete orthonormal set and depend also parametrically on the nuclear coordinates. The eigenvalues $E_i(\mathbf{R})$ are real and form the Potential Energy Surface (PES) of the system. The total wave function can be defined using the Born representation [74],

---

[3] A system of natural units of measurement especially convenient for Quantum Chemistry. In atomic units four fundamental constant are unity by definition: the unit of action is the reduced Planck constant $\hbar$, the unit of charge is the elementary charge e, the unit of length is the Bohr radius a_0, and the unit of mass is the electron mass m_e.

[4] This equation will be discussed in more detail in the next section. For *stationary states*, the equation can be obtained from the TDSE by separation of variables.

which makes use of the eigenvectors of Equation 2.6:

$$\Psi(\mathbf{R},\mathbf{r}) = \sum_i \chi_i(\mathbf{R})\Phi_i(\mathbf{r};\mathbf{R}) \tag{2.7}$$

where the $\chi_i(\mathbf{R})$ are nuclear functions that act as expansion coefficients. If we introduce the Equation 2.7 into the TDSE, multiply from the left by one particular electronic function (Φ_j), and integrate over the electronic coordinates we obtain the following set of coupled equations on the expansion coefficients:

$$[\hat{T}_n + E_j(\mathbf{R})]\chi_j(\mathbf{R}) - \sum_i \hat{\Lambda}_{ji}(\mathbf{R})\chi_i(\mathbf{R}) = i\hbar\frac{\partial\chi_j(\mathbf{R})}{\partial t} \quad j = 1,\ldots,N \tag{2.8}$$

where N is the number of states and the matrix elements $\hat{\Lambda}_{ji}$ are the non-adiabatic coupling operators:

$$\hat{\Lambda}_{ji} = \delta_{ji}\hat{T}_n - \langle\Phi_j|\hat{T}_n|\Phi_i\rangle \tag{2.9}$$

In the above expression we have dropped the explicit dependence on the nuclear coordinates for simplicity in the notation. The coupling operators describe the dynamical interaction between the electronic and nuclear motions. They can be rewritten in a more convenient way after a somehow elaborated algebraic procedure[5], which leads to the more compact expressions:

$$\hat{\Lambda}_{ji} = \frac{1}{2M}(2\mathbf{F}_{ji}\cdot\nabla + G_{ji}) \tag{2.10}$$

where we have introduced the derivative coupling vector:

$$\begin{aligned}\mathbf{F}_{ji} &= \langle\Phi_j|\boldsymbol{\nabla}\Phi_i\rangle \\ &= \frac{\langle\Phi_j|\boldsymbol{\nabla}\hat{H}_e|\Phi_i\rangle}{E_i - E_j} \text{ for } i \neq j\end{aligned} \tag{2.11}$$

and the scalar coupling:

$$G_{ji} = \langle\Phi_j|\nabla^2\Phi_i\rangle \tag{2.12}$$

The Born-Oppenheimer approximation will hold when the diagonal nonadiabatic coupling term is added to the potential energy surface provided by the electrons:

$$[\hat{T}_n + E - \hat{\Lambda}_{jj}]\chi_j(\mathbf{R}) = i\hbar\frac{\partial\chi_j(\mathbf{R})}{\partial t} \tag{2.13}$$

The previous expression is a direct consequence of using a single product in the Born expansion:

$$\Psi(\mathbf{R},\mathbf{r}) = \chi(\mathbf{R})\Phi(\mathbf{r};\mathbf{R}) \tag{2.14}$$

---

[5] The interested reader is invited to consult the material from references [73, 75, 76]

In the original development, Born and Oppenheimer used perturbation theory to justify the previous *ansatz* [73, and references therein]. Furthermore, the separability of the electronic and nuclear motion is justified by the large difference of electronic and nuclear masses (see Equation 2.10). The adiabatic approximation will hold when the nonadiabatic term is completely neglected:

$$[\hat{T}_n + E]\chi_j(\mathbf{R}) = i\hbar\frac{\partial\chi_j(\mathbf{R})}{\partial t} \tag{2.15}$$

Throughout our treatment we will consider that the Born-Oppenheimer approximation is in use, though it is now clear than in many important cases (for example, in the study of the photochemistry of polyatomic molecules) this assumption breaks. This becomes particularly intuitive if we consider Equation 2.11, in which the denominator approaches infinity in the vicinity of a degeneracy. If two PESs meet at one point (for example in a conical intersection [76]), the approximation becomes meaningless. To summarize, the Born-Oppenheimer approximation leads to the possibility of computing a PES with which quantum dynamics calculations can be performed by propagating a nuclear wave packet. The previous problem will be discussed in detail in the next sections.

2.2 Electronic Structure

Electronic Structure theory deals with the distribution of the energy levels of bounded electrons in atomic and molecular systems, defining many of the distinctive properties of them. Most of the Electronic Structure machinery is designed to find the numerical solution of the electronic Time-Independent Schrödinger Equation (TISE):

$$\hat{H}\Psi = E\Psi \tag{2.16}$$

This notorious linear partial differential equation can be solved by finding the discrete spectrum of the self-adjoint Hamiltonian operator ($\hat{H}$) in the Hilbert space. The eigenvalues of the equation (E) correspond to the energy levels of the system. A distinction should be made on the nature of the methods employed to solve Equation 3.1. If only first principles assumptions are used in the process, the methods are called *ab initio*. On the other hand, if some experimental data is used then they are called *semiempirical*. In a somehow more diffuse region we find the Density Functional Theory methods, because despite making use of (in some cases) first principles assumptions, they are separated from the *ab initio* wave function methods due to historical reasons [77]. They are also typically *ad hoc* optimized using set of empirical data or sets of *ab initio* reference calculations. In the following sections methods on the three categories (that have been used in this work) will be presented, starting from the seminal Hartree-Fock approximation.

2.2.1 The Hartree-Fock approximation

The TISE can be solved analytically only for one-electron systems. For the vast majority of practical cases, approximations should be made in order to obtain a numerical solution. In the previous section, the non-relativistic Born-Oppenheimer approximation was introduced. The above serves as a starting point for the next group of approximations: mean-field approaches. In the context of Quantum Chemistry, the most important of these approaches is arguably the Hartree-Fock method [68, 78, 79], whose particularities will be discussed in this section.

Before going into details of the Hartree-Fock approximation, some contextual facts should be presented. First of all, since we are restricting our analysis to the non-relativistic case, the electron spin has to be introduced as an *ad hoc* quantum effect. Electrons are particles with spin quantum number $\pm 1/2$, and as it was shown by the Stern-Gerlach experiment [80], the presence of a magnetic field deflect them into two possible states. Another cornerstone of the theory is the concept of Molecular Orbital (MO) introduced by Mulliken [81] in 1932. The fact that a direct relationship with experimental quantities can be established for MOs (see Koopmans' theorem [82]) made them a *de facto* approach in Quantum Chemistry in general.

In the Molecular Orbital approximation an electron i is represented by a spin-orbital function:

$$\chi_i(\mathbf{x_i}) = \psi_i(\mathbf{r_i})\sigma(\omega_\mathbf{i}) \tag{2.17}$$

where we have introduced the grouped coordinates $\mathbf{x_i} = \{\mathbf{r_i}, \omega_\mathbf{i}\}$, the spatial orbitals $\psi_i(\mathbf{r_i})$ as a function of the electronic coordinates $\mathbf{r_i}$ and the function $\sigma(\omega_\mathbf{i})$ depending on the spin coordinates $\omega_\mathbf{i}$. There are two possible values for the latter, namely $\alpha(\omega_\mathbf{i})$ and $\beta(\omega_\mathbf{i})$, in concordance with the properties of the electron spin. Without any loss of generality, the spin-orbitals can be considered to form a complete orthonormal basis set:

$$\langle\chi_i|\chi_j\rangle = \delta_{ij} \ \wedge \ \mathbb{1} = \sum_i |\chi_i\rangle\langle\chi_i| \tag{2.18}$$

where we have removed the explicit dependence on the grouped coordinates for simplicity. The simplest N-electron wave function that be constructed from a set of N spin-orbitals is called Hartree product:

$$\Psi_{HP} = \prod_{i=1}^{N} \chi_i \tag{2.19}$$

This function however does not fulfill the Antisymmetry Principle (or Pauli's Exclusion Principle), which states that two or more identical fermions (particles with half-integer spin) cannot have the same set of quantum numbers within a particular quantum system simultaneously [83]. To solve this deficiency, an intrinsically antisymmetric *ansatz* can be introduced, the Slater determinant (SD):

$$\begin{aligned}\Psi_{SD} &= \frac{1}{\sqrt{N!}}\begin{vmatrix}\chi_1(1) & \chi_2(1) & \cdots & \chi_N(1)\\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2)\\ \vdots & \vdots & & \vdots\\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N)\end{vmatrix}\\ &= \frac{1}{\sqrt{N!}}\sum_{n=1}^{N!}(-1)^{p_n}\mathcal{P}_n\left\{\chi_1(1)\chi_2(2)\ldots\chi_N(N)\right\}\end{aligned} \tag{2.20}$$

where the index n runs over the $N!$ permutations of the N single particle states, $\mathcal{P}_n$ is the permutation operator and p_n the number of transpositions required to obtain the running permutation. The above determinant can be expressed more succinctly as the symbolic ket:

$$\Psi_{SD} = |\chi_1\chi_2\ldots\chi_N\rangle \tag{2.21}$$

One important remark here is that the usage of a single Slater determinant as a trial wave function constitutes an approximation *per se*. Indeed, in this scheme the electron-electron repulsion is only included as an average effect, making it impossible to retrieve *electron correlation*.

The base of our approximated solutions is the Variational Principle [84], which states that any approximate wave function has an energy above or equal to the exact energy. For a normalized wave function, the energy can be computed as:

$$E = \langle\Psi|\hat{H}_e|\Psi\rangle \tag{2.22}$$

In order to obtain an analytical expression for the energy in Equation 2.22, one can make use of the Slater-Condon rules, which provide a set of recipes to find the matrix elements of one- and two-electron operators. One particularly elegant way of deriving the rules using the Second Quantization formalism can be found in the book of Schirmer [85]. After carrying out the algebraic procedure, the following expression arises (in Physicists' notation):

$$E = \sum_{i}^{N}\langle i|h|i\rangle + \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}(\langle ij|ij\rangle - \langle ij|ji\rangle) + V \tag{2.23}$$

In the right hand side of the above equation, the first term corresponds to the one-electron contribution to the total energy, whereas the second one corresponds to the Electron Repulsion Integrals (ERIs). The third term V is the constant nucleus-nucleus repulsion energy, and the for the purposes of the subsequent analysis it can be obviated. In most non-relativistic numerical calculations, the spin contributions in the equation are integrated out (they act as stubs included in the formulation to bring physical meaning). After this process, we are left out with spatial integrals only. The new expression reads:

$$E = 2\sum_{i}^{N/2} \langle i|h|i\rangle + \sum_{i}^{N/2}\sum_{j}^{N/2}(2\langle ij|ij\rangle - \langle ij|ji\rangle) \tag{2.24}$$

We can further simplify our notation by introducing the average kinetic and nuclear attraction energies of an electron (h_{ii}), and the *Coulomb* (J_{ij}) and *Exchange* (K_{ij}) integrals:

$$\begin{aligned} h_{ii} &= \langle i|h|i\rangle = \int d\mathbf{r_1}\, \psi_i^*(\mathbf{r_1})(-\frac{1}{2}\boldsymbol{\nabla}_1^2 - \sum_A \frac{Z_A}{r_{1A}})\psi_i(\mathbf{r_1}) \\ J_{ij} &= \langle ij|ij\rangle = \int d\mathbf{r_1} d\mathbf{r_2}\, |\psi_i(\mathbf{r_1})|^2 r_{12}^{-1} |\psi_j(\mathbf{r_2})|^2 \\ K_{ij} &= \langle ij|ji\rangle = \int d\mathbf{r_1} d\mathbf{r_2}\, \psi_i^*(\mathbf{r_1})\psi_j(\mathbf{r_1}) r_{12}^{-1} \psi_j^*(\mathbf{r_2})\psi_i(\mathbf{r_2}) \end{aligned} \tag{2.25}$$

The first one of the two-electrons integrals accounts for the interaction between the two electronic densities in two centers, hence the name of Coulomb integrals. The Exchange integrals have no direct physical interpretation (in classical mechanics) and their appearance in the expression of the energy is a consequence of the indistinguishability of fermions, which in the single determinantal approach is equivalent to say that the motion of electrons with parallel spin is correlated.

2.2.1.1 The Self-Consistent Field procedure

With the expression of the energy of the single determinantal approach in hand, we can proceed to discuss the details of the Hartree-Fock approximation. The problem consists in minimizing Equation 2.23 with respect to the changes in the orbitals, in accordance with the Variational Principle. The choice of this equation and not Equation 2.24 is justified by the fact that with the former is possible to derive a General Hartree-Fock formalism, that can be extended to restricted and unrestricted approaches. A very convenient way of performing the optimization is the use of the Lagrange's multipliers method. For this purpose, we define the functional $\mathcal{L}$ as:

$$\mathcal{L}[\{\psi_i\}] = \sum_{i}^{N} h_i - \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}(J_{ij} - K_{ij}) - \sum_{i}^{N}\sum_{j}^{N}\lambda_{ij}(\langle\psi_i|\psi_j\rangle - \delta_{ij}) \tag{2.26}$$

where we have made use of the orthonormality of the spatial orbitals and introduced the Lagrange multipliers λ_{ij}. If now we request the functional derivative of $\mathcal{L}$ with respect to ψ_k^* to be zero for all possible values of k, we obtain after some algebra[6]:

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\mathcal{L}[\psi_k^*+\epsilon\delta\psi_k^*]-\mathcal{L}[\psi_k^*]}{\epsilon} &= \int\Bigg[\hat{h}\psi_k(x)+\sum_j\Bigg(\psi_k(x)\int dx'\frac{|\psi_j(x')|^2}{|\mathbf{r}-\mathbf{r}'|} \\ &-\psi_j(x)\int dx'\frac{\psi_j^*(x')\psi_k(x')}{\mathbf{r}-\mathbf{r}'}\Bigg)-\sum_j\lambda_{kj}\psi_j(x)\Bigg]\delta\psi_k^*(x)dx\end{aligned} \tag{2.27}$$

where the dummy integration variables have been set to $\{x, x'\}$ for the sake of simplicity in the notation[7]. By simple inspection of the right hand side of the above equation, we can recognize the expression of the functional derivative inside the integral (see [86] for the mathematical discussion):

$$\frac{\delta\mathcal{L}[\{\psi_i\}]}{\delta\psi_k^*}=\Bigg[\hat{h}+\sum_j(\hat{J}_j[\psi]-\hat{K}_j[\psi])\Bigg]\psi_k(x)-\sum_j\lambda_{kj}\psi_j(x) \tag{2.28}$$

where the Coulomb and Exchange operators have been defined by their action on an arbitrary function $\phi(x)$ as:

$$\begin{aligned}\hat{J}_j[\psi]\phi(x) &= \left(\int dx'\frac{|\psi_i(x')|^2}{|\mathbf{r}-\mathbf{r}'|}\right)\phi(x) \\ \hat{K}_j[\psi]\phi(x) &= \left(\int dx'\frac{\psi_j^*(x')\phi(x')}{|\mathbf{r}-\mathbf{r}'|}\right)\psi_k(x)\end{aligned} \tag{2.29}$$

Consequently, we can define the Fock operator as:

$$\hat{F}[\psi]=\hat{h}+\sum_j(\hat{J}_j[\psi]-\hat{K}_j[\psi]) \tag{2.30}$$

The initial form of the Hartree-Fock equations is then:

$$\hat{F}[\psi]\psi_i(x)=\sum_j\lambda_{ij}\psi_j(x) \tag{2.31}$$

Now, by taking the functional derivative with respect to ψ_k and subtracting the result with respect to the complex conjugate of the above equation, one can show that the matrix of the Lagrange multipliers $\Lambda = \lambda_{ij}$ is Hermitian. This implies that there exists a unitary matrix U that diagonalizes Λ. Using this matrix to transform Equation 2.31, and provided the Fock operator is invariant to a unitary transformation, we arrive to set of canonical General Hartree-Fock equations:

$$\hat{F}[\psi]\psi_i(x)=\varepsilon_i\psi_i(x), \qquad i=1,\ldots,N \tag{2.32}$$

---

[6] The interested reader is warmly encouraged to consult the article from Echenique *et al* [79]

[7] To be more precise, notation is simplified by using $\int dx f(x) \equiv \int_{\mathbb{R}^3\times\mathbb{Z}_2} d\mathbf{x} f(\mathbf{x}) = \sum_{\sigma\in\mathbb{Z}_2}\int_{\mathbb{R}^3} f(\mathbf{r},\sigma)d\mathbf{r}$

This set of N coupled non-linear integro-differential equations does not have analytical solution [79], and since the Fock operator depends on the own solutions of the system, an iterative process needs to be applied. This process starts by proposing a starting guess for the set of N spin-orbitals $\{\psi_i^0\}$. The Fock operator $\hat{F}[\psi_i^0]$ is then constructed using the guess spin-orbitals and the set of N coupled equations is solved as an eigenvalue problem. Afterwards, the $\{\psi_i^1\}$ corresponding to the N lowest eigenvalues ε_i^1 are used to form a new Fock operator $\hat{F}[\psi_i^1]$. This process is iterated until a given termination condition is met. The solutions at this point are called self-consistent, and the process as a whole is known as a *self-consistent field* (SCF). The self-consistent field is an NP-complete problem [87] and its solution scales to $O(M^3)$, being the most important bottleneck the computation of the ERIs, which scales to $O(M^4)$, where M is the number of basis functions used.

2.2.1.2 The Roothaan-Hall equations

As a last step in our discussion about the Hartree-Fock method, we will address one of the most common approaches for solving the set of Equations 2.32. Although this equations could be solve by finite differences, this is not very practical for medium-large systems, and an extra approximation must be made. The later was proposed by Roothaan [88] and Hall [89] independently the same year, and it proposes a discretization of the Hilbert space. In other words, the spatial orbitals are approximated by a finite sum of K known *basis functions* $\{\phi_\mu\}$:

$$\psi_i = \sum_{\mu}^{K} C_{\mu i}\phi_\mu \tag{2.33}$$

By introducing the previous expression into Equation 2.32, and multiplying from the left by ϕ_ν we arrive at the Roothaan-Hall equations:

$$\sum_{\mu} F_{\nu\mu}C_{\mu i} = \varepsilon_i \sum_{\mu} S_{\nu\mu}C_{\mu i} \tag{2.34}$$

where we have introduced the *Fock matrix* and the *overlap matrix* defined as:

$$\begin{aligned} F_{\nu\mu} &= \langle \phi_\mu | \hat{F}[\phi] | \phi_\nu \rangle \\ S_{\nu\mu} &= \langle \phi_\mu | \phi_\nu \rangle \end{aligned} \tag{2.35}$$

As it can be appreciated, the problem now has been reduced to find the set coefficients $\{C_{\mu i}\}$ that minimizes the expression of the energy. Since no orthogonality requirements were set for the basis functions (which are *atom centered*), the overlap matrix is in general not diagonal. The final matrix form of the equations reads:

$$\boldsymbol{FC} = \boldsymbol{SCE} \tag{2.36}$$

An optional extra step can be preformed in order to simplify the previous expression, and it is know as symmetric orthogonalization (or Löwdin orthogonalization [90]). After this transformation the equation becomes:

$$\tilde{\boldsymbol{F}}\tilde{\boldsymbol{C}} = \tilde{\boldsymbol{C}}\boldsymbol{E} \tag{2.37}$$

where we have introduced the following matrices:

$$\begin{aligned} \tilde{\boldsymbol{F}} &= \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{S}^{-\frac{1}{2}} \\ \tilde{\boldsymbol{C}} &= \boldsymbol{S}^{\frac{1}{2}}\boldsymbol{C} \end{aligned} \tag{2.38}$$

The previous treatment is completely analogous to a change of basis set of the form:

$$\tilde{\psi}_i = \sum_k S_{\mu i}^{-\frac{1}{2}}\phi_\mu \tag{2.39}$$

where the $\{\tilde{\psi}_i\}$ are obviously not centered in any atom. The notation used so far can be made more convenient if we introduce the density matrix $\boldsymbol{P}$ as:

$$P_{\mu\nu} = 2\sum_i^{N/2} C_{\mu i}C_{\nu i}^* \tag{2.40}$$

With this new definition, the Fock matrix can be rewritten (omitting the label of the spatial orbitals for simplicity) as:

$$\begin{aligned} F_{\mu\nu} &= \langle\mu|h|\nu\rangle + \sum_{\lambda\sigma} P_{\lambda\sigma}[\langle\mu\nu|\lambda\sigma\rangle - \frac{1}{2}\langle\mu\sigma|\lambda\nu\rangle] \\ &= H_{\mu\nu}^{core} + G_{\mu\nu} \end{aligned} \tag{2.41}$$

where we have defined the one ($\boldsymbol{H}$) and two electron ($\boldsymbol{G}$) matrices.

The model we have presented so far serves as a base for further developments, in which additional constraints may be imposed to the spin-orbitals. The first scenario is precisely the absence of restrictions on the form of the spatial orbitals, and it is known as Unrestricted Hartree–Fock (UHF). If we impose that each spatial orbital should have two electrons (one α and one β), then we obtain the Restricted Hartree–Fock (RHF) formalism, which is normally applied to *closed shell* systems. It is important to notice that open-shell systems can also be treated by a restricted formalism, if we force the spatial part of the doubly occupied orbitals to be the same, in what is known as Restricted Open-shell Hartree–Fock (ROHF). One drawback of the last approach is that it is not possible to choose a unitary transformation that makes the matrix of Lagrange multipliers diagonal, so orbital energies are not uniquely defined, breaking the applicability of

Koopmans' theorem.

Despite its unquestionable interest both conceptually an practically, the Hartree-Fock approach in any of its forms is very limited. Indeed, the electron repulsion is described as the interaction of a given electron with the mean field generated by the rest of the electrons, and as a consequence of the single determinantal approach, the motion of electrons of with parallel spins is correlated (a *Fermi hole* exists around them) but not the motion of electrons with opposite spin. Consequently, the Hartree-Fock energy constitutes an upper limit to the exact non-relativistic total energy, and the difference between them is called *correlation energy*:

$$\begin{aligned} E_{corr} &= E_0 - E_{HF} \\ &= E_{static} + E_{dynamic} + E_{dispersion} \end{aligned} \tag{2.42}$$

One of the main objectives of the Electronic Structure theory is to design methods that retrieve as much correlation energy as possible. One can identify three contributions to this energy: the static (or non-dynamic) correlation which is the energetic contribution resulting from the multi-configurational character of the wave function, the dynamic correlation which arises mainly because of the interelectronic cusp conditions (see Kato's Theorem [91]), and the dispersion correlation, which is only relevant in the long range and arises from instantaneous multipole-multipole interactions between electrons on different centers [92].

The Hartree-Fock approaches serves however as the branching point [69] for further approximations (like semiempirical methods) or more sophisticated correlation-retrieving methods (usually called *Post Hartree-Fock methods*), which at the same time are divided into Multi-configurational and Perturbative approaches (or a combination of them). For the sake of completeness, it should be mentioned that methods like variational Monte Carlo (VMC) modify the Hartree-Fock *ansatz* by multiplying it by a function called Jastrow factor, which explicitly contains the the physics of electron correlation [93]. Although computationally equivalent to the Hartree-Fock method, Density Functional Theory (DFT) approaches can be considered as improvements on the former, providing significantly better results. In the following sections, some of the previously mentioned methods (the ones having direct relevance to this work) will be discussed in more detail.

2.2.2 Basis sets and Basis Set Superposition Error

Before entering into details about the methods derived from the Hartree-Fock approach, one subject of capital importance to the whole wave function theory needs to be discussed: the selection of the basis functions to represent the molecular orbitals and, in turn, the wave function. We will base our reasoning in the Roothaan-Hall process introduced in the previous section, but the arguments can be readily extended to Post Hartree-Fock and semiempirical methods. A detailed description of the subject is out of the scope of this work though[8].

One can in principle choose any function ϕ_μ as a basis to solve Equation 2.36. In practice however, the use of chemical intuition can help to speed up the convergence of the process. Indeed, the analytical solutions of the *hydrogen-like* atom can be an excellent starting point, if we think that isolated atoms are not extremely different from atoms in molecules. The spatial wave function for an *hydrogen-like* atom of nuclear charge Z has the form:

$$\phi_{nlm}(r,\theta,\varphi) = \sqrt{\left(\frac{2Z}{n}\right)^{l+5} \frac{(n-l-1)!}{2n[(n+l)!]^3} r^{l+2}} \cdot L_{n-l-1}^{2l+1} e^{-Zr/n} Y_{lm}(\theta,\varphi) \tag{2.43}$$

where n, l, m are the energy, total angular momentum and z-angular momentum quantum numbers respectively. The functions L_{n-l-1}^{2l+1} are the generalized Laguerre polynomials and $Y_{lm}(\theta,\varphi)$ are the spherical harmonics. One of the first basis sets employed to solve molecular problems, the Slater-type Orbitals [95] (STO), was designed to resemble the behavior of ϕ_{nlm}. The STOs are defined as:

$$\phi_i^{STO}(\mathbf{r};\mathbf{R}_{\alpha_i}) = \mathcal{N}_i^{STO} \tilde{Y}_{l_i m_i}^{c,s}(\theta_{\alpha_i},\varphi_{\alpha_i}) |\mathbf{r}-\mathbf{R}_{\alpha_i}|^{n_i-1} e^{-\zeta_i|\mathbf{r}-\mathbf{R}_{\alpha_i}|} \tag{2.44}$$

where $\mathcal{N}_i^{STO}$ is a normalization constant and ζ_i is known as the Slater exponent. The pass from Equation 2.43 to Equation 2.44 implies the inclusion of explicit dependency with the nucleus at which the function is centered (index α_i). Additionally, only the leading term of the Laguerre polynomial is kept and the real spherical harmonics ($\tilde{Y}_{l_i m_i}^{c,s}$) substitute the standard spherical harmonics (c and s stand for cosine and sine respectively).

The STOs have an excellent physical behavior. They present the *cusp* required by Kato's Theorem and also tend asymptotically to zero when the separation between the nuclei and electrons is large. Unfortunately, they are not convenient from the computational point of view. Indeed, the four center electronic integrals arising from the STOs cannot be computed analytically, and this can be a major drawback in terms of accuracy of the calculations. To circumvent the previous issue, the Cartesian Gaussian-type orbitals (GTO) were introduced:

$$\phi_i^{GTO}(\mathbf{r};\mathbf{R}_{\alpha_i}) = \mathcal{N}_i^{GTO} (r_x - R_{\alpha_i,x})^{l_i^x} (r_y - R_{\alpha_i,y})^{l_i^y} (r_z - R_{\alpha_i,z})^{l_i^z} e^{-\zeta_i|\mathbf{r}-\mathbf{R}_{\alpha_i}|^2} \tag{2.45}$$

where the subscripts x, y, z represent the Cartesian components of the electronic and nuclear position vectors, and l_i^x, l_i^y, l_i^z are the so called orbital quantum numbers. Even

---

[8] The interested reader is invited to consult, for example, the Chapter 2 of reference [94]

when GTOs do not exhibit an accurate physical behavior (i.e. they do not fulfill the conditions of Kato's Theorem), they have become a standard choice in SCF calculations due to their enormous computational advantages, as all the integrals arising from their use are analytical. To further improve the Physics of this basis functions, one can employ linear combinations of them. The newly defined functions are then named contracted GTOs or cGTOs. Starting from the cGTOs, a huge variety of basis functions have been introduced over the years, all with different sizes and qualities. The definition and selection of the appropriate basis function for the system in hand has in all cases a strong heuristic component [94].

The simplest cGTO basis set that can be defined contains only the minimal number of functions required to accommodate all of the electrons and at the same time keep the overall spherical symmetry, hence the name *minimal basis*. Examples of the above are the STO-KG basis sets, were $K = 2 - 6$ represents the number of primitive GTOs used to describe a single STO. One can further increase the number of functions in the contraction to provide additional shells of valence, including p and d functions that are useful to describe anisotropic effects in molecules. Since the effect of inner-shell electrons is less significant, it is customary to increase only the number of basis functions representing the valence region. Example of these basis sets is the *split-valence* 6-31G, which belongs to the family of the Pople[9] basis sets. Here the number six indicates that the each inner-shell atomic orbital is described by a single function, that in turns is conformed by 6 GTOs. The amount of numbers after the hyphen indicates how many cGTOs describe each valence shell orbital, and the corresponding numbers dictate the respective amount of primitive GTOs. For modern calculations, *split-valence* basis sets are still not good enough, and one needs to improve even more their structure. An example of such improvements is the inclusion of highly diffuse functions (designed by a "+" in the name of the basis set), which account for the correct description of electron affinities, Rydberg states and binding energies of van der Waals complexes. In addition, higher angular momentum functions might be included to deal with the polarization effects.

Another kind of basis sets (commonly used in Post Hartree-Fock calculations) are the so called *correlation consistent* ones, which are designed to retrieve electronic correlation in a systematic way [94, and references therein]. They are usually named using the convention (aug-)cc-pV{D,T,Q,5,6}Z, were the optional *aug* string indicates the presence of diffuse functions and the label inside curly braces makes reference to the amount of shells in the basis. It is precisely by adding shells of cGTOs to a core set of atomic Hartree-Fock functions that these basis sets are formed. For the sake of completeness, one also has to mention the more recently developed *polarization consistent* basis sets [96], which were

---

[9] In homage to Nobel Laureate John Pople.

designed with similar goals as the their correlation consistent counterparts, but with the additional aim of reducing computational effort. For a very recent study on the performance of both kinds of basis sets[10] in different systems, consult reference [97].

To finalize our discussion, we will inquire on the effect that the mixture of basis sets of different sizes (an consequently, qualities) can have in the overall accuracy of a given calculation. This approach can look very appealing, especially when the system in hand contains atoms of many types. The results can be remarkably wrong though. Indeed, the use of unbalanced basis sets can lead to an unwanted "redistribution" of the cGTOs, in which the atoms with poorer descriptions take basis functions from the remaining ones. The aforementioned imbalance is known as Basis Set Superposition Error (BSSE), and can be dealt with using several approaches [98]. The one with the biggest relevance to this work is the Counterpoise correction [99] (CP). For the particular case of a molecular dimer AB, the CP correction is defined as:

$$\Delta E_{CP} = E(A)_{AB} + E(B)_{AB} - [E(A)_A + E(B)_B] \tag{2.46}$$

where the subscript AB implies the use of the basis sets of both molecules, in contrast with the subscripts A and B which suggest the use of only the basis functions of the corresponding molecule. Despite the fact that the BSSE can occur in any imbalanced calculation, the usage of bigger basis sets can greatly reduce its magnitude, which tends to zero at the Complete Basis Set [11] (CBS) limit [97].

2.2.3 Semiempirical Methods

Semiempirical methods appeared in the middle of the last century as an alternative to the more expensive *ab initio* methodologies. Nowadays, the technological development in Computer Science and Physics has granted us with a tremendous (yet not enough) computational power. As a consequence, the original area of application of semiempirical methods has moved from small molecular systems to the study of large biomolecular systems [100] and huge molecular clusters [101]. The family tree of semiempirical methods is large, with a common trunk of main approximations which are extended into several branches. In this particular study, we have focused on the methods of the Modified Neglect of Diatomic Overlap (MNDO) branch, introduced by Dewar and Thiel [102] in 1977. Throughout the discussion, the ideas developed for the Hartree-Fock method will serve as a starting point.

---

[10] These basis sets are also known by the respective names of their creators, i.e. Dunning basis sets and Jensen basis sets.

[11] The Complete Basis Set limit is the defined as the energy obtained with a given method when the the amount of basis functions tends to infinity.

The MNDO method and derivatives[12] find their roots in the Neglect of Diatomic Differential Overlap approximation [104]. The above assumption is shared by many semiempirical methods, and it can be formulated as:

$$\langle \tilde{\psi}_\mu \tilde{\psi}_\nu | \tilde{\psi}_\lambda \tilde{\psi}_\sigma \rangle \approx \delta_{IJ}\delta_{KL} \langle \phi_\mu^I \phi_\nu^J | \phi_\lambda^K \phi_\sigma^L \rangle \tag{2.47}$$

In the previous equation, we have respected the notation introduced in Subsection 2.2.1.2, and the superscripts I, J, K, L make reference to a particular atomic center. As can be inferred by the presence of the Kronecker deltas, the number of ERIs that effectively enters the SCF process gets greatly reduced. The condition in Equation 2.47 is a consequence of an "emulated" symmetric orthogonalization of the SCF equations [103]. The approximation can look drastic at a first glance, but it is heuristically justified by the fact that the lost information will be retrieved by using parameters obtained from experimental (or *ab initio*) data in the model.

In addition to NDDO, two other approximations are usually introduced in MNDO methods. The first one is the use of a minimal (valence-shell only) basis set. The second one is the reduction of the effective number of electrons used in the calculations:

$$Q_I = Z_I - N_{c,I} \tag{2.48}$$

where we have introduced the effective change Q_I and the number of core electrons $N_{c,I}$. As a consequence of both approximations, one can consider that in the description of a given atom there are at most two s, six p and ten d electrons distributed in the corresponding minimal cGTOs. To have a more clear picture of how semiempirical methods diverge from Hartree-Fock, we will gradually introduce the approximations in Equation 2.41. After the aforementioned considerations, we are left with the expression:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma}^{NDDO} [\delta_{IJ}\delta_{KL} \langle \phi_\mu^I \phi_\nu^J | \phi_\lambda^K \phi_\sigma^L \rangle - \frac{1}{2}\delta_{IL}\delta_{KJ} \langle \phi_\mu^I \phi_\sigma^L | \phi_\lambda^K \phi_\nu^J \rangle] \tag{2.49}$$

Moving back to the MNDO method, in its original implementation [102] it only included a minimal description (with all the previously discussed approximations) of the hydrogen, carbon, nitrogen, oxygen, and fluorine. The STO-KG functions were modified to introduce an exponential parameter (ζ^{Z_I}) which depended only in the type of atom in consideration, and whose value was shared by the s and p functions indistinctly. Regarding the ERIs that survived the NDDO approximation, their values are substituted by parameters depending on the amount of centers involved in the integral.

---

[12] See Figure 1 in reference [103] for an overview of the different methods in the MNDO family.

For the one-center case, six unique non-zero integrals exists, and the first five are given by:

$$\begin{aligned}
\gamma_{ss}^{Z_I} &= \langle s^I s^I | s^I s^I \rangle \\
\gamma_{pp}^{Z_I} &= \langle p^I p^I | p^I p^I \rangle \\
\gamma_{sp}^{Z_I} &= \langle s^I s^I | p^I p^I \rangle \\
\gamma_{pp'}^{Z_I} &= \langle p^I p^I | p'^I p'^I \rangle \\
\tilde{\gamma}_{sp}^{Z_I} &= \langle s^I p^I | s^I p^I \rangle
\end{aligned} \tag{2.50}$$

where the spatial orbitals are labeled with the letters s and p and the gammas are parameters fitted from experimental or *ab initio* data. Primed orbitals (p') have a magnetic quantum number different from the unprimed ones. The sixth once center integral can be computed as:

$$\langle p^I p'^I | p^I p'^I \rangle = \frac{1}{2}(\gamma_{pp}^{Z_I} - \gamma_{pp'}^{Z_I}) \tag{2.51}$$

The two-center case is a bit more complicated. The key here is to understand the integrals as the electrostatic interaction between two densities of charge centered at two different atoms. Depending on the values of the orbital quantum numbers, the interaction can be approximated by a corresponding truncated classical multipole expansion [103]. In simpler words, one can imagine the process of computing a two center ERI as a mapping of the molecular configuration space to a distribution of point charges. In mathematical terms we can formulate these previous ideas as:

$$\langle \phi_\mu^I \phi_\nu^I | \phi_\lambda^J \phi_\sigma^J \rangle \approx \sum_{t=1}^{T_{\mu\nu}} \sum_{s=1}^{T_{\lambda\sigma}} U(\Theta_t^{\mu\nu}, \Theta_s^{\lambda\sigma}) \tag{2.52}$$

where the potential energy U is approximated by the interaction of the $T_{\mu\nu}$ multipoles $\Theta_t^{\mu\nu}$ generating the charge density $\phi_\mu^I \phi_\nu^I$ and the $T_{\lambda\sigma}$ multipoles $\Theta_t^{\lambda\sigma}$ generating the charge density $\phi_\lambda^J \phi_\sigma^J$. For the MNDO method, these multipoles are restricted to be a monopole, a dipole, a linear quadrupole and a square quadrupole. There are some other considerations to take into account to generate the correct point charge distribution (for example, the Coulomb interaction needs to be regularized), and this constitutes by far the most involved section of the MNDO-SCF. For a detailed description of this process, consult the Appendix 10.5 of the excellent review by Husch *et al.* [103].

In the MNDO framework, the one-electron integrals are also greatly simplified. If only one atomic center is taken into account, then the first addend on Equation 2.49 can be written as:

$$H_{\mu\nu}^{core} = \langle \phi_\mu^I | -\frac{1}{2}\boldsymbol{\nabla}^2 | \phi_\nu^I \rangle - \langle \phi_\mu^I | \frac{Q_I}{|\mathbf{r}_i - \mathbf{R}_I|} | \phi_\nu^I \rangle - \sum_{J \neq I} \langle \phi_\mu^I | \frac{Q_I}{|\mathbf{r}_i - \mathbf{R}_J|} | \phi_\nu^I \rangle \tag{2.53}$$

The corresponding semiempirical approximation to the above equation is:

$$H_{\mu\nu}^{MNDO} = U_{l(\mu)l(\nu)}^{Z_I} - \sum_{J \neq I} Q_J [\phi_\mu^I \phi_\nu^I | s^J s^J] \tag{2.54}$$

where we have introduced the parameters $U_{l(\mu)l(\nu)}^{Z_I}$, which approximate the first two addends of Equation 2.53 (one-center one-electron contribution) and depends on the type of orbital in consideration. The third addend is equivalent to the electrostatic interaction between the charge distribution $\phi_\mu^I \phi_\nu^I$ and all atomic centers but I, and it is approximated by the ERI $[\phi_\mu^I \phi_\nu^I | s^J s^J]$ weighted by the charge Q_J. The usage of square brackets here indicates that the integrals have been computed using the Klopman formula [103].

Finally, when two atomic centers are considered the MNDO one-electron matrix is assumed to be proportional to the overlap matrix[13]:

$$H_{\mu\nu}^{MNDO} = \frac{\beta_{l(\mu)}^{Z_I} + \beta_{l(\nu)}^{Z_J}}{2} S_{\mu\nu} \tag{2.55}$$

where we have introduced two additional parameters $\beta_{l(\mu)}^{Z_I}$ depending on both the atom and orbital type.

One last approximation made in the MNDO models is the empirical modification of the core-core repulsion energy. The expected Coulomb interaction energy is substituted by:

$$\begin{aligned} V^{MNDO} &= \sum_I \sum_{J>I} Q_I Q_J [s^I s^I | s^J s^J] f_{IJ}^{MNDO} \\ &= \sum_I \sum_{J>I} Q_I Q_J \left[R_{IJ}^2 + \left(\frac{1}{2\gamma_{ss}^{Z_I}} + \frac{1}{2\gamma_{ss}^{Z_J}} \right)^2 \right]^{-\frac{1}{2}} f_{IJ}^{MNDO} \end{aligned} \tag{2.56}$$

where the point charge interaction is calculated as the charge-weighted interaction of the densities of charge $s^I s^I$ and $s^J s^J$ evaluated in the so called Klopman approximation [102]. The term f_{IJ}^{MNDO} is a suitable scaling factor.

Of particular interest for this work are the MNDO derivatives known as Parametric Methods (PM$_x$), largely developed by James Stewart [105, 106, and references therein]. Starting from the inclusion of d orbitals into the model described so far, the PM$_x$ introduce the important idea that the semiempirical parameters are unique to the type of atom in consideration (i.e. there is a different parameter set for each atom). The PM$_x$ also includes several new parameters with respect to MNDO (notably the element-pair-dependent parameters), and a further modification of the core-core repulsion energy expression[14]. As it can be inferred at this point, one of the most important aspects in

---

[13] Note that in the MNDO method one makes use of the overlap matrix to approximate integrals, but in the SCF equations it is considered to be $\mathbb{1}$.

[14] In the particular case of the PM7 method, the core-core repulsion energy includes a dispersion term.

this family of methods is to have an adequate description of the electrostatic nuclear interaction, which acts as a compensation to the other approximations.

Despite the seemingly extreme character of the assumptions in the MNDO method and derivatives, they have been successfully applied to a wide range of problems over the last few decades [103, and references therein]. They are also prove of the artfulness with which the first quantum chemists needed to tackle the inherent computational limitations of their times, by introducing thoughtful approximations. As we will see further in this work, the behavior of semiempirical methods can be improved even more by performing reparametrizations that target specific families of compounds, i.e. yielding Specific Reaction Parameters (SRPs) Hamiltonians. In Chapter 3 we will present a method developed by us that allows for performing the aforementioned task, the SRP-MGPF[7].

2.2.4 Overview of the Second Quantization formalism

Before passing to the description of electron correlation methods, we will briefly discuss the Second Quantization (SQ) formalism[15]. Lets consider the non-relativistic, Born-Oppenheimer Hamiltonian operator in this new formulation:

$$\hat{H} = \sum_{i,j} h_{ij}\hat{a}_i^\dagger \hat{a}_j + \frac{1}{2}\sum_{i,j,k,l} g_{ijkl}\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k + V \tag{2.57}$$

In the above equation, the terms h_{ij} and g_{ijkl} make reference to the one-electron integrals and the ERIs respectively, and V is the nuclear repulsion energy. We have also introduced the *creation* $\hat{a}_i^\dagger$ and *annihilation* $\hat{a}_i$ operators (or fermion operators), which are defined by their action on the Fock space[16] basis sets:

$$\begin{aligned} \hat{a}_q^\dagger \left|q_1 \dots q_N\right\rangle &= \left|q_1 \dots q_N q\right\rangle \\ \hat{a}_q \left|q_1 \dots q_N\right\rangle &= \sum_{q'_1 < q'_2 < \dots < q'_{N-1}} \left|q'_1 \dots q'_{N-1}\right\rangle \left\langle q_1 \dots q_N \middle| \hat{a}_q^\dagger \middle| q'_1 \dots q'_{N-1}\right\rangle^* \end{aligned} \tag{2.58}$$

In simple terms, the fermion operators generate states with $N+1$ and $N-1$ electrons respectively. The action of the operators in the vacuum state $|\varnothing\rangle$ is given by:

$$\begin{aligned} \hat{a}_q^\dagger \left|\varnothing\right\rangle &= \left|q\right\rangle \\ \hat{a}_q \left|\varnothing\right\rangle &= \mathbf{0} \end{aligned} \tag{2.59}$$

---

[15] An excellent introduction to the topic can be found in Chapter 2 of Reference [85]

[16] The Fock space is defined as the direct sum of N-electron Hilbert spaces $\mathcal{F} = \mathcal{H}_0 \bigoplus \mathcal{H}_1 \dots$

where **0** is the null vector. Finally, the very important anticommutation equalities hold:

$$\begin{aligned} &\{\hat{a}_q^\dagger, \hat{a}_p\} = \hat{a}_p\hat{a}_q^\dagger + \hat{a}_q^\dagger\hat{a}_p = \delta_{pq} \\ &\{\hat{a}_q^\dagger, \hat{a}_p^\dagger\} = 0 \qquad \{\hat{a}_q, \hat{a}_p\} = 0 \end{aligned} \tag{2.60}$$

The use of SQ formalism in Electronic Structure implies a slight change on one's mathematical mindset, but this is arguably its biggest disadvantage. With this new formalism it is possible to consider systems of undetermined number of particles, it provides a natural way of extending mean-field methods to multiconfigurational approaches, and the corresponding computational implementations are notably simpler than in its first quantization counterpart. As we already noticed in Equation 2.57, the electronic integrals are included in the Hamiltonian in a very intuitive and concise manner. The "extreme simplification" of electronic integrals in semiempirical methods can be very useful here, as it ensures a reasonably good accuracy at a very low computational cost, which could allow for treating the electronic and nuclear problem altogether (attochemistry scale) [107].

2.2.5 Wave function electron correlation methods

The landscape of electron correlation methods (or Post Hartree-Fock methods) is vast, with dozens of different methods contained in years of dense literature [68, 69, 108, and references therein]. For the purposes of this work, we will limit our discussion only to key aspects of the methods, leaving in each case the suitable references for a possible in-dept analysis.

We will start our discussion with an example of great historical relevance: the Hylleraas Configuration Interaction (HCI) method [109]. The original idea of the method was to use a trial wave function containing the interelectronic r_{12} distance explicitly. The method had a huge success in the accurate prediction of the ionization potential of the Helium atom, but its application to larger systems was limited by the computational power of the time. The general expression of the wave function is in this case:

$$\Psi_{HCI} = \sum_i C_i \hat{A} \left[\prod_{m,n \in [1..N]} r_{mn}^{u_i^{(mn)}} \Psi_{SD}^{(i)} \right] \tag{2.61}$$

in the above equation $\hat{A}$ is an antisymmetrization operator acting on the the Slater Determinants $\Psi_{SD}^{(i)}$ multiplied by the interelectronic distances r_{mn} of all possible pairs of electrons m, n. The configuration interaction coefficients C_i and the exponents $u_i^{(mn)}$ are the variational parameters in the expansion. If now in Equation 2.61 we set the latter to zero we obtain the expression of the well-know Configuration Interaction *ansatz*:

$$\Psi_{CI} = \sum_i C_i \Psi_{SD}^{(i)} \tag{2.62}$$

If in the above equation all the $\binom{2K}{N}$ possible Slater Determinants are included ($2K$ is the number of spin-orbitals), then the method is called *full* CI (FCI). In the CBS limit, the FCI method is *exact* [110]. Even with the huge computational power that we have available nowadays, the FCI expansion of medium-large system is not doable[17]. The equations for the FCI method can be obtained using the Variational Principle, in a process very similar to the one presented in the Section 2.2.1.1. The notable differences are obviously the form of the wave function, and the constraints used with the Lagrange multipliers. After performing the variation, we are left with the expression [113]:

$$\boldsymbol{HC} = E\boldsymbol{SC} \tag{2.63}$$

where the matrix $\boldsymbol{C}$ contains the configuration interaction coefficients, $H_{ij} = \langle \Psi_{SD}^{(i)} | \hat{H} | \Psi_{SD}^{(j)} \rangle$, and $S_{ij} = \langle \Psi_{SD}^{(i)} | \Psi_{SD}^{(j)} \rangle$. The Hamiltonian in 2.57 is exactly that of a FCI [108]. The matrix elements of $\boldsymbol{H}$ are greatly simplified due to the Slater-Condon rules and the Brillouin theorem[18] [68]:

$$\boldsymbol{H} = \begin{array}{c|cccccc|}
 & |\Psi_0\rangle & |S\rangle & |D\rangle & |T\rangle & |Q\rangle & \cdots \\
\hline
\langle\Psi_0| & \langle\Psi_0|\hat{H}|\Psi_0\rangle & & & & & \cdots \\
\langle S| & 0 & \langle S|\hat{H}|S\rangle & & & & \cdots \\
\langle D| & \langle D|\hat{H}|\Psi_0\rangle & \langle D|\hat{H}|S\rangle & \langle D|\hat{H}|D\rangle & & & \cdots \\
\langle T| & 0 & \langle T|\hat{H}|S\rangle & \langle T|\hat{H}|D\rangle & \langle T|\hat{H}|T\rangle & & \cdots \\
\langle Q| & 0 & 0 & \langle Q|\hat{H}|D\rangle & \langle Q|\hat{H}|T\rangle & \langle Q|\hat{H}|Q\rangle & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array} \tag{2.64}$$

In the above equation we have labeled the Hartree-Fock ground state wave function as Ψ_0, and S, D, T, Q represent the corresponding singly, doubly, triply, and quadruply excited determinants respectively. The previous matrix is also the starting point for several approximations that can be made to solve 2.63, which are in many cases truncations of the expansion 2.62. This approximations are named according to the amount of excitations included in the model, i.e. CIS, CISD, CISDT, and so on. Since only the double excitations interact directly with the ground state energy, they are expected to have the largest contribution to the correlation energy. The energy contribution of quadruple excitations is higher than the one of triple excitations [113]. One significant drawback of

---

[17] To the best of our knowledge, the largest FCI calculation registered is that of the N_2 molecule [111, 112]

[18] $\langle S|\hat{H}|\Psi_0\rangle = 0$. See Chapter 11 of Reference [114] for a rigorous proof of the theorem in SQ formalism.

the truncated approaches is their lack of size-extensivity, i.e. the proper (linear) scaling of the energy of the system with its size. Finally we have to mention that in many implementations of the CI method and derivatives it is customary to use Configuration State Functions (CSFs) instead of Slater Determinants. CSFs are linear combinations of Slater Determinants that are eigenfunctions of the $\hat{S}^2$ and $\hat{S}_z$ spin operators, and belong to the proper irreducible representation of the symmetry group of $\hat{H}$ [115]. By using CSFs one can greatly reduce the amount of computations due to the aforementioned properties.

An improvement can be made to the truncated CI theory if in addition to the configuration interaction coefficients, one also includes the MO coefficients in the optimization with the variational principle. This new approach is called Multi-Configuration Self-Consistent Field (MCSCF) [69]. MCSCF is an iterative method like Hartree-Fock, and it fact, it is equivalent to the latter if the number of configuration is one. The MC-SCF wave functions are obviously smaller than their truncated CI counterparts, and also much more difficult to converge to the global minimum. The optimization of the wave function is normally carried out by expanding the energy to second order in the variational parameters, and using using methods from the Newton-Raphson family [69]. Since MCSCF methods do not excel in retrieving large amounts of correlation (orbital relaxation is not so important as adding configurations), its application is centered in offering qualitatively accurate results where other mono-determinantal approaches would fail.

One of the most successful variants of MCSCF is the Complete Active Space Self-Consistent Field (CASSCF) method [116]. In the latter the selection of the configurations to include in the *ansatz* is done by dividing the MOs in *active* and *inactive* spaces. The former are MOs in the regions of highest occupied and lowest unoccupied, whereas the latter are always either doubly occupied or unoccupied [69]. A FCI calculation is performed with all the MOs in the active space (AS). The problem of selecting the correct AS largely remains in the domains of chemical intuition. As for notation, these calculations are usually labeled $[n, m]-$CASSCF, where n, m are the number of electrons and orbitals respectively. A further approximation can be made in order to improve the computational reach of the CASSCF method, and it is know as Restricted Active Space Self-Consistent Field (RASSCF) [117]. In the latter, the AS is subdivided in three regions (RAS1, RAS2 and RAS3) which incorporate additional restrictions on the occupation numbers allowed. In any of the previous approaches, the size of the AS is very limited, with the maximum size being around $[20, 20]-$CASSCF [118]. A very promising alternative that allows for numerically exact solutions of calculations with AS up to 40 electrons in 40 are the methods from the family of the Density Matrix Renormalization Group (DMRG) [119, and references therein].

Finally, we could also enhance the truncated CI methods by using CSFs that are generated from a MCSCF wave function instead of a single Hartree-Fock reference. This approach is called Multi-Reference Configuration Interaction (MRCI), and it usually uses the CISD truncation [120]. MRCI methods can retrieve a large amount of correlation energy, but at a very high computational cost. Another weak point of the approximation is the lack of size-extensivity.

Another important family of Post Hartree-Fock methods is the one deriving from the Many-Body Perturbation Theory (MBPT). The main idea behind this approaches is that the solution of the problem at hand (retrieving electron correlation) differ only slightly from the problem that has been already solved (for example, the Hartree-Fock wave function) [69]. One of the simplest and most useful method of this kind is the Møller-Plesset Perturbation Theory (MPPT), particularly in its second order expansion (MP2) [121]. In the framework of this theory, the Hamiltonian can be represented as:

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \tag{2.65}$$

where $\hat{H}_0$ is the Hartree-Fock Hamiltonian, λ is an arbitrary parameter, and $\hat{H}'$ is a small perturbation. Under this conditions, the perturbed wave function and energy can be represented as a power series of λ:

$$\begin{aligned} \Psi &= \sum_{i=0}^{\infty} \lambda^i \Psi^{(i)} \\ E &= \sum_{i=0}^{\infty} \lambda^i E^{(i)} \end{aligned} \tag{2.66}$$

where $\Psi_0 \equiv \Psi^{(0)}$, $\Psi_0 \equiv E^{(0)}$ are the Hartree-Fock wave function and energies. If we now plug the previous equation into the TISE, and group terms according to the powers of λ, we get:

$$\begin{aligned} \lambda^0 &\rightarrow \hat{H}_0 \Psi_0 = E_0 \Psi_0 \\ \lambda^{n>0} &\rightarrow \hat{H}_0 \Psi^{(n)} + \hat{H}' \Psi^{(n-1)} = \sum_{i=0}^{n} E^{(i)} \Psi^{(n-1)} \end{aligned} \tag{2.67}$$

If we multiply the previous equation from the left by Ψ_0 and integrate, we obtain the n-th order correction energy:

$$E^{(n)} = \langle \Psi_0 | \hat{H}' | \Psi^{(n-1)} \rangle \tag{2.68}$$

In the MPPT (which is a special case Rayleigh–Schrödinger perturbation theory [122]) we consider the $\hat{H}^{(0)}$ to be equal to a sum of Fock operators, and it can be shown that under this conditions, the energy of the first order correction is zero[19], whereas the second order one is given by [69]:

$$E_{MP2} = -\frac{1}{4}\sum_{ab}^{\text{virt}}\sum_{ij}^{\text{occ}} \frac{|\langle ab|ij\rangle - \langle ab|ji\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{2.69}$$

where the indexes a, b refer to unoccupied (virtual) orbitals, i, j to occupied orbitals, and the ε_i to the corresponding orbital energies. The MP2 method has the same computational complexity as the CIS method, i.e. $O(M^5)$ with M being the number of spin orbitals. However, the results obtained with the former are considerably more accurate [69, 122]. Another advantage of the MPPT over the truncated CI methods is that the former are size extensive. With arguments similar to the ones presented so far, we can obtain the MPPT corrections of order higher than two, which come with great increase on the complexity of the algorithms. The understanding of corrections beyond MP6 is to the date very limited [122]. It is very important to emphasize that even for well behaved closed shell system the MP$_n$ series can diverge [123].

One of the most powerful applications of MPPT (and MBPT in general) is the computation of energy corrections of multi-configurational methods. Example of the above are the CASMP2 and CASPT2 methods [124], which could be improved even more by adding a multi-dimensional reference space that is spanned by two or more state-average[20] CASSCF (SA-CASSCF) states [125]. One drawback shared by all perturbative approaches is that they are not sustained in any variational principle and, even when this rarely happens, one can end with energies lower than the exact one.

The next important group of electron correlation methods are the ones derived from the Coupled Cluster (CC) Theory. The CC methods have their roots in nuclear physics, and were imported to Quantum Chemistry in the second half of last century [126, 127, and references therein]. The main idea behind this theory is the use of an exponential *ansatz* of the form:

$$\Psi_{CC} = e^{\hat{T}}\Psi_0 \tag{2.70}$$

where Ψ_0 is again the Hartree-Fock ground state wave function and $\hat{T}$ is the *cluster operator*. The latter is defined as the sum of cluster operators of different excitation levels:

$$\hat{T} = \sum_{l}^{N} \hat{T}_l \tag{2.71}$$

---

[19] This implies that the Hartree-Fock wave function corrects the energy up to the MP1 level. Similar results can be obtained for dipole moments and electron densities.

[20] In a SA-CASSCF calculation one single set of molecular orbitals is used to compute all the states of a given spatial and spin symmetry.

The $\hat{T}_j$ cluster operator acting on Ψ_0 generates the l-th excited Slater Determinant, which can be represented in SQ as:

$$\begin{aligned}\hat{T}_1 &= \sum_{a}^{\text{virt}}\sum_{i}^{\text{occ}} t_i^a \hat{a}_a^\dagger \hat{a}_i \\ \hat{T}_2 &= \frac{1}{4}\sum_{a,b}^{\text{virt}}\sum_{i,j}^{\text{occ}} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j \\ &\vdots\end{aligned} \tag{2.72}$$

where the coefficients t_i^a, t_{ij}^{ab} are the single and double excitation *amplitudes* respectively. To better understand the action of $\hat{T}$ on the Hartree-Fock wave function, we will expand the exponential in 2.70 in power series:

$$\begin{aligned}e^{\hat{T}} &= \hat{1} + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \cdots \\ &= \hat{1} + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \cdots\end{aligned} \tag{2.73}$$

where the $\hat{C}_l$ operator generates l excitations. Since all of the $\hat{T}_l$ commute between each other, we can define the former operators as:

$$\begin{aligned}\hat{C}_1 &= \hat{T}_1 \\ \hat{C}_2 &= \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 \\ \hat{C}_3 &= \hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3 \\ &\vdots\end{aligned} \tag{2.74}$$

Under these conditions, the CC wave function can be written as:

$$\Psi_{CC} = \Psi_0 + \sum_{a}^{\text{virt}}\sum_{i}^{\text{occ}} c_i^a \Psi_i^a + \frac{1}{4}\sum_{a,b}^{\text{virt}}\sum_{i,j}^{\text{occ}} c_{ij}^{ab}\Psi_{ij}^{ab} + \cdots \tag{2.75}$$

with coefficients:

$$\begin{aligned}c_i^a &= t_i^a \\ c_{ij}^{ab} &= t_{ij}^{ab} + t_i^a * t_j^b \\ &\vdots\end{aligned} \tag{2.76}$$

where the * symbol represents an antisymmetric product making the obtained coefficients antisymmetric with respect to the exchange of two occupied spin orbitals or two virtual spin orbitals [128]. One can then conclude that, from a formal point of view, the CC *ansatz* is equivalent to the FCI one[21].

The computational advantages of CC become evident when dealing with truncated expansions. In this case, the operator $\hat{T}$ is formed by a finite number of cluster operators (typically two). The coefficients 2.76 must be constructed using only amplitudes that correspond with the number of cluster operators in the expansion. In this way, a truncated CC wave function contains much more excited determinants than the corresponding truncated CI one, by keeping the same number of parameters. In addition, it can be shown that, contrary to truncated CI, the truncated CC expansion is size extensive [127].

The deduction of the CC equations differs a bit from the rest of the methods that we have presented so far. Indeed, instead of plugging the wave function 2.75 into the Variational Principle and optimizing the parameters on the expansion, one makes use of the so called *projection method* [128]. This choice is motivated by the fact that equations arising from the application of the variational principle are way to complicated to be solved efficiently [129], although important efforts are being made to circumvent this issue [130, and references therein]. That being said, the following system of equations need to be solved:

$$\begin{aligned} \langle \Psi_0 | \hat{H} - E | \Psi_{CC} \rangle &= 0 \\ \langle \Psi_{\{i\}}^{\{a\}} | \hat{H} - E | \Psi_{CC} \rangle &= 0 \end{aligned} \tag{2.77}$$

where the kets in the second equality are all possible excitations of the Hartree-Fock wave function, formed with the set of occupied ($\{i\}$) and virtual ($\{a\}$) spin-orbitals. If we now introduce 2.75 in the first of the above equations, and apply the Slater-Condon rules and Brillouin's Theorem we get the following expression for the CC energy:

$$E = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle + \frac{1}{4} \sum_{a,b}^{\text{virt}} \sum_{i,j}^{\text{occ}} (t_{ij}^{ab} + 2 t_i^a t_j^b) [\langle ab | ij \rangle - \langle ab | ji \rangle] \tag{2.78}$$

As can be seen, the CC energy is completely determined by the amplitudes of single and double excitations and the ERIs of the system. The remaining equations on 2.77 determine the CC amplitudes, and the are usually solved by performing the transformation:

$$\langle \Psi_{\{i\}}^{\{a\}} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle = 0 \tag{2.79}$$

The previous equations are known as *linked CC amplitude equations*. It is precisely the multiplication from the left by $e^{-\hat{T}}$ what ensures the size-extensivity of CC, because the

---

[21] To be more precise, the CC wave function is a reparametrization of the FCI one.

expansion of $e^{-\hat{T}}\hat{H}e^{\hat{T}}$ exactly terminates at the fourth order[22]

$$e^{-\hat{T}}\hat{H}e^{\hat{T}} = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}] \quad (2.80)$$

In addition to enforce the size-extensivity of the method, the above result greatly reduces the complexity of the equations [127].

CC Theory and MPPT are closely related from the mathematical point of view [69]. Since one of the biggest limitations of the CCSD method is the lack of connected triple excitations, a natural improvement to the method is to include this effect in a perturbative manner. The most popular and successful of this extensions in known as CCSD(T) [131]. When extrapolated to the CBS limit, the latter is known as the *gold-standard in Quantum Chemistry*, and together with MP4, it is one of the few highest-exponent polynomial time algorithms (it scales as $O(M^7)$) that are systematically used [132]. The inherent computational challenge that implies the use of this algorithm can be mitigated by using explicitly correlated basis sets, like in the case of the CCSD(T)-F12 method [133].

Powerful as they are, many of the Post Hartree-Fock methods that we have presented suffer from computational limitations, being the major bottleneck the calculation of the ERIs. In the previous section, we saw that semiempirical methods deal with this issue by introducing rather "drastic" approximations. For highly correlated systems, the accuracy of semiempirical methods might be not acceptable. In the field of electron correlation methods, somehow milder approximations have been introduced, allowing for the treatment of larger systems with a reasonable compromise between quality and computational effort. One of the most frequently used between these approximations is Resolution of the Identity [134] (RI). In this approach, the ERIs are approximated by:

$$\langle ij|kl\rangle = \sum_t \langle ij|t\rangle \langle t|kl\rangle \quad (2.81)$$

where we have use the second equality in 2.18. If a finite number of terms is taking into account, an error is introduced and should be minimized. As it can be noticed, the RI effectively reduces the complexity of the computation of the ERIs from $O(M^4)$ to $O(M^3)$. While RI has found applications in many methods, it is particularly useful when direct four-center integrals are involved in the calculations, like in the MP2 method [135].

As we mention at the beginning of this section, the previous lines do not attempt to be a comprehensive guide to electron correlation methods. Notable examples that were excluded from the discussion are the Variational Quantum Monte Carlo methods [93, 136, 137, and references therein], the Random Phase Approximation (RPA) [138], the

---

[22] Since the Hamiltonian contains at most two-body operators, the Baker-Campbell-Hausdorff expansion is finite and ends in the fourth term.

Green's Function methods (GW) [85], and the Algebraic–Diagrammatic Construction (ADC) [85]. The interested reader in invited to consult the suggested bibliography in order to have a more complete picture of the field.

2.2.6 Density Functional Theory

To finalize our discussion about Electronic Structure theory we will present one of the most successful approaches in the field: the Density Functional Theory (DFT) [139, 140, and references therein]. The theory is based on a reformulation of the Variational Principle (Equation 2.22) in terms of the one-electron density, which is defined as:

$$\rho(\mathbf{r}) = N \int d\sigma d\mathbf{x}_2 d\mathbf{x}_3 \cdots \mathbf{x}_N |\Psi(\mathbf{x}, \mathbf{x}_2, \cdots, \mathbf{x}_N)|^2 \tag{2.82}$$

where we have employed the grouped coordinates $\mathbf{x}$ introduced in Section 2.2.1. The mathematical advantages of ρ with respect to the wave function Ψ are enormous. Indeed, the number of arguments in the target function for the Variational Principle has been reduced from $3N$ to three, which implies that the complexity of the calculations will increase linearly with the system size [141]. DFT has its solid theoretical foundations in the Hohenberg-Kohn (HK) theorems [142], which were later transformed into a practical computational scheme by Kohn and Sham (KS) [143]. For the purposes of this work it is sufficient to introduce the HK theorems without demonstration[23], being the focus of the section on the KS equations. Throughout the discussion we will follow the arguments of Sun *et al.* in Reference [94].

In their general form, the HK theorems are:

Theorem 2.1 (Hohenberg-Kohn I). *Let $\rho_0(\mathbf{r})$ be the, possibly degenerate, ground-state density for an N-electron system. Then $\rho_0(\mathbf{r})$ determines not only the electron number,*

$$N = \int d\mathbf{r} \rho_0(\mathbf{r})$$

but also the external potential[24] *$v(\mathbf{r})$ and thus the Hamiltonian $\hat{H}$ and thereby everything about this system (e.g., the ground and excited-state wave functions).*

Theorem 2.2 (Hohenberg-Kohn II). *There exists a universal functional of the density, $F[\rho]$, such that for any N-representable density (i.e., any density that comes from some wave function for an N-electron system) $\rho(\mathbf{r})$, which yields a given number of electrons*

---

[23] See Section 3.2 of Reference [139] for a rigorous derivation.

[24] This is an arbitrary external local potential and not only the Coulomb potential generated by the nuclei treated in the Born-Oppenheimer approximation.

N, the energy functional is,

$$E_v[\rho] = F[\rho] + \int d\mathbf{r}\rho(\mathbf{r})v(\mathbf{r}) \geq E_0$$

The equality holds when the density $\rho(\mathbf{r})$ is the, possibly degenerate, ground-state density for the external potential $v(\mathbf{r})$.

As we can see in Equation 2.2, the universal functional is system independent (from there its name). The part of the energy functional that depends on the structure of the system is exclusively the integral $\int d\mathbf{r}\rho(\mathbf{r})v(\mathbf{r})$. The HK theorems offer (in principle) a way to obtain the ground state density. Unfortunately, in practice this reformulation of the Variational Principle is harder to manipulate than its wave function counterpart. The two main problems are: finding a systematic way to construct the universal functional $F[\rho]$, and given the latter, minimize Equation 2.2 to obtain the ground state energy. The second problem leads to the KS equations. The optimization proceeds in a similar way than in the Hartree-Fock case, using the Lagrange multipliers method:

$$\delta\left\{E_v[\rho] - \mu\int d\mathbf{r}\rho(\mathbf{r})\right\} = 0 \tag{2.83}$$

were we have introduced the multipliers μ. By using Equation 2.2 we get:

$$\int d\mathbf{r}\left\{\frac{\delta F[\rho]}{\delta\rho(\mathbf{r})} + v(\mathbf{r}) - \mu\right\}\delta\rho(\mathbf{r}) = 0 \tag{2.84}$$

and which gives the Euler-Lagrange equation [144]:

$$\frac{\delta F[\rho]}{\delta\rho(\mathbf{r})} + v(\mathbf{r}) = \mu \tag{2.85}$$

The previous expression is a corroboration of the HK Theorem I, i.e. the external potential $v(\mathbf{r})$ is uniquely determined by the ground-state density. Lets now introduce the so called Kohn-Sham noninteracting system [94]. In order to do this, we first set the general expression for $F[\rho]$ using Levy's constrained-search formulation [145]:

$$F[\rho] = \langle\Psi_\rho|\hat{T}_e + \hat{V}_{ee}|\Psi_\rho\rangle \tag{2.86}$$

where we have used the operators defined in Section 2.1, and have indicated the implicit association between wave function and density. In a noninteracting system the second operator in the previous equation vanishes, so we are left with:

$$T_s[\rho] = \min_{\Psi_\rho\to\rho_0}\langle\Psi_\rho|\hat{T}_e|\Psi_\rho\rangle = \langle\Psi_{SD}|\hat{T}_e|\Psi_{SD}\rangle \tag{2.87}$$

where Ψ_{SD} is the Slater Determinant that minimizes the expression for a given ρ. If we plug this result in Equations 2.85 we get:

$$\frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} + v_s(\mathbf{r}) = \mu_s \tag{2.88}$$

where we have introduced the noninteracting external $v_s(\mathbf{r})$ and chemical μ_s potentials. We can now define the Exchange-Correlation energy as:

$$E_{xc}[\rho] = F[\rho] - T_s[\rho] - U[\rho] \tag{2.89}$$

where $U[\rho]$ is the Hartree Coulomb interaction:

$$U[\rho] = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{2.90}$$

If we take the variation of Equation 2.89, use the results from Equations 2.85 and 2.88, and then rearrange the terms we get:

$$v_s(\mathbf{r}) = v(\mathbf{r}) + u[\rho] + v_{xc}[\rho] \tag{2.91}$$

In the previous equation, the difference between the chemical potentials of the interaction and noninteracting systems get adsorbed by the $v_s(\mathbf{r})$. We have also defined the Hartree and exchange-correlation potentials as:

$$u[\rho] = \frac{\delta U[\rho]}{\delta\rho(\mathbf{r})}; \qquad v_{xc}[\rho] = \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} \tag{2.92}$$

The Hamiltonian for the noninteracting system is given by:

$$\hat{H}_s = \sum_i \left[-\frac{1}{2}\boldsymbol{\nabla}_i^2 + v_s(\mathbf{r}_i)\right] \tag{2.93}$$

The ground state wave function of this system Ψ_s must satisfy the auxiliary Schrödinger Equation $\hat{H}_s\Psi_s = E_s\Psi_s$. Since $\hat{H}_s$ is just a sum of one-electron Hamiltonians (see Section 2.2.1) and Ψ_s can be considered as a Slater Determinant, we are left with the set of equations:

$$\hat{h}_s\chi_i = \left[-\frac{1}{2}\boldsymbol{\nabla}_i^2 + v_s(\mathbf{r}_i)\right]\chi_i = \epsilon_i\chi_i \tag{2.94}$$

correspondingly, the electron density can be computed as:

$$\rho(\mathbf{r}) = \sum_i |\chi_i|^2 \tag{2.95}$$

The total energy of the system of interacting electrons can be computed as:

$$E = T_s[\rho] + U[\rho] + E_{xc} + \int d\mathbf{r}\rho(\mathbf{r})v(\mathbf{r}) \tag{2.96}$$

This concludes the derivation of the KS formalism.

The Hartree-Fock and KS equations have the same structure, but the latter is a more general formalism due to the flexibility of incorporate exchange-correlation effects [139]. In fact, KS would be exact if $v_{xc}[\rho]$ was known precisely. There are two ways in which the above can be achieved: an empirical approach in which experimental data is used to generate the functional, and a nonempirical one in which one relies on the mathematical properties of the exact functional. The previous two approaches have generated three main families of approximations:

- Local Density Approximation (LDA), in which the exchange-correlation potential is approximated by the exchange-correlation energy of an electron in an homogeneous electron gas of the same density $\rho(\mathbf{r})$:

  $$E_{xc}^{LDA} = \int d\mathbf{r}\rho(\mathbf{r})\epsilon_{xc}(\rho) \tag{2.97}$$

  where $\epsilon_{xc}(\rho)$ is the exchange-correlation energy per particle.

- Generalized Gradient Approximation (GGA), which is an improvement with respect to LDA in the sense that they explicitly include the gradient of the electronic density in the expression of the functional:

  $$E_{xc}^{GGA} = \int d\mathbf{r} f(\rho, \nabla\rho) \tag{2.98}$$

- Hybrid methods, in which part of the exchange energy is taken from a Hartree-Fock calculation.

Generally, DFT methods provide very precise geometries and frequencies at very low computational cost. They also allow the study of systems of medium-large size with reasonably good accuracy [140]. Two of the main drawbacks from the theory are the impossibility of systematic improvement of the results and the incorrect description of dissociative processes [146]. Finally, it is worth mentioning that an DFT has been extended to treat time-dependent electronic densities, in what is known as time-dependent density-functional theory (TDDFT) [147]. Despite its widespread success, the validity of TDDFT has been questioned by some authors [148, 149], being one of the most controversial points the non-stationarity of the action functional integral [150].

2.3 Quantum Dynamics

We have seen so far that in Electronic Structure theory the electrons are treated in a full quantum mechanical manner. If we now attempt to include the nuclei in our model we have two possible alternatives. In the first one of them, which is called *Molecular Dynamics*, the movement of nuclei is described by solving the classical Newton's Equations of Motion (EOM). PESs of different qualities can be used in this case, ranging from the computationally less involved Force Fields[25] up to more elaborated *ab initio* approaches. In the latter case, the PES can be computed in a previous stage (fit) or can be generated in each time step of the calculations (Born-Oppenheimer or *on-the-fly* Molecular Dynamics). In between the two previous approaches lies the so called Car-Parrinello molecular dynamics (CPMD) [151]. In the second scenario, both electrons and nuclei are described using quantum mechanical formalisms, and it is called *Nuclear Quantum Dynamics* [152]. In the present work we have employed both methodologies, the first one in the vdW-TSSCDS method (see Section 2.5) and the second one in the MCTDH method (see Section 2.3.2.1). In the following pages we will go through the most important aspects of both cases.

2.3.1 Classical Molecular Dynamics

The classical description of the movement of nuclei in molecular systems is often carried out by finding the numerical (step-by-step) solutions of the Newton EOM [153]:

$$M_\alpha \ddot{\mathbf{r}}_\alpha = \mathbf{F}_\alpha \tag{2.99}$$

were the subscript α runs over the nuclei of mass M_α, and $\mathbf{F}_\alpha$ is the force acting on it. The $\mathbf{r}_\alpha$ represent a complete set of N generalized coordinates, that for simplicity in the discussion will be considered Cartesian coordinates.

The Newtonian mechanics can be generalized using the Lagrangian or the Hamiltonian formulations. The latter offer a lot of mathematical advantages (for example, the ability to deal with Cartesian, cylindrical, spherical, and any other coordinate systems with ease) with respect to the former[26]. The Lagrangian function of the system can be defined as:

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t) = \mathcal{T}(\dot{\mathbf{q}}) - \mathcal{U}(\mathbf{q}) \tag{2.100}$$

---

[25] Representations of the electronic energy as a parametric function of the nuclear coordinates. Molecules are described by a "ball and springs" model that does not consider electrons as individual particles, and the solution of the TISE is omitted [69].

[26] For a complete description of the different formalism in Classical Mechanics please consult the standard textbook by Goldstein *et al.* [154].

where we have introduced $3N$ generalized coordinates $\mathbf{q}$ and the corresponding generalized velocities $\dot{\mathbf{q}}$. The kinetic energy $\mathcal{K}$ is defined as:

$$\mathcal{K}(\dot{\mathbf{q}}) = \sum_{\alpha} \frac{1}{2} M_{\alpha} \dot{q}_{\alpha}^{2} \tag{2.101}$$

The potential energy function $\mathcal{U}$ depends on the $3N$ generalized coordinates. Note that even in the case where the Lagrangian does not have time dependent constraints or influence of external fields, there would be an implicit dependence on the time t due to the use of generalized coordinates. If we now introduce the generalized momenta as:

$$p_{\alpha} = \frac{\partial \mathcal{L}}{\partial \dot{q}_{\alpha}} \tag{2.102}$$

we can define the Hamiltonian function as the Legendre transformation of the Lagrangian:

$$\mathcal{H}(\mathbf{q}, \mathbf{p}, t) = \sum_{\alpha} p_{\alpha} \dot{q}_{\alpha} - \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{t}) \tag{2.103}$$

The Hamiltonian obeys the Hamilton equations:

$$\frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t}; \quad \dot{q}_{\alpha} = \frac{\partial \mathcal{H}}{\partial p_{\alpha}}; \quad -\dot{p}_{\alpha} = \frac{\partial \mathcal{H}}{\partial q_{\alpha}} \tag{2.104}$$

Provided the Newtonian and Hamiltonian formulations are equivalent [154], the expression of the force in Equation 2.99 can be readily found by [155]:

$$\mathbf{F}_{\alpha} = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}_{\alpha}} = -\frac{\partial \mathcal{U}}{\partial \mathbf{r}_{\alpha}} \tag{2.105}$$

where we have set our generalized coordinates to be equivalent to the Cartesian coordinates $\mathbf{r}_{\alpha}$.

The time evolution of the coordinates and momenta of a system is called *trajectory*. The study of a time-dependent classical observable O is usually carried out by averaging it over time along a trajectory [156]:

$$\langle O \rangle = \lim_{t \to \infty} \frac{1}{t} \int_{0}^{t} d\tau O[\mathbf{r}(\tau), \dot{\mathbf{r}}(\tau)] \tag{2.106}$$

The above assumption is usually complemented with the Ergodic Hypothesis, which states that the average over long periods of time along a given trajectory is identical to the ensemble average[27] of the observable. Obviously, to obtain the trajectories associated to a given molecular system, one needs to solve the Newton or Hamilton EOM. Both

---

[27] The average of a quantity that is function of the individual states (microstate) of the system [157].

systems of equations are time-reversible and the Liouville Theorem[28] holds for them [156]:

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \sum_\alpha \left(\frac{\partial\rho}{\partial q_\alpha}\dot{q}_\alpha + \frac{\partial\rho}{\partial p_\alpha}\dot{p}_\alpha \right) = 0 \tag{2.107}$$

Numerical algorithms must ensure these two conditions in order to obtain physically meaningful solutions. In order to design a suitable finite differences algorithm we will transform Hamilton equations by introducing the Liouville operator, which defines a propagator for the generalized coordinates:

$$iL = \sum_\alpha \left[\frac{\partial\mathcal{H}}{\partial p_\alpha}\frac{\partial}{\partial q_\alpha} - \frac{\partial\mathcal{H}}{\partial q_\alpha}\frac{\partial}{\partial p_\alpha} \right] \tag{2.108}$$

$$\mathbf{q}(t) = \mathbf{q}(0)\cdot e^{iLt} \tag{2.109}$$

In Cartesian coordinates the Liouville operator can be written as:

$$iL = iL_1 + iL_2; \tag{2.110}$$

$$iL_1 = \sum_i \frac{\mathbf{p}_i}{M_i}\cdot\frac{\partial}{\partial\mathbf{r}_i} \qquad iL_2 = \sum_i \left(-\frac{\partial\mathcal{U}}{\partial\mathbf{r}_i}\right)\cdot\frac{\partial}{\partial\mathbf{p}_i} \tag{2.111}$$

were the index i runs over the number N of Cartesian coordinates. We could now study the time evolution of the system by propagating Equation 2.109. Since the individual components of the Liouville operator do not commute, we use the Suzuki-Trotter decomposition [158] to approximate the exponential:

$$e^{iL\Delta t} = e^{i(L_1+L_2)\Delta t} \approx e^{iL_2\frac{\Delta t}{2}} e^{iL_1\Delta t} e^{iL_2\frac{\Delta t}{2}} + o(\Delta t^3) \tag{2.112}$$

We can further approximate each exponential by a Taylor series, which transforms Equation 2.109 into:

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) - \frac{1}{M_i}\frac{\partial\mathcal{U}}{\partial\mathbf{r}_i}\Delta t^2 + o(\Delta t^4)\bigg|_{i=1,\ldots,N} \tag{2.113}$$

The previous expression is the well known Verlet algorithm, and the reasoning presented so far can be use to derive many other integration algorithms commonly employed in Molecular Dynamics [156].

Despite the logical arguments that motivate the use of Molecular Dynamics, and the enormous success of the methodology [159, and references therein], its applicability is limited. Indeed, there are systems for which the nuclear quantum effects cannot be neglected, to mention some of them [152]: molecular spectroscopy (particularly infrared

---

[28] The distribution function $\rho(p,q,t)$ is constant along any trajectory in *phase space* $\Gamma = \prod_i \{q_i, p_i\}$. See Reference [157] for a demonstration.

spectroscopy), in which transition between the vibrational states is quantized; the quantum tunneling, in which particles can traverse barriers impossible to overcome in classical terms, hence having a direct influence in the accurate determination of rate constants in chemical reactions; and the strong vibronic coupling between electronic states due to the nuclear motion, particularly conical intersections.

Of particular relevance to the present work is the accurate estimation of the adsorption energy of a small molecule placed over a large molecular surface (computed as the opposite of the difference between the Zero-Point Energy (ZPE) and the energy of the isolated fragments). In (quasi-)classical terms, the theoretical computation of this energy for a system with a highly corrugated PES (many stationary points), would likely involve averaging the ZPEs of the independent wells, weighted by their corresponding residence times. The sagacious reader could readily infer the meaning of the *quasi* prefix in the previous sentence: the ZPE is an exclusively quantum effect. Moreover, the computation of absorption energies can be greatly improved using quantum dynamical reasonings, as with a wave packet relaxation one can directly obtain the Ground State (GS) distribution of the system, and the *exact* ZPE. Due to its relevance, the biggest part of the present section will be dedicated to the description of Nuclear Quantum Dynamics, with particular emphasis on one of the most powerful and successful methods in the field: Multi-Configuration Time Dependent Hartree (MCTDH).

2.3.2 Nuclear Quantum Dynamics

Nuclear Quantum Dynamics methods provide a numerical solution of the Time Dependent Schrödinger Equation (see 2.1). Although exact solutions to the above equation can only be obtained in an infinite dimensional Hilbert space, in practice one needs to truncate the basis sets to a finite dimension (Galerkin's Method). Given a parametric representation of the total wave function of the system Ψ, one can find the optimal solution using the Dirac-Frenkel Variational Principle [160]:

$$\langle \delta\Psi | \hat{H} - i\frac{\partial}{\partial t} | \Psi \rangle = 0 \tag{2.114}$$

The simplest way of representing the wave function is given by a product of time independent basis sets, with time dependent coefficients, in what is usually called the Standard Method:

$$\Psi(q_i, \ldots, q_f, t) = \sum_{j_1=1}^{N_1} \cdots \sum_{j_f=1}^{N_f} C_{j_1 \ldots j_f}(t) \prod_{\kappa=1}^{f} \varphi_{j_\kappa}^{(\kappa)}(q_\kappa) \tag{2.115}$$

where f represents the number of degrees of freedom (DOFs) of the system, and the $\varphi_{j_\kappa}^{(\kappa)}$ form an orthonormal set of N_κ functions. Examples of the latter are the particle-in-a-box eigenfunctions (used to describe wave functions that vanish at finite boundaries), the Hermite functions solution of the one-dimensional harmonic oscillator (used for vibrational modes), the normalized Legendre polynomials (used for planar angles), and the two dimension Spherical Harmonics (used when the Kinetic Energy Operator have singularities) [70]. These basis sets are called Finite Basis Representation (FBR). The solution of the TDSE in this approach is fairly simple from the mathematical point of view, and shares the spirit of the Configuration Interaction method introduced in the Electronic Structure section[29]:

$$\begin{aligned} i\dot{C}_L &= \sum_J \langle \varphi_L | \hat{H} | \varphi_J \rangle \, C_J \\ \mathbf{C}(t) &= e^{-i\mathbf{H}t}\mathbf{C}(0) \end{aligned} \tag{2.116}$$

In the above equations we have introduced the composite indexes $J = (j_1, \ldots, j_f)$ (the same for L) and the Hartree products $\varphi_J = \prod_\kappa \chi_{j_\kappa}$. In addition, the bold fonts indicate the vector and matrix form of the coefficients tensor and the Hamiltonian respectively. Owing to its conceptual simplicity, the standard method is greatly affected by the Curse of Dimensionality [161], being its application very unpractical for systems with more than 4 atoms. One the next sections we will introduce some extra approximations that allow for greatly climbing on this "dimensionality hill".

One aspect of great relevance to forthcoming discussions is the concept of Discrete Variable Representation (DVR) [162]. A DVR basis consists on a set of N functions, localized on space (grid points), which can be obtained by unitary transformation of the FBR, making both representations mathematically equivalent. To obtain a DVR one needs to diagonalize the matrix representation of the coordinate operator in the FBR basis φ:

$$\begin{aligned} Q_{jk} &= \langle \varphi_j | x | \varphi_k \rangle \\ \mathbf{Q} &= \mathbf{U}\mathbf{X}\mathbf{U}^{+} \end{aligned} \tag{2.117}$$

and subsequently:

$$\begin{aligned} \chi_\alpha^{DVR}(x) &= \sum_j \varphi_j(x) U_{j\alpha} \\ \langle \chi_\alpha^{DVR} | \chi_\beta^{DVR} \rangle &= \delta_{\alpha\beta} \end{aligned} \tag{2.118}$$

DVR functions act as Dirac's deltas on the grid points, and have the important property that in this representation any multiplicative (local) operator is diagonal. The latter

---

[29] The algebraic details in the forthcoming sections might be occasionally abbreviated. For a complete analysis consult the book of Gatti *et al.* [70]

allows for great simplification in the computation of the integrals arising from the potential energy operator, at the expense of computing them in quadratures. One might want to switch back to FBR to compute the action of differential (non-local) operators like the Kinetic Energy Operator (KEO), because the latter is analytical in the FBR.

Before presenting the MCTDH method, we will briefly discuss some of the algorithms that are commonly used to integrate the TDSE. The formal solution of the latter (Equation 2.1) is given by:

$$\begin{aligned} |\Psi(t)\rangle &= e^{-i\hat{H}t}\,|\Psi(0)\rangle \\ |\Psi(t+\tau)\rangle &= e^{-i\hat{H}\tau}\,|\Psi(t)\rangle \end{aligned} \tag{2.119}$$

The temporal evolution of the wave function is then studied by propagating it with the operator e^{-iHt}. Here we mention three algorithms that can be used to carry out this process: the Split Operator [163], which makes use of the the Suzuki-Trotter decomposition (see Section 2.3.1); the Second Order Difference [70]; and the Short Iterative Lanczos (SIL) [164], which belongs to the family of Krylov subspace methods. As an example, we will describe the third one, particularly in the Arnoldi-Lanczos variant (which is used in MCTDH) [165].

We will first introduce the truncated Taylor expansion of the second equality in 2.119, and the corresponding Krylov subspace spanned by the Hamiltonian matrix:

$$|\Psi(t+\tau)\rangle \approx \sum_{n=0}^{L} \frac{1}{n!}\left[-i\hat{H}\tau\right]^{n} |\Psi(t)\rangle \tag{2.120}$$

and accordingly:

$$\mathcal{K}_{L+1}(\mathbf{H},\Psi) = \mathrm{span}\{\Psi, \mathbf{H}\Psi, \ldots, \mathbf{H}^{L}\Psi\} \tag{2.121}$$

We can then find an orthonormal set of L vectors q_k in $\mathcal{K}_{L-1}$ by using a modified Gram-Schmidt algorithm, whose starting vector is $q_0 = |\Psi(t)\rangle / \| |\Psi(t)\rangle \|$. The wave function and Hamiltonian can now be approximated as:

$$\begin{aligned} \Psi(t+\tau) &\approx \sum_{j=0}^{L} |q_j\rangle \langle q_j|\Psi(t+\tau)\rangle \\ \hat{H} &\approx \sum_{j=0}^{L}\sum_{k=0}^{L} |q_j\rangle \langle q_j|\hat{H}|q_k\rangle \langle q_k| \end{aligned} \tag{2.122}$$

In the above expression, the matrix element $\langle q_j|\hat{H}|q_k\rangle$ corresponds to the reduced Hamiltonian $\mathbf{H_L}$. The latter can be diagonalized with a conjugation of the form $\mathbf{P^{-1}H_LP} = \mathbf{\Lambda}$.

By combining 2.122 and 2.119 we get the final expression of the propagated wave function:

$$\begin{aligned}|\Psi(t+\tau)\rangle &\approx \sum_{j=0}^{L} |q_j\rangle \langle q_j| e^{-i\hat{H}\tau} |\Psi(t)\rangle \\ &\approx \sum_{j=0}^{L} |q_j\rangle \langle q_j| \sum_{k=0}^{L}\sum_{l=0}^{L} |q_k\rangle \left[e^{-i\mathbf{H_L}\tau}\right]_{kl} \langle q_l|q_0\rangle \, \||\Psi(t)\rangle\| \\ &= \sum_{j=0}^{L} |q_j\rangle \sum_{k=0}^{L} P_{jk} e^{-i\lambda_k \tau} P^{-1}_{k0} \||\Psi(t)\rangle\|\end{aligned} \tag{2.123}$$

where we have made use of the orthonormality of the q_k and introduced the matrix elements λ_k of $\mathbf{\Lambda}$. The pseudocode for the algorithm can be summarized as:

```
Algorithm 1: Lanczos-Arnoldi
Result: ${q_k}$ and $\langle q_j|\hat{H}|q_k\rangle$
$q_0 = \Psi(t)/\|\Psi(t)\|$;
for $k \leftarrow 0$ to $L$ do
    $\chi^{(0)}_{k+1} = \hat{H} q_k$;
    for $j \leftarrow 0$ to $k$ do
        $\beta_{jk} = \langle q_j|\chi^{(j)}_{k+1}\rangle$;
        $\chi^{(j+1)}_{k+1} = \chi^{(j)}_{k+1} - q_j\beta_{jk}$;
    end
    $\beta_{k+1,k} = \|\chi^{(k+1)}_{k+1}\|$;
    $q_{k+1} = \chi^{(k+1)}_{k+1}/\beta_{k+1,k}$;
end
```

The reduced Hamiltonian elements form a a complex upper Hessenberg matrix:

$$\langle q_j|\hat{H}|q_k\rangle = \begin{cases} \beta_{jk} & \text{if} \quad j \leq k+1 \\ 0 & \text{else} \end{cases} \tag{2.124}$$

Once constructed and diagonalized the Hessenberg matrix (which is considerably smaller than the original Hamiltonian matrix), one can proceed with the propagation as dictated by 2.123.

2.3.2.1 The Multiconfiguration Time-Dependent Hartree method (MCTDH)

Once described the generalities of the numerical solution of the TDSE, we can present two further approximations that can be introduced to improve the Standard Method, namely the Time-Dependent Hartree [166] (TDH) and the Multiconfiguration Time-Dependent Hartree [9, 167] (MCTDH) methods. The focus of the section will be on the latter.

The TDH *ansatz* is:

$$\Psi(q_1, \ldots, q_f, t) = a(t) \prod_{\kappa=1}^{f} \varphi_\kappa(q_\kappa, t) \tag{2.125}$$

The functions $\varphi_\kappa(q_\kappa, t)$ are called single-particle functions (SPFs) and are expressed in a time-independent basis set. The derivation of the TDH Equations of Motion (EOM) constitutes an excellent pedagogical example of the application of the Dirac-Frenkel variational principle, and it can be found in Section 8.2.1 of Reference [70]. Here we will limit our discussion to mention that TDH represents a computational improvement with respect to the Standard Method, but its application its limited because for many problems the wave function has a pronounced multiconfigurational character[30] in order to describe (even qualitatively) some problems [168].

A more sophisticated *ansatz* is needed in order to increase our reach in the "dimensionality hill". Sharing the spirit of the Multiconfigurational Electronic Structure methods (see Section 2.2.5), in 1990 Meyer *et al.* proposed to use a wave function of the form[31]:

$$\Psi(q_1, \ldots, q_f, t) = \sum_{j_1=1}^{n_1} \cdots \sum_{j_f=1}^{n_f} A_{j_1,\ldots,j_f}(t) \prod_{\kappa=1}^{f} \varphi_{j_\kappa}^{(\kappa)}(q_\kappa, t) \tag{2.126}$$

where the SPFs are given by:

$$\varphi_{j_\kappa}^{(\kappa)}(q_\kappa, t) = \sum_{\mu=1}^{N_\kappa} c_\mu^{(\kappa,j_\kappa)}(t) \cdot \chi_\mu^{(\kappa)}(q_\kappa) \tag{2.127}$$

with $\chi_\mu^{(\kappa)}(q_\kappa)$ being typically a DVR function. MCTDH is a general approach the has as limit cases both the TDH and the Standard Method, and is certainly an improvement of the former. Indeed, if in Equation 2.126 we set $n_1 = n_2 = \ldots = n_f = 1$ then we are left with Equation 2.125. If, on the other hand, we set $n_1 = N_1, n_2 = N_2, \ldots, n_f = N_f$ then MCTDH is equivalent to the Standard Method. Note however that the number of configurations in MCTDH is smaller, which brings obvious computational advantages. It must be pointed out that in the MCTDH formalism the wave function is not (anti)-symmetrized by default, so the treatment of fermions and bosons should take this into account and hence the MCTDHF [169] and MCTDHB [170] methods have been developed.

The MCTDH *ansatz* is determined up to a linear transformation of the SPFs, so in order to derive the corresponding equations of motion a constraint must be included. The

---

[30] Retrieving nuclear correlation is a harder task than its electronic counterpart, due to the higher nuclear masses.

[31] An interesting analogy to Electronic Structure methods cen be found in Reference [70]. We can consider the *ansatz* of the Standard Method to be similar to the one of a FCI, the TDH method to a Hartree-Fock, and the MCTDH to a MCSCF.

above can be done by setting:

$$i\,\langle\varphi_l^{(\kappa)}|\dot{\varphi}_j^{(\kappa)}\rangle = \langle\varphi_l^{(\kappa)}|\hat{g}^{(\kappa)}|\varphi_l^{(\kappa)}\rangle \tag{2.128}$$

where $g^{(\kappa)}$ is an arbitrary constraint operator. With the previous condition in mind, and after an involved algebraic procedure [70], we arrive to the MCTDH-EOM:

$$\begin{aligned} i\dot{A}_J &= \sum_L \langle\Phi_J|\hat{H}|\Phi_L\rangle\, A_L \\ i\dot{\varphi}_j^{(\kappa)} &= (1-\hat{P}^{(\kappa)})\sum_{k,l=1}^{n_\kappa}(\boldsymbol{\rho}^{(\kappa)^{-1}})_{jk}\,\langle\hat{H}\rangle_{kl}^{(\kappa)}\,\varphi_l^{(\kappa)} \end{aligned} \tag{2.129}$$

Several remarks regarding the previous set of coupled integro-differential equations need to be made. First, they have been obtained for the particular case in which $\hat{g}^{(\kappa)}$ is zero for all κ. As in previous discussions, the letters J, L represent composite indexes. The quantities Φ_J are called configurations (or Hartree products) and are computed as:

$$\Phi_J = \prod_{\kappa=1}^{f}\varphi_{j_\kappa}^{(\kappa)} \tag{2.130}$$

In addition, we have introduced the densities and mean fields defined as:

$$\begin{aligned} \rho_{kl}^{(\kappa)} &= \langle\Psi_k^{(\kappa)}|\Psi_l^{(\kappa)}\rangle = \sum_{J^\kappa} A^*_{J_k^\kappa} A_{J_l^\kappa} \\ \langle\hat{H}\rangle_{kl}^{(\kappa)} &= \langle\Psi_k^{(\kappa)}|\hat{H}|\Psi_l^{(\kappa)}\rangle \end{aligned} \tag{2.131}$$

where the $\Psi_k^{(\kappa)}$ are called single-hole functions, and are given by:

$$\Psi_l^{(\kappa)} = \sum_{J^\kappa} A_{J_l^\kappa}\Phi_{J^\kappa} \tag{2.132}$$

In the above equation we extended the nomenclature with:

$$\begin{aligned} J^\kappa &= (j_1, \ldots, j_{\kappa-1}, j_{\kappa+1}, \ldots, j_f) \\ J_l^\kappa &= (j_1, \ldots, j_{\kappa-1}, l, j_{\kappa+1}, \ldots, j_f) \\ \Phi_{J^\kappa} &= \prod_{\nu\neq\kappa}^{f}\varphi_{j_\nu}^{(\nu)} \end{aligned} \tag{2.133}$$

Finally, the quantity $\hat{P}^{(\kappa)}$ is called the MCTDH projector, and is computed as:

$$P^{(\kappa)} = \sum_{j=1}^{n_\kappa}|\varphi_j^{(\kappa)}\rangle\,\langle\varphi_j^{(\kappa)}| \tag{2.134}$$

If we now consider the Hamiltonian operator to be:

$$\hat{H} = \sum_{\kappa} \hat{h}^{(\kappa)} + \hat{H}_R \tag{2.135}$$

where $\hat{h}^{(\kappa)}$ represents an operator acting on only on the κ-th DOF, and $\hat{H}_R$ the non-separable part of the operator, then the MCTDH-EOM can be written:

$$\begin{aligned} i\dot{A}_J &= \sum_{L} \mathcal{K}_{JL} A_L \\ i\dot{\varphi}_j^{(\kappa)} &= (1-\hat{P}^{(\kappa)})\{\hat{h}^{(\kappa)}\varphi_j^{(\kappa)} + \sum_{k,l=1}^{n_\kappa} (\boldsymbol{\rho}^{(\kappa)^{-1}})_{jk} \langle \hat{H}_R \rangle_{kl}^{(\kappa)} \varphi_l^{(\kappa)}\} \end{aligned} \tag{2.136}$$

where we have introduced the quantity $\mathcal{K}_{JL} = \langle \Phi_J | \hat{H} | \Phi_L \rangle$.

The MCTDH-EOM 2.129 that we have presented have been obtained with the simplest choice of constraints, but in some situations a more sophisticated choice might be needed [167]. To find accurate solutions of this system of coupled[32] equations is a challenging task. On the other hand, the number of equations is greatly reduced with respect to the Standard Method, which is due to the inclusion of less configurations in the expansion of the wave function as $n_\kappa < N_\kappa$ (compare Equations 2.115 and 2.126). Even when MCTDH needs (at least formally) to solve many multi-dimensional integrals for every time-step (see the MCTDH-EOM 2.136), the usage of Sum of Products (SOP) form simplifies this process considerably. To illustrate this, let's suppose we want to determine the matrix elements of an arbitrary Hermitian operator $\hat{O}$ that can be represented in Tucker form, over some configuration functions Φ_J, then we have:

$$\begin{aligned} \hat{O} &= \sum_{\alpha=1}^{S} c_\alpha \prod_{\kappa=1}^{f} \hat{o}_\alpha^{(\kappa)} \\ \langle \Phi_J | \hat{O} | \Phi_L \rangle &= \sum_{\alpha=1}^{S} c_\alpha \prod_{\kappa=1}^{f} \langle \varphi_{j_\kappa}^{(\kappa)} | \hat{o}_\alpha^{(\kappa)} | \varphi_{l_\kappa}^{(\kappa)} \rangle \end{aligned} \tag{2.137}$$

As we can see, the multi-dimensional integrals have been reduced to sums of products of one-dimensional integrals, easily solvable by quadratures. The Kinetic Energy Operator (KEO) is most of the time representable in SOP form [171, 172]. In the case of the potential operator, special techniques (such as tensor-decomposition methods) need to be applied. A large part of the present Doctoral work is focused on the development of efficient methodologies that allow for obtaining PES in the SOP form. For more details on the subject consult the introduction to this problem in Section 2.4 and the developed methods in Chapter 3.

---

[32] Indeed, $\dot{A}$ depends on φ and at the same time $\dot{\varphi}$ depends on A.

The system in 2.129 could be solved with a general purpose integrator like Runge-Kutta or or Adams-Bashforth-Moulton. In this scheme, the matrix elements $\mathcal{K}_{LJ}$ and the mean fields $\langle \hat{H}_R \rangle_{kl}^{(\kappa)}$ would need to be recalculated at every time step, which in turn needs to be small enough to ensure an accurate description of the oscillatory behavior of the wave function. Since the mean fields are not strongly oscillating, a good approximation to solve the Equations 2.136 would be to take them as constant over a large time-step interval τ, and integrate the A vectors and the SPPs in shorter time intervals. The previous integration scheme is know as constant mean-field (CMF). It effectively decouples the differential equations in 2.136 in a given time interval τ (barred quantities below):

$$\begin{aligned}
i\dot{A}_J &= \sum_L \bar{\mathcal{K}}_{JL} A_L \\
i\dot{\varphi}_j^{(1)} &= (1-\hat{P}^{(1)})\{\hat{h}^{(1)}\varphi_j^{(1)} + \sum_{k,l=1}^{n_1} (\boldsymbol{\rho}^{(1)^{-1}})_{jk} \langle \bar{H}_R \rangle_{kl}^{(1)} \varphi_l^{(1)}\} \\
&\vdots \\
i\dot{\varphi}_j^{(f)} &= (1-\hat{P}^{(f)})\{\hat{h}^{(f)}\varphi_j^{(f)} + \sum_{k,l=1}^{n_f} (\boldsymbol{\rho}^{(f)^{-1}})_{jk} \langle \bar{H}_R \rangle_{kl}^{(f)} \varphi_l^{(f)}\}
\end{aligned} \tag{2.138}$$

The above equations can be solved independently using the most convenient integrator in each case. For example, the first of the equations in 2.138 is linear and time-independent, so it can be solved using the SIL algorithm introduced in the previous section.

2.3.2.2 Mode combination and Multilayer MCTDH

In the previous section it was shown that the MCTDH-EOM can be solved efficiently using the CMF approximation. In order to increase the reach of the algorithm (which still suffers from exponential scaling), several strategies have been followed. The first one of them is based on the use of combined modes. Due to the flexibility of the SPFs, they can be taken to depend on more than one variable. In this way, several physical coordinates can be regrouped into one logical particle [70]:

$$\begin{aligned}
Q_\kappa &\equiv (q_{\kappa,1}, q_{\kappa,1}, \ldots, q_{\kappa,d}) \\
\varphi_j^{(\kappa)}(Q_\kappa, t) &= \varphi_j^{(\kappa)}(q_{\kappa,1}, q_{\kappa,1}, \ldots, q_{\kappa,d}, t)
\end{aligned} \tag{2.139}$$

Under these conditions, the MCTDH *ansatz* will take the form:

$$\Psi(q_1, \ldots, q_f, t) \equiv \Psi(Q_1, \ldots, Q_p, t) = \sum_{j_1=1}^{n_1} \cdots \sum_{j_p=1}^{n_p} A_{j_1,\ldots,j_p}(t) \prod_{\kappa=1}^{p} \varphi_{j_\kappa}^{(\kappa)}(Q_\kappa, t) \tag{2.140}$$

On the other hand, the SPPs can be expanded as:

$$\varphi_j^{(\kappa)}(Q_\kappa, t) = \sum_{i_1 \dots i_d} C_{i_1 \dots i_d}^{(\kappa, j)}(t) \prod_{\nu=1}^{d} \chi^{(\kappa,\nu)}(q_{\kappa,\nu}) \tag{2.141}$$

For the above scheme to have actual impact on the calculations, one needs to ensure that the total number of SPFs would be smaller than in the non-combined case. The previous conditions ensures that the length of the A-vector would be reduced, greatly diminishing the computational effort. As can be readily inferred, all correlations between DOFs in the same particle need to be addressed at the SPF level. Mode combination effectively transfers the complexity of the A-vector towards the SPF (which are now multidimensional, hence more complicated to propagate). The above constitutes the principal motivation for the Multilayer (ML) approach, i.e. to propagate the multidimensional SPFs with MCTDH [173–175].

To better grasp the idea behind ML-MCTDH, the reader is invited to recall the evolution on the form of the wave function in the Standard Method, and in MCTDH (see previous section). With the same spirit of the approximations that led to the MCTDH *ansatz*, we could propose the following generalization with *three layers*:

$$\Psi(q_1, q_2, q_3, t) = \sum_{j_{12}=1}^{n_{12}} \sum_{j_3=1}^{n_3} A_{j_{12} j_3}(t) \varphi_{j_{12}}^{(12)}(q_1, q_2, t) \varphi_{j_3}^{(3)}(q_3, t) \tag{2.142}$$

where we have set the number of DOFs to three for simplicity, and we have introduced:

$$\varphi_{j_{12}}^{(12)}(q_1, q_2, t) = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} B_{k_1,k_2}^{(12,j_{12})}(t) \prod_{\mu=1}^{2} \xi_{k_\mu}^{(\mu)}(q_\mu, t) \tag{2.143}$$

and:

$$\xi_{k_\mu}^{(\mu)}(q_\mu, t) = \sum_{i_\mu=1}^{N_\mu} c_{i_\mu}^{(\mu,k_\mu)}(t) \chi_{i_\mu}^{(\mu)}(q_\mu) \tag{2.144}$$

As we can see, for each combined SPFs one performs an additional MCTDH expansion instead of the primitive basis. The approach is extensible to as many layers as needed, always ensuring that all time-dependent functions are optimized at the same time. At this point it becomes obvious that MCTDH is a *two layer* scheme. Our analysis of the ML-MCTDH method is rather qualitative here, for a detailed derivation of the corresponding EOMs and subsequent discussion of the effectiveness of the approach consult Reference [174–176]. As a final remark, we should mention that the usage of mode combination becomes advantageous when there are more than five DOFs, whereas the ML scheme becomes more efficient than MCTDH when there are more than ten DOFs approximately.

2.3.3 Relaxation and (Block) Improved Relaxation

It is also possible to use the MCTDH algorithm to solve the TISE. This process (which is analogous to vibrational MCSCF) converges to a set of *vibrational orbitals*. The ground state (GS) wave function of a system can be obtained in a time-dependent fashion by performing relaxation, i.e. propagation in negative imaginary time $\tau = -it$. Under this conditions, the TISE becomes:

$$\dot{\Psi} = -\hat{H}\Psi \tag{2.145}$$

The wave function can be expanded in terms of its eigenstates as:

$$\Psi(t) = \sum_n a_n e^{-E_n t}\Psi_n \tag{2.146}$$

As time increases in the propagation, the state with lowest energy (E_0) will prevail. Since the norm must be restored, one needs to change the Schrödinger equation to:

$$\dot{\Psi}(t) = -[\hat{H} - E(t)]\Psi(t) \qquad \text{with} \qquad E(t) = \langle\Psi(t)|\hat{H}|\Psi(t)\rangle \tag{2.147}$$

The above conditions imply that:

$$\langle\Psi(t)|\dot{\Psi}(t)\rangle = 0 \quad \Longrightarrow \quad \frac{d}{dt}\|\Psi\|^2 = 0 \tag{2.148}$$

The energy E can be understood as a Lagrange multiplier introduced to preserve the norm of the wave function. If we now differentiate $E(t)$ we get:

$$\dot{E} = -2\,\langle\Psi(t)|[\hat{H} - E(t)]^2|\Psi(t)\rangle \tag{2.149}$$

As we can see, the energy decreases with relaxation time and converges if Ψ becomes an eigenstate of $\hat{H}$. This convergence should lead to the ground state of the system. However, if one chooses an initial state that is orthogonal to the GS, the calculations may converge to an excited state.

The relaxation algorithm performs well if the initial guess has sufficient overlap with the GS, and if the latter is well separated energetically from the rest of the states. Unfortunately, relaxation can have slow convergence if the energy of the first excited state (E_1) is close to E_0. The relaxation can be sped up if the A-vector of MCTDH is not computed by relaxation but by diagonalization. This new formalism is called *improved relaxation*, and it allows for the computation of excited states as well.

The new algorithm can be derived by applying the time-independent variational principle with Lagrange multipliers (similar to the process in Section 2.2.1.1):

$$\delta\{\langle\Psi|\hat{H}|\Psi\rangle - E(\sum_J A_J^* A_J - 1) - \sum_{\kappa=1}^{f}\sum_{j,l=1}^{n_\kappa} \epsilon_{jl}^{(\kappa)}[\langle\varphi_j^{(\kappa)}|\varphi_l^{(\kappa)}\rangle - \delta_{jl}]\} = 0 \tag{2.150}$$

The first Lagrange multiplier (E) warrants that the A-vector is normalized and $\epsilon_{jl}^{(\kappa)}$ that the SPFs are orthonormal. After taking the variations with respect to the complex conjugate of both the A-vector and the SPFs independently, we get the following equations [70]:

$$\begin{aligned} \sum_K H_{JK} A_K &= E A_J \\ \frac{\partial \varphi_j^{(\kappa)}}{\partial \tau} &= -(1 - \hat{P}^{(\kappa)}) \sum_{k,l} (\rho^{(\kappa)^{-1}})_{jk} \, \langle \hat{H} \rangle_{kl}^{(\kappa)} \, \varphi_l^{(\kappa)} = 0 \end{aligned} \tag{2.151}$$

The second of these equations implies that we can obtain the updated SPFs simply by relaxation. The A-vector in the first equation can be obtained with the Davidson diagonalization algorithm [177]. Similar to the already described SIL (see Section 2.3.2), the Davidson algorithm is an iterative procedure that do not makes use of Krylov subspace, but uses the uses the Rayleigh-Ritz approach.

To summarize, an improved relaxation calculation starts with an initial guess with reasonably resemblance to the targeted state. Then the matrix representation of the Hamiltonian H_{JK} is built and diagonalized with the Davidson algorithm, which is followed by the construction of the mean fields and the relaxation of the SPFs. After this step, H_{JK} is rebuilt in the space of the relaxed SPFs and the whole process is repeated until a given convergence criterion is met. Improved relaxation can be also used in block form, in other words, it is possible to start the calculations with a set of initial vectors which should converge collectively to a set of eigenstates. The methodology has been successfully applied to a wide range of problems. In the case of tetra-atomic systems (6D) it is generally possible to compute all the desired eigenstates, but the process becomes more challenging for higher dimensions.

2.4 The problem of the PES

One of the most challenging parts of any quantum dynamical calculation is the development of an accurate PES in the suitable form to be efficiently used by the available algorithms. A large amount of fitting procedures have developed over the years [7, 11, and references therein][33]. This approaches range from direct fitting to physically intuitive functions (not necessarily in sums of products form) of *ab initio* data, to sophisticated machine learning and tensor decomposition approaches. In any of the cases, a suitable discretization of the configuration space must be performed.

The PES of a molecular system in its discretized form can be expressed as a f-dimensional tensor whose elements read:

$$V_{i_1,\ldots,i_f} \equiv V_I = V(q^{(1)}_{i_1}, \ldots, q^{(f)}_{i_f}) \ , \tag{2.152}$$

Without loss of generality, we will consider that the grid points have been generated with a DVR. Each DOF (κ) has an associated set of N_κ grid points (and DVR functions). The value of the coordinate ($q^{(\kappa)}$) at the specific grid point (i_κ) is then represented by $q^{(\kappa)}_{i_\kappa}$.

As we already mentioned, this direct mapping onto a grid suffers from the Curse of Dimensionality and its application is limited to systems of low dimensionality. On the other hand, PESs are in many cases smooth functions of the nuclear coordinates, thus allowing for approximations. In other words, one can neglect some information of the global PES without missing its relevant features. The above opens the path to global optimizations of PESs from a set of judiciously chosen points and, correspondingly, to the employment of efficient tensor decomposition algorithms leading to a more compact expression on the grid. Although there are many ways to achieve the latter, the algorithms suffer from lack of uniqueness in the obtained solutions, which complicates the search for a global minima [161]. Furthermore, many of the building blocks (independent subalgorithms) of these methods attempt to solve NP-hard problems, which adds extra layers of complexity to their effective usage.

In the Nuclear Quantum Dynamics field, probably the most widely used tensor decomposition algorithm is the n-Mode Representation (n-MR) [178], developed by Bowman and coworkers, but originally conceived (cut-HDMR) by Rabitz *et al.* [179]. The n-MR expands the potential hierarchically, as the sum of terms of increased dimensionality:

$$V(q_1, q_2, \ldots, q_f) = V^{(0)} + \sum_{j=1}^{f} V^{(1)}_j(q_j) + \sum_{j<k} V^{(2)}_{jk}(q_j, q_k) + \sum_{j<k<l} V^{(3)}_{jkl}(q_j, q_k, q_l) + \ldots \tag{2.153}$$

One particularly impressive application of the n-MR is the computation of the 21D PES of malonaldehyde [180]. Despite its unquestionable success, the n-MR is known to yield PESs with deviated physical behavior if for a given order some terms have been pruned

---

[33] The second of these references correspond to an article recently submitted (and in review at the preset moment) by Panadés-Barrueta and Peláez to The Journal of Chemical Physics. Both articles are included in Chapter 3.

(it might generate holes) [180, 181]. As a matter of fact, it is customary to select only the most relevant terms of a given order when dealing with very large systems. Doing the above, one breaks the variational character of the method, and no error control is granted. In what follows, we will restrict our analysis to variational approaches like the POTFIT method [182, 183] and derivatives.

For the sake of completeness, some alternative approaches of dealing with the problem of the PES in quantum dynamical calculations will be briefly mentioned, leaving suitable references for the interested reader. The first of them is the partially grid-based G-MCTDH method [184], in which a subset of the time-dependent basis functions are substituted by frozen (or thawed) Gaussian functions. If we start from the MCTDH *ansatz* (Section 2.3.2.1), the G-MCTDH would be:

$$\Psi(q_i, \ldots, q_f, t) = \sum_{j_1=1}^{n_1} \cdots \sum_{j_f=1}^{n_f} A_{j_1,\ldots,j_f}(t) \prod_{\kappa=1}^{m} \varphi_{j_\kappa}^{(\kappa)}(q_\kappa, t) \prod_{\kappa=m+1}^{f} g_{j_\kappa}^{(\kappa)}(q_\kappa, t) \tag{2.154}$$

where the $g_{j_\kappa}^{(\kappa)}$ are the aforementioned Gaussian functions. Related to the latter is the Variational Multiconfigurational Gaussian method (vMCG) with its corresponding direct-dynamics (DD) extension [185], which are grid-free and use exclusively Gaussian functions. In this case, the *ansatz* is:

$$\Psi(q_i, \ldots, q_f, t) = \sum_i A_i(t) g_i(q_i, \ldots, q_f, t) \tag{2.155}$$

Finally, a number of methods make use of *on-the-fly* strategies in which global knowledge of PES is sacrificed for the sake of using a local (ideally of low computational cost) approximation of it. Unfortunately, though very promising *on-the-fly* methods are still limited to very modest Electronic Structure levels (see for instance Reference [186]). For a comprehensive discussion of the different approaches, we recommend the introduction of Reference [7].

2.4.1 Tensor Decompositions

In the present section we will address tensor decomposition algorithms in formal way, using the standard mathematical notations. This ideas will be useful to complement the content of the next sections, in which the POTFIT algorithm and derivatives are discussed. The section does not aim to be an introduction the field though. For a complete immersion, the interested reader can consult the excellent review by Kolda *et al.* [161], the articles by De Lathauwer *et al.* [187, 188], and the book by Khoromskaia and Khoromskij [189]. As we mentioned in previous sections, the SOP form is of fundamental importance for the efficiency of the calculations with MCTDH method. The tensor decomposition algorithms here presented allow for transforming an initial reference tensor into the aforementioned form.

We will start by assuming that the reader is familiar with the concept of tensor, and common operations with them, like matricization and vectorization [161]. One of the key features of the algorithms that we will present is the so called *tensor n-mode product*, which is defined as:

$$(\mathcal{X} \times_n \mathbf{U})_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_N} u_{j i_n} \tag{2.156}$$

In the above expression, the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ is multiplied by the matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$. In terms of unfolded tensors, the expression becomes:

$$\mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)} \tag{2.157}$$

where $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{(n)}$ are the corresponding matrix forms of the tensors in mode (dimension) n. Two important properties of this product are:

$$\begin{aligned} &\mathcal{X} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n \mathbf{B} \times_m \mathbf{A} \quad (m \neq n) \\ &\mathcal{X} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n \mathbf{A}\mathbf{B} \end{aligned} \tag{2.158}$$

The concept of tensor n-mode product can be extended also to vectors, but in this case a reduction on the number of dimensions of the original tensor occurs:

$$(\mathcal{X} \times_n \mathbf{v})_{i_1 \dots i_{n-1} i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_N} v_{i_n} \tag{2.159}$$

With this ideas in mind, we can define the Tucker decomposition as:

$$\mathcal{V} = \mathcal{C} \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n \tag{2.160}$$

As we can see, the dimensions of the reference tensor $\mathcal{V}$ and the core tensor $\mathcal{C}$ are the same if the $\mathbf{U}_i$ are matrices of the appropriate shape. The Tucker decomposition can be also reformulated in terms of matrices and vectors in a couple of ways, for example:

$$\begin{aligned} \text{vec}(\mathcal{V}) &= (\mathbf{U}_n \otimes \cdots \otimes \mathbf{U}_1) \cdot \text{vec}(\mathcal{C}) \\ \mathbf{V} &= \mathbf{\Omega} \cdot \mathbf{C} \end{aligned} \tag{2.161}$$

where the matrix $\mathbf{\Omega}$ is equal to the Kronecker product of all the factor matrices in reversed order. Another important property is the so called core tensor unfolding, which reads:

$$\mathcal{C}_{(k)} = \mathbf{U}_k^{-1}\mathcal{V}_{(k)}(\mathbf{U}_n \otimes \cdots \otimes \mathbf{U}_{k+1} \otimes \mathbf{U}_{k-1} \otimes \cdots \otimes \mathbf{U}_1) \tag{2.162}$$

We are now ready to introduce the first of the algorithms that can be used to perform the Tucker decomposition, the Higher-Order Singular Value Decomposition (HOSVD) [187]. The procedure is sketched below: As we can appreciate, the idea of HOSVD is fairly simple. It consists of several matricizations of the reference tensor, followed by the corresponding Singular Value Decompositions (SVD) of the form:

```
Algorithm 2: HOSVD
Result: $\mathcal{C}, \mathbf{U_1}, \ldots, \mathbf{U_n}$
Input: $\mathcal{V}$;
for $k \leftarrow 1$ to $n$ do
    $\mathbf{U}_k \leftarrow R_k$ leading left singular vectors of $\mathbf{V}_{(k)}$
end
$\mathcal{C} \leftarrow \mathcal{V} \times_1 \mathbf{U_1}^{-1} \cdots \times_n \mathbf{U_n}^{-1}$
```

$$\mathbf{V}_{(k)} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^T \tag{2.163}$$

The process is repeated for each of the modes and the left singular vectors (contained in the orthogonal matrix U_k) are kept. At the end of the process, the core tensor is constructed by overlap. One can readily realize that the HOSVD is not optimal in the sense of of giving the best fit to a given selected norm, i.e. it is not variational. The algorithm however offers a great starting for any alternating least-squares approach that aims to refine the solution. This is precisely the motivation behind the Higher-Order Orthogonal Iteration (HOOI) algorithm [188]. Here is very important to mention that a very similar approach was first proposed by Jäckle and Meyer in 1996, namely the POTFIT algorithm [182], four years before HOOI was released. In essence, HOOI performs the following constrained optimization:

$$\begin{gathered}\min_{\mathcal{C},\mathbf{U}_1,\ldots,\mathbf{U}_n} \|\mathcal{V} - \mathcal{C} \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n\| \\ \mathcal{C} \in \mathbb{R}^{R_1 \times \ldots \times R_n} \\ \mathbf{U}_n \in \mathbb{R}^{I_n \times R_n}\end{gathered} \tag{2.164}$$

In the above equations, R_n represents the rank of the corresponding matrix unfolding $\mathbf{V}_{(n)}$. The algorithm goes like this: In this case the algorithm makes use of the intermediate solution tensor $\mathcal{Y}$, and the factor matrices are improving in each iteration until the norm of difference between the reference and the reconstructed tensor reaches a threshold value. For additional details about HOOI, and further extensions to the methodology in general see References [161, 190].

```
Algorithm 3: HOOI
Result: $\mathcal{C}, \mathbf{U_1}, \dots, \mathbf{U}_n$
Input: $\mathbf{U}_1, \dots, \mathbf{U}_n$ from HOSVD;
repeat
    for $k \leftarrow 1$ to $n$ do
        $\mathcal{Y} \leftarrow \mathcal{V} \times_1 \mathbf{U}_1^{-1} \cdots \times_{k-1} \mathbf{U}_{k-1}^{-1} \times_{k+1} \mathbf{U}_{k+1}^{-1} \cdots \times_n \mathbf{U}_n^{-1}$
        $\mathbf{U}_k \leftarrow R_k$ leading left singular vectors of $\mathbf{Y}_{(k)}$
    end
until norm smaller than threshold or maximum number
  of iterations reached;
$\mathcal{C} \leftarrow \mathcal{V} \times_1 \mathbf{U_1}^{-1} \cdots \times_n \mathbf{U_n}^{-1}$
```

Finally, as a complement to Tucker form related methods, we will briefly mention another family of tensors decompositions that finds applications in many branches of natural sciences: the tensor-train decomposition (TT) [191]. In this case the reference tensor is factorized into a product of third order tensors:

$$\mathcal{V}(i_1, \dots, i_n) = G_1(i_1, :)\mathcal{G}_2(:, i_2, :)\mathcal{G}_3(:, i_3, :) \cdots G_n(i_n, :) \tag{2.165}$$

$$= G_1(i_1)\mathcal{G}_2(i_2)\mathcal{G}_3(i_3) \cdots G_n(i_n) \tag{2.166}$$

where the $G_1, \mathcal{G}_2, \dots$ are called TT-cores. The previous expression can be rewritten in index form as:

$$\mathcal{V}(i_1, \dots, i_n) = \sum_{\alpha_1, \dots, \alpha_{n-1}} G_1(i_1, \alpha_1)\mathcal{G}_2(\alpha_1, i_2, : \alpha_2)\mathcal{G}_3(\alpha_2, i_3, \alpha_3) \cdots G_n(i_n, \alpha_{n-1}) \tag{2.167}$$

A graphical representation of the decomposition can be found in References [191, 192]. For a third order tensor, it becomes obvious that the TT and HOSVD are formally equivalent. There exists a number of algorithms that be used to perform the TT. Here we present the one that keeps closer similarity with HOSVD, and it is precisely the TT-SVD algorithm. In the pseudocode below, it is assumed that the TT-cores are of size $r_{k-1} \times N_k \times r_k$ and that the reference tensor is of size $N_1 \times \cdots \times N_n$. The main idea behind the procedure is to perform n sequential truncated SVDs that start from suitably reshaped tensors. We recommend the seminal work by Oseledets [191] for a complete analysis.

Many of the concepts we have discussed in this section are of central importance for the present work, in which two new methodologies that allow for obtaining PESs in the suitable SOP form has been developed. Tensor decomposition algorithms as introduced here and in the next two sections are a fundamental element of the meta-algorithms that were conceived together with the aforementioned methodologies. For more details about the subject, the reader is invited to consult Chapter 3.

```
Algorithm 4: TT-SVD
Result: $G_1, \mathcal{G}_2, \dots, G_n$
Input: $\mathcal{V}$;
Set: $\mathcal{W} = \mathcal{V}$, $r_0 = 1$;
for $k \leftarrow 1$ to $n-1$ do
    $W \leftarrow \texttt{reshape}(\mathcal{W}, [r_{k-1}N_k, \frac{\mathrm{len}(W)}{r_{k-1}N_k}])$;
    Compute truncated SVD of reshaped $\mathcal{W}$:
     $W = U\Sigma V^T + E$;
    $G_k \leftarrow \texttt{reshape}(U, [r_{k-1}, N_k, r_k])$;
    $W \leftarrow \Sigma V^T$;
end
$G_n \leftarrow W$
```

2.4.2 The POTFIT algorithm

It was mentioned at the end of Section 2.3.2.1 that the PES is not necessarily in the SOP form needed by MCTDH, and an additional step is often necessary in order to fulfill this requirement. The POTFIT algorithm [182, 183] (or HOOI in mathematical community) transforms a general tensor into Sum of Products (SOP) form. The algorithm is variational and grid based, and the working SOP expression is the so called Tucker form:

$$V^{PF}_{i_1,\dots,i_f} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_f=1}^{m_f} C_{j_1\dots j_f} v^{(1)}_{i_1\ j_1} \cdots v^{(f)}_{i_f\ j_f} \tag{2.168}$$

As we can see, the reference tensor is decomposed into a smaller tensor of the same dimensionality (C) multiplied by a set of matrices representing the basis functions evaluated on the grid. This decomposition is represented graphically in Figure 1 of Reference [11] for the case of a 3D tensor. One can infer from Equation 2.168 that the POTFIT expansion needs to store $m^f + fNm$ data points, which is way smaller than the original N^f reference data points required for the PES on the full grid.

In POTFIT the coefficients are obtained by overlap (*tensor n-mode product*, Equation 2.156) of the reference tensor with the matrices of basis functions expressed on the grid, which are usually called *single-particle potentials* (SPPs):

$$C_{j_1\dots j_f} = \sum_{i_1\dots i_f} V_{i_1\dots i_f} v^{(1)}_{i_1\ j_1} \cdots v^{(f)}_{i_f\ j_f} \tag{2.169}$$

As it was mentioned before, the POTFIT algorithm is closely related to other tensor decomposition methods, in particular the Higher-Order Singular Value Decomposition (HOSVD) [187] and the Higher-Order Orthogonal Iteration (HOOI) [188]. In order to homogenize POTFIT notation with the one in the mathematical community, we need to mention that the operation in Equation 2.169 can be also written as:

$$\mathcal{C} = \mathcal{V} \times_1 \mathbf{v}^{(1)} \times_2 \mathbf{v}^{(2)} \cdots \times_f \mathbf{v}^{(f)} \tag{2.170}$$

where $\mathcal{C}$ and $\mathcal{V}$ are the core and reference tensors, and the $v^{(\kappa)}$ are called factor matrices. The number of expansion terms in the previous expressions (which are the main computational limitation in MCTDH method) can be reduced by performing a contraction of the coefficients:

$$D^{(\nu)}_{j_1\dots j_{\nu-1} i_\nu j_{\nu+1}\dots j_f} = \sum_{j_\nu} C_{j_1\dots j_f}\; v^{(\nu)}_{i_\nu\, j_\nu} \tag{2.171}$$

After this step, the POTFIT expansion becomes:

$$V^{PF}_{i_1,\dots,i_f} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_f=1}^{m_f} D^{(\nu)}_{j_1\dots j_{\nu-1} i_\nu j_{\nu+1}\dots j_f}\, v^{(1)}_{i_1\, j_1} \cdots v^{(\nu-1)}_{i_{\nu-1}\, j_{\nu-1}} v^{(\nu+1)}_{i_{\nu+1}\, j_{\nu+1}} \cdots v^{(f)}_{i_f\, j_f} \tag{2.172}$$

where the number of elements of the core tensor has changed from $\prod_\kappa j_\kappa$ to $\prod_{\kappa\neq\nu} j_\kappa \cdot i_\nu$, leading to the contracted core tensor D. Moreover, the SPPs are eigenvectors of the potential density matrices whose elements are defined as:

$$\rho^{(\kappa)}_{kk'} := \sum_{I^\kappa} V_{I^\kappa_k}\, V_{I^\kappa_{k'}} \tag{2.173}$$

The above matrices are squared and positive semi-definite. One can see that there is one matrix for each DOF with dimensions $N_\kappa \times N_\kappa$. Once diagonalized, each density matrix yields a set of N_κ eigenvectors and eigenvalues. The eigenvectors are called *natural potentials*, and are employed as the set of SPPs in the POTFIT expansion. If all natural potentials are used in the expansion, then it reproduces the original PES, i.e the expansion 2.168 is exact. In practice however, it is possible (and convenient) to trim down the full expansion up to a given accuracy. In order to do this, the corresponding eigenvalues (which are called *natural weights*) are taken as deciding factors, because they assess the relevance of the associated eigenvector in the expansion. Finally, it should be noted that the POTFIT error is typically computed by iterating over all grid points and comparing the value obtained after the decomposition with the value of the reference PES. It has been demonstrated that the error is proportional to the sum of the neglected natural weights [182, 183].

The POTFIT algorithm has proven to be very useful in a wide range of problems (see for example Reference [193]). Unfortunately, in order to perform the tensor decomposition and obtain the expansion coefficients and corresponding SPPs, one has to use the full grid. This fact prevents an efficient usage of the algorithm for systems with dimensionality larger than 6-7D. Motivated by this fact, further approximations have been developed in order to overcome the “dimensionality hill”. The most prominent examples are the Multigrid POTFIT (MGPF) algorithm [8], the Monte-Carlo POTFIT algorithm (MCPF) [194] and its Canonical Polyadic Decomposition (CPD) formulation [195], and the Multilayer POTFIT [196, 197] algorithm. Due to its relevance to the present work, MGPF will be discussed in more detail on the next section.

2.4.3 The Multigrid POTFIT (MGPF) algorithm

As we mentioned before, the application of the POTFIT algorithm is limited to systems with dimensionality lower than 7D, which is around 10^9 grid points. On the other hand, the method has very convenient error control properties due to its variational character. The Multigrid POTFIT algorithm was designed to push the dimensionality boundaries while keeping some of the desirable features of the original POTFIT [8]. The method itself is based on a series of underlying POTFIT calculations, and it the same spirit of its predecessor, it allows for decomposing a reference tensor (up to $\sim 12D$) into Tucker form:

$$V_{i_1 \dots i_f} \approx V^{MGPF}_{\tilde{i}_1 \dots \tilde{i}_f} = \sum_{j_1=1}^{m_1} \dots \sum_{j_f=1}^{m_f} C_{j_1 \dots j_f} \, \tilde{v}^{(1)}_{\tilde{i}_1 j_1} \dots \tilde{v}^{(f)}_{\tilde{i}_f j_f} \tag{2.174}$$

The MGPF decomposition is formally equivalent to the one of POTFIT, but relevant quantities (core tensor and factor matrices) are computed in a different manner. As can be inferred from its name, MGPF uses several grids in the calculations. In order to continue our discussion, the nomenclature used so far needs to be extended. The new definitions are: the fine grid ($\tilde{I}$), which is the grid needed to fully converge a MCTDH calculation, and as in POTFIT, each κ-th DOF has N_κ fine grid points; the coarse grid (I), which is a subset of the fine grid and each κ-th DOF has n_κ coarse grid points; and the so called partial grids ($I^\kappa_{\tilde{i}_\kappa}$), in which the κ-th DOF lies on the fine grid whereas the rest lie on the coarse [8]. If one selects a number of grid points per DOF (N or n) that ensures physical meaning for the calculations performed on an f-dimensional system, then the fine grid will contain N^f grid points, the coarse grid n^f, and the partial grids $N \cdot n^{f-1}$.

In the framework of MGPF, the coefficients are obtained as in POTFIT, by overlapping the reference PES with the factor matrices (Equation 2.169). *Au contraire*, the factor matrices are obtained by optimizing the $\mathcal{L}^2$ norm of the difference between the reference potential and an approximated one, both expressed on a partial grid:

$$\sum_{I^\kappa} \sum_{\tilde{i}_\kappa} \left(V_{I^\kappa_{\tilde{i}_\kappa}} - V^{app}_{I^\kappa_{\tilde{i}_\kappa}} \right)^2 = \min \quad \kappa = 1, \dots, f \tag{2.175}$$

After an elaborated algebraic procedure[34], and taking as many coarse grid points as MGPF SPPs, the Equation 2.175 can be transformed to yield:

$$\tilde{\mathbf{v}}^{(\kappa)}_{N\times n} = \boldsymbol{\rho}^{(\kappa)'}_{N\times n}\, \boldsymbol{\rho}^{(\kappa)^{-1}}_{n\times n}\, \mathbf{v}^{(\kappa)}_{n\times n} \qquad \kappa = 1,\ldots,f \tag{2.176}$$

The above are the MGPF working equations. The new quantity ρ' is called semi-extended potential density matrix, and it is computed as in Equation 2.173 but with the left index running over the fine grid. At this point it can inferred that MGPF is exact on the coarse grid. By analyzing Equation 2.176, one can realize that what the algorithm effectively does is to interpolate the SPPs on the coarse grid (v) to SPPs on some fine grid ($\tilde{v}$). MGPF enjoys of an excellent linear scaling with respect to the system size. In contrast to POTFIT, in which one needs to iterate over all grid points to obtain the core tensor and SPPs, this new method operates on the coarse grid or on the partial grids, respectively.

One of the key aspects of the method is the selection of a coarse grid that captures the physics of the problem accurately. There is not a universal guideline to do this, except for the fact that larger coarse grids will yield better results. Nonetheless, the particularities of the PES at hand may lead to dramatic differences in the quality of the results obtained with different coarse grids. Motivated by this fact, Peláez *et al.* proposed a DVR-like process for the optimization of the coarse grid [8]. To do this, MGPF builds coordinate matrices (Q) defined as (using index space notation):

$$Q^{(\kappa)}_{jk} = \sum_{\tilde{i}_\kappa} \hat{v}^{(\kappa)}_{\tilde{i}j} \tilde{i}_\kappa \hat{v}^{(\kappa)}_{\tilde{i}k} \tag{2.177}$$

where the $\hat{v}$ are the orthogonalized MGPF SPPs. The matrix Q is then diagonalized, and its eigenvalues are rounded-off to match the closest integer corresponding to the index of the coarse grid point. This creates two possible operation modes for MGPF, the *bottom-up* (bu) and *top-down* (td) approaches. The former uses a small coarse grid, whereas the latter uses the largest coarse grid possible. For a complete analysis of the performance of the different MGPF flavors with POTFIT, see Reference [198].

The MGPF algorithm represent a substantial improvement of POTFIT that keeps its more valuable features, i.e. the variational character. The method allows for a fast and accurate discrete representation of PES in Tucker form. Most importantly, it helps to overcome in some extent the curse of dimensionality that limits POTFIT, allowing for the treatment of systems with up to 12D.

---

[34] See the original work by Peláez *et al.* [8] for a complete derivation.

2.5 Automated topographical studies: the vdW-TSSCDS method

Topographical knowledge of the global (inter)-molecular PES is a prerequisite for the simulation of a dynamical process. The characterization of a PES is usually a tedious task, which often relies on large amounts of chemical intuition for the elucidation of the guess structures from which the stationary points will be optimized. The process becomes even harder for systems of large dimensionality, or when dealing with highly flexible systems of many stationary points. Motivated by this, a number of approaches have been developed in order to find individual minima and transition states (TSs).

Starting from some *a priori* information about minima (pivotal structures), the corresponding TSs can be determined by using single-ended [199, 200] or double-ended [201–204] methods. The crucial step for this task is the automatic identification of suitable guess structures. With this purpose different algorithms have been proposed in recent years [4, and references therein]. Among the latter, a recent and very promising one is the *Transition State Search using Chemical Dynamics Simulations* (TSSCDS) method [5, 6, 205], which combines accelerated direct dynamics with an efficient algorithm of geometry processing to determine stationary points on the PESs, starting from a single (or several if desired) input geometries [4]. Many of the previous methods are based on chemical reactivity arguments, i.e. the breaking and formation of covalent bonds [5, 6, 206, 207]. The above implies that they are not suitable (or still not adequate in their present status) to treat non-covalently bound systems. These weakly interacting systems (like van der Waals complexes or hydrogen bonding) are of great importance in many biologically, chemically and physically relevant problems [4, 208, and references therein]. Examples of this are many astrochemical [209] and atmospherical systems [2, 210]. It is precisely this fact what motivated Kopec *et al.* to extend the TSSCDS formalism in order to include explicit support for weakly bound systems, in what is known as vdW-TSSCDS. The contents on this section are largely based on their work [4].

Before going into the details of the aforementioned extension, the TSSCDS method should be described. A schematic representation of the ideas behind this algorithm can be found in Figure 2 of Reference [4]. Starting from one initial random geometry (or a set of them), the software performs a geometry optimization and corresponding frequency calculation, employing a low-level (LL) electronic structure method. A very common choice for the LL is a semiempirical Hamiltonian (see Section 2.2.3), but the use of other methods with low computational requirements (like DFTB, [211]) is also possible.

The next step is efficient sampling of the configuration space by launching a set of high-energy classical trajectories (microcanonical or canonical), whose initial conditions are obtained after normal mode or canonical sampling. Consecutive geometries in each trajectory are periodically investigated by the Bond Breakage / Formation Search (BBFS), which studies the temporal evolution of chemical bonds in the molecular system to offer guess TSs structures[35]. The BBFS algorithm relays on the graph theory concept of the *connectivity* or *adjacency* matrix **C** [212]. The matrix elements c_{ij} are defined in Equation 2.178. As it can be appreciated, they depend on the normalized distance δ_{ij}, defined as the ratio of the distance d_{ij} between every two atoms i and j and a reference distance d_{ij}^{ref}. The latter should be chemically meaningful, and in TSSCDS it is chosen to be 10 % greater than the sum of the corresponding covalent radii for the sake of flexibility. If the value of δ_{ij} is smaller than one, the two atoms are taken as chemically bound and are therefore identified as neighbors.

$$c_{ij} = \begin{cases} 1 & \text{if } \delta_{ij} < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{, with } \delta_{ij} = \frac{d_{ij}}{d_{ij}^{ref}} \tag{2.178}$$

In the next step, the guess TSs are used as starting points for geometry optimizations at LL. The optimized structures are then screened in order to detect duplicates. This process compares a weighted connectivity matrix as well as a set of SPRINT (social permutationally invariant) coordinates for each structure [213], and compares them with those of the remaining structures in order to select the unique ones. Intrinsic reaction coordinate (IRC) [214] calculations are then performed for the unique TSs, followed by optimizations of the final IRC structures, which are then classified into minima and reaction products on the PES. This information allows for constructing the reaction network (RXN) of the system, which compiles the connections between the different SPs.

With the optimized LL transition states in hand, the algorithm proceeds to reoptimize them using a suitable high-level (HL) *ab initio* electronic structure method. In this way, the low-level TSs act as guess structures for the corresponding high-level optimizations. In the same spirit as for the LL, the IRC calculations are launched for all HL transition states leading to minima and products. The final step consists in the construction of the reaction network at the HL, allowing for a complete topographical characterization of the PES.

---

[35] See Figure 3 of Reference [4] for a graphical representation of the time evolution of an example molecular graph. The changes on the lengths of the edges are analyzed and compared with the reference distance, and the corresponding change on the connectivity matrix determines the presence or not of a possible TS.

Even though the TSSCDS is a very robust algorithm, it lacks support for detection of van der Waals bonds. The method is therefore *blind* to TSs corresponding to weak intermolecular interactions. To improve this weak point, Kopec *et al* [4] proposed to generalize the connectivity matrix in a block structure accounting for the different fragments that make up the system as well as their interactions. This generalization can be better understood with a bimolecular example[36] [4]. We will consider a system with two well-identifiable fragments A and B. As shown in Eq. 2.179, the diagonal blocks in the connectivity matrix $\mathbf{C}$ refer to the intramolecular distances of the individual molecules (fragments A and B), whereas the off-diagonal ones correspond to intermolecular distances (the interacting system AB). As in the TSSCDS case, the matrix elements C_{ij} are computed according to Eq. 2.178, but based on normalized interatomic distances.

The major difference of vdW-TSSCDS with respect to its predecessor, is that bonds of different nature are taken into account by adjusting of their reference distance d_{ij}^{ref}. For the case of the diagonal blocks, the normalized distance δ_{ij} is obtained using the sum of the covalent radii as the reference distance, whereas in the off-diagonal blocks AB, it is chosen to be (10 % greater than) the sum of the van der Waals radii of each two atoms [4]. By inserting this changes, a weak interaction can be treated on the same way as a valence or a hydrogen bond, enabling the detection of structures with changes on the connectivity between individual fragments A and B as a possible transition state (first order saddle point). An important remark here is that, by simply taking the sum of the van der Waals radii as the overall reference distance, one can simulate additional bonds within the covalently bound fragments, which could be considered as transition state guess structures.

$$\mathbf{C} = \begin{pmatrix} \boxed{\mathbf{A}} & \boxed{\mathbf{AB}} \\ \boxed{\mathbf{AB}} & \boxed{\mathbf{B}} \end{pmatrix} \tag{2.179}$$

The BBFS algorithm has been adapted to generate the vdW-TSSCDS block-structured connectivity matrix, which remains unaltered during the whole process. After suitable initial conditions are established, a set of classical trajectories (generated without constraints regarding the nature of the bonds) are propagated during a given time. The geometries generated along each of the trajectories are then analyzed at a fix time-interval as in the TSSCDS case. The analysis of the connectivity matrices is done element-wise, checking in each case if the element changed from zero to one or from one to zero. As described by Kopec *et al* [4], four different situations are possible:

---

[36] The extension to the n-molecular case is straightforward

1. Off-diagonal and on-diagonal elements are stationary. BBFS cannot detect any TS.
2. Stationary on-diagonal elements with non-stationary off-diagonal elements. BBFS detects a guess TS structure for a weakly bound system.
3. Stationary off-diagonal elements with non-stationary on-diagonal elements. BBFS detects a guess TS structure for one (or several) independent fragment.
4. Off-diagonal and on-diagonal elements are non-stationary.

When facing cases 2 to 4, an unconstrained LL optimization is performed, which may lead to any type of TS (non-covalent, reactive, etc.). It is noteworthy that reaction among fragments can be reflected in cases 2 and 3 independently, but also (and most likely) as a combination of both. The resulting TSs are specific to the system in hand, and cannot be predicted a priori. The use of an initial guess that is non-covalently bound is of great relevance for an efficient usage of vdW-TSSCDS. If the study of a bound system is desired, the algorithm switches automatically to the original TSSCDS implementation.

One can attempt to describe molecular systems consisting of two or more well-identifiable fragments in several ways, ranging from the rigid approaches to the fully flexible *ansatz*. The former, keeps all intramolecular parameters frozen while allowing intermolecular ones to relax [215], thus providing an upper bound to the interaction energy and allowing for the accurate study of intermolecular vibrations. The latter allows for a complete description of the interaction energy, spectroscopy and reactivity of the complexes, but due to the necessary increase on dimensionality, it is only practical for small-medium systems. The vdW-TSSCDS algorithm can be used in all of this scenarios [215].

2.6 A survey on non-linear optimization algorithms

Optimization algorithms are present in almost every area of human activity. Our understanding of economy, natural sciences, engineering and mathematics would be utterly limited without the development of efficient optimization strategies [216–218]. Obviously, quantum chemistry is not an exception: finding the energy of a molecular system (one of the principal objectives of the field) is always an optimization problem! In the current section the generalities of numerical optimization will be introduced, followed by the discussion of some algorithms that have direct relevance to this work.

The general optimization problem can be stated as follows:

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{array}{lll} c_i(x) & = 0, & i \in \mathcal{E} \\ c_j(x) & \geq 0, & j \in \mathcal{I} \end{array} \tag{2.180}$$

The objective is to minimize the target function $f(x)$, which is a mapping of the kind $\mathbb{R}^n \rightarrow \mathbb{R}$, possibly subjected to a set of equality and inequality constraints $c_i(x)$. Of course, the above definition is shared by a large number of different problems. We can have constrained and unconstrained problems, continuous and discrete, global and local, stochastic and deterministic, and much more.

In our particular case, we are interested in the global optimization of non-linear problems, possibly of *black box* functions whose derivatives are not available [219]. This sort of task in never an easy one, and depending on the algorithm, there might be or not reassurance that a global minimum has been found at the end of the process. Moreover, there is not much of a systematic approach for choosing the right algorithm for these kind of problems, but a useful guideline is provided by the *No Free Lunch* Theorems for Optimization (NFL) [220]. The first of these theorems reads:

Theorem 2.3. *NFL 1 For any pair of algorithms a_1 and a_2*

$$\sum_f P(d_m^y | f, m, a_1) = \sum_f P(d_m^y | f, m, a_2) \tag{2.181}$$

In the above equation the sums run over all possible objective functions f, and the sums of the probabilities $P(d_m^y | f, m, a_1)$ of obtaining minimum d_m^y after m iterations is computed for each algorithm. The principal conclusion that we can extract from the theorem is that there is no *a priory* answer for the best algorithm for a given problem, because what one algorithm gains in performance on a given type of problem, is necessarily offset by its performance on another kind of problems [220]. With this ideas in mind, we will proceed to show some of the algorithms that we have used in the present work. We will present first the global optimizers (which are of stochastic nature in our case) and then the local ones.

The first of the global optimizers is the well known Basinhopping algorithm (BH) [221], whose underlaying principle is to perform global searches of the full configuration space using a stochastic Metropolis Monte Carlo (MC) procedure. The strong point of BH is that it combines the heuristic component with local searches that are launched after a series of perturbations. A simple implementation of the algorithm, inspired by the one proposed by Ferreiro *et al.* in Reference [222], is shown in Algorithm 5.

As it can be appreciated, BH is an iterative procedure in which every time a random uniform sample is drawn from a multidimensional ball B of radius r_k and centered in x_k (the current minima), then the difference between the function evaluated in the new sample and the previous minima is evaluated and submitted to a MC acceptance test. The radius and the temperature T are optionally adjustable, and a cooling factor can be introduced. Normally a combo of termination conditions can be used, involving the total number of iterations N, the minimal temperature T_{min}, the relative tolerance of the found solution or the maximum number of rejected solutions J. Note that in every step, only the knowledge of the current best solution is strictly required, which greatly alleviates memory consumption in case of large number of iterations. As outlined before, BH makes heavy use of a local optimizer which should be efficient and accurate enough to make the procedure viable. Some of this algorithms will be presented further on this section.

Algorithm 5: Basinhopping

```
Result: $x_{min}$
Input: $x_{guess}$ possibly random starting point;
$j \leftarrow 0$ number of successive rejections;
$x_i \leftarrow$ localopt($x_{guess}$);
while $j < J \vee T < T_{min}$ do
    for $i \leftarrow 0$ to $N$ do
        $y_i \leftarrow$ random uniform $B(x_k, r_k)$;
        $u \leftarrow$ random uniform $[0, 1]$;
        $\Delta \leftarrow f(\texttt{localopt}(y_i)) - f(\texttt{localopt}(x_i))$;
        if $u < exp(-\Delta/kT)$ then
            $x_{i+1} \leftarrow$ localopt($y_i$);
            $j = 0$;
        else
            $j \leftarrow j + 1$;
        end
        $i \leftarrow i + 1$;
    end
    Update radius $r_k$ and temperature $T$;
end
$x_{min} \leftarrow y_i$;
```

The next important global optimizer that we will show here is the Multi-Level Single Linkage (MLSL) [223] algorithm. This procedure was derived from clustering methods, which enable the exploration of the accessible region of the configuration space by performing random samplings followed by local optimizations. Although loosely similar in concept to BH, the MLSL algorithm differs substantially from the former. The simplest implementation of MLSL [224] is represented in Algorithm 6. The first step is to define a subset $X_s \subset X$ of the whole search set X that will contain the candidate solutions. In each iteration random samples of N points are drawn from X and added to a cumulative sample set. Then the best γkN solutions are taken from the latter set and added to the reduced sample X_r, with $0 < \gamma \geq 1$ being a constant chosen a constant chosen *a priory*. Each of the elements in X_r are then submitted to a local optimization, making sure none of the points in this set are separated more than the critical radius r_k, which is computed as [224]:

$$r_k(x) = \frac{1}{\sqrt{\pi}}\left(\Gamma(1+\frac{n}{2}) \cdot m(X) \cdot \frac{\zeta \ln(kN)}{kN}\right)^{\frac{1}{n}} \tag{2.182}$$

In the above equation, n represents the number of dimensions of the problem in hand, Γ is the extended factorial function, $m(X)$ is the Lebesgue measure of the domain X, ζ is an arbitrary constant, k is the number of iterations, and kN the total number of sampled points. As we can see, the usage of r_k represents a solid heuristic tool to avoid exploring twice the same region of the configuration space.

```
Algorithm 6: MLSL
Result: $x_{min}$
Input: search domain $X \subset \mathbb{R}^n$;
$k \leftarrow 0$, $X_s \leftarrow \varnothing$;
repeat
    $k \leftarrow k + 1$;
    $x_{(k-1)N+1}, \dots, x_{kN} \leftarrow$ uniform random sample of $N$
      points from $X$;
    $X_r \leftarrow$ reduced sample of the $\gamma kN$ best points from
      cumulative sample $x_1, \dots, x_{kN}$;
    for $i \leftarrow 0$ to $len(X_r)$ do
        if $\nexists j : f(x_j) < f(x_i) \wedge \|x_j - x_i\| < r_k$ then
            $y_i \leftarrow$ localopt($x_i$);
            $X_s \leftarrow X_s \cup \{y_i\}$;
        end
    end
until relative error smaller than threshold or maximum
  number of iterations reached;
$x_{min} \leftarrow \min f(X_s)$
```

The local minima that survived this search procedure are then added to the set of solutions X_s. The process repeats until one of the local minima has a relative tolerance lower than a threshold value, or when the total number of iterations is reached. As in the case of BH, the selection of the local optimizer is critical here, as well as the random sampling from the set X. Regarding this last aspects, one usually uses Sobol quasi-random sequences [225], which are known to sample the configuration space in a more homogeneous way than classical pseudo-random numbers generators.

The global optimizers that we presented here are not a representative part of the vast landscape of algorithms of this kind available. Some notable exceptions are the Genetic Algorithms (GA) [226] and Differential Evolution (DE) [227], Particle Swarm Optimization (PSO) [228], and Generalized Stimulated Annealing (GSA) [229]. The interested reader is invited to consult the recommended literature.

The next part of this section is dedicated to local optimizers, which as the name suggests, allow for obtaining the minimum of a specific region of the configuration space. We should mention here than in many practical cases one does not need to use a global optimizer, either because of the physics of the problem at hand or because a good starting guess solution is known. For example, the optimization of the cost functions in Neural Networks usually proceeds with a simple Gradient Descend (GD) algorithm [230], whose great popularity and efficiency is justified by the availability of analytical derivatives computed with the robust back-propagation algorithm [231].

The first local optimizer we show is the so called Powell algorithm (or Powell's method) [232]. Named in this way to honor its developer M.J.D. Powell[37], this algorithm belongs to the family of the conjugate direction methods, and is derivative free. The method relays on the properties of a hypothetic convex and quadratic objective function (although the last requirement is only formal, and the method is easily extended to non-quadratic functions):

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \tag{2.183}$$

In this conditions, two directions $\mathbf{d}_i$ and $\mathbf{d}_j$ are mutually conjugate if:

$$\mathbf{d}_j\mathbf{H}\mathbf{d}_i = 0 \quad i \neq j \tag{2.184}$$

The method proceeds by performing consecutive optimizations on a series of directions $\mathbf{d}_i$ that lead to a new updated solution, and a corresponding updated set of directions $\mathbf{d}_{i+1}$. After n iterations it is ensured that all the directions $\mathbf{d}_i$ are conjugate [217]. The algorithm can be implemented as follows:

It was shown by Powell in its original work, that a local minima can be found by looking at each of the conjugate directions only once [232]. The great advantage here is that Hessians are not needed, and with the assistance of 1D algorithms based on function evaluations for line searches, the calculation of gradients is also skipped. The strategy of the algorithm is simple and yet quite powerful, which makes it one of the most popular derivative free optimizers [217]. A well-known drawback of the method is the appearance

---

[37] Michael James David Powell (1936-2015) was a a British mathematician who worked extensively on numerical analysis, developing several algorithms that have become standards on the field, like COBYLA, NEWUOA and BOBYQA [233].

```
Algorithm 7: Powell
Result: $x_{min}$
Input: $x_{guess}$ possibly random starting point;
Define search vectors $d_i : i = 1, \dots, n$;
$x_0 \leftarrow x_{guess}$;
repeat
    for $k \leftarrow 1$ to $n$ do
        $\lambda_k \leftarrow \min_{\lambda_k} f(x_{k-1} + \lambda_k d_k)$;
        $x_k \leftarrow x_{k-1} + \lambda_k d_k$;
    end
    for $j \leftarrow 1$ to $n-1$ do
        $d_j \leftarrow d_{j+1}$;
    end
    $d_n \leftarrow x_n - x_0$;
    $\lambda_n \leftarrow \min_{\lambda_n} f(x_{n-1} + \lambda_n d_n)$;
    $x_0 \leftarrow x_0 + \lambda_n d_n$;
until relative error smaller than threshold or maximum
  number of iterations reached;
$x_{min} \leftarrow x_0$
```

of spurious linear dependences between the search directions, which results in automatic lost of convergence.

The second local optimizer that we will presents belongs to the family of the quasi-Newton methods, so it only requires the gradient and not the Hessian of the objective function. The algorithm is usually called BFGS after his discoverers Broyden, Fletcher, Goldfarb, and Shanno [234, and references therein]. Similarly to Powell's method, we start our analysis with a quadratic model including the objective function f at the iteration k:

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \tag{2.185}$$

where B_k is a symmetric positive definite matrix that will be updated in every iteration. The previous problem is convex, and the minimizer p_k can be expressed as:

$$p_k = -B_k^{-1} \nabla f_k \tag{2.186}$$

This minimizer is used as a search direction:

$$x_{k+1} = x_k + \alpha_k p_k \tag{2.187}$$

where the α_k is a minimization parameter that satisfies suitable conditions [216]. The main difference with respect to Newton's method is that B_k is not rebuild in each iteration, but it is updated in a clever manner that accounts for the curvature measured in the previous step. The proposed updating scheme is:

$$B_{k+1}\alpha_k p_k = \nabla f_{k+1} - \nabla f_k \tag{2.188}$$

To simplify the notation we will introduce the following vectors:

$$s_k = x_{k+1} - x_k = \alpha_k p_k; \qquad y_k = \nabla f_{k+1} - \nabla f_k \tag{2.189}$$

Equation 2.188 implies extra constraints to the curvature of the objective function which always verified if the problem is convex, but must be imposed otherwise. In practice, the BFGS works with the inverses $H_k = B_k^{-1}$, so the determination of H_{k+1} proceeds by the minimization problem:

$$\begin{aligned} H_{k+1}y_k &= s_k \\ \min_H &\|H - H_k\| \\ \text{subject to} \quad H &= H^T; \quad Hy_k = s_k \end{aligned} \tag{2.190}$$

It can be shown (see Reference [216] for a proof) that the above problem has a unique solution of the form:

$$\begin{aligned} H_{k+1} &= (I - \rho_k s_k y_k^T)H_k(I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \\ \rho_k &= \frac{1}{y_k^T s_k} \end{aligned} \tag{2.191}$$

The above expression is the the core of the BFGS algorithm. The question of finding an adequate initial guess for the inverse Hessian H_0 does not have a simple answer, and in many cases it is computed by finite differences at the initial point. The BFGS procedure can be then resumed as:

```
Algorithm 8: BFGS
Result: $x_{min}$
Input: $x_{guess}$ possibly random starting point, threshold $\epsilon$;
$x_0 \leftarrow x_{guess}$;
$H_0 \leftarrow$ finite differences approximation on $x_0$;
$k \leftarrow 0$;
while $\|\nabla f_k\| < \epsilon$ do
    $p_k \leftarrow -H_k \nabla f_k$;
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$ line search procedure;
    $s_k \leftarrow x_{k+1} - x_k$;
    $y_k \leftarrow \nabla f_{k+1} - \nabla f_k$;
    $H_{k+1} \leftarrow (I - \rho_k s_k y_k^T)H_k(I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T$;
    $k \leftarrow k + 1$
end
$x_{min} \leftarrow x_k$
```

In the present section we have discussed some algorithms that are frequently employed to solve non-lineal optimization problems. As suggested before, the content can be greatly complemented by consulting the reference literature. However, what we have described in the last two sections of this Chapter will become useful when reading the articles presented in Chapter 3. The algorithms that we have analyzed in both sections are a fundamental part of the workflow of the methods developed in this thesis.

Chapter 3

Representation of the interaction potential

On peut braver les lois humaines, mais non résister aux lois naturelles.
Jules Verne, *Vingt mille lieues sous les mers*

3.1 Introduction to the study

In the present chapter we deal with the problem introduced in Section 2.4. As previously discussed, one of the most important and laborious parts of any nuclear quantum dynamics calculation is the development of the PES of the system in consideration, possibly in the suitable mathematical form required by the method of choice. In order to meet the goals of this work (see Section 1.3), we have developed two methodologies and corresponding software packages that aim at solving the aforementioned task. Both methods are introduced here by means of the corresponding articles[1], one of them already published and the other one in reviewing status.

The first method is named Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF) [7]. This automated approach allows for obtaining global PESs (or molecular properties surfaces in general) starting from a set of judiciously chosen reference geometries. The SRP-MGPF method can be divided in three main logical stages: the automated (global) topographical characterization of the PES based on the Transition State Search Using Quantum Dynamical Simulations (TSSCDS) family of methods; the reparametrization

---

[1] The author of the present doctoral dissertation is the first author of both publications.

of a semiempirical Hamiltonian using a global optimization algorithm in order to reflect the chemistry of a given level of electronic structure theory; and an efficient tensor decomposition process that transforms the obtained semiempirical PES into sum of products (Tucker) form using the Multigrid POTFIT algorithm. As it was explained in Sections 2.4.2 and 2.4.3, this last step is necessary for performing quantum dynamical studies with the MCTDH method. The approach was validated with a well know benchmark system: the *cis-trans* isomerization of the HONO molecule in full dimensionality (6D) [14]. The reader is encouraged to complement the lecture of this publication (included in Section 3.2) with the discussions about semiempirical methods in Section 2.2.3, and optimization algorithms in Section 2.6.

The second method is called Sum-of-Products Finite-Basis-Representation (SOP-FBR) [11]. As in the previous case, the approach was designed with automaticity and simplicity of usage in mind, without sacrificing accuracy on the calculations. The method can be employed in a *black box* fashion if desired, but the corresponding software package will be available upon request for more advanced users. The current implementation of SOP-FBR generates PESs in the so-called Tucker Sum-Of-Products form, but it can be easily extended to support other *ansätze*. The originality of this new approach relies in the fact that the optimization is carried out in terms of a direct product of *Schmidt* basis, which are commonly known as *natural potentials* in the MCTDH literature (see Section 2.4.2). This basis set contains the Physics of the problem encoded in a non-trivial way, which sustains the hypothesis of removing the usual additional *ad hoc* modifications of the obtained PES. Additionally, the intermediate *refitting* step prevalent in other tensor-decomposition methods with applications in nuclear quantum dynamics is suppressed.

The PESs generated by SOP-FBR are analytical and infinitely differentiable. The method is compatible with most molecular dynamics codes commonly used. In our specific case, we have interfaced it with the Heidelberg implementation of the MCTDH method. The lecture of the corresponding article (see 3.2) can be complemented with the content about tensor decomposition algorithms in Section 2.4.1, and the discussion about optimization algorithms in Section 2.6. In addition to this, in the next section we examine the SOP-FBR *ansatz* in more detail, and show a transformation of it that could be applied to future studies.

3.1.1 Transformation of Tucker expansions of products of Chebyshev series

In this subsection we will present an interesting transformation that can be applied to the SOP-FBR *ansatz* in order to simplify the optimization process. For the sake of completeness, both the algebraic and the tensor approaches will be described in detail. We will developed our ideas in terms of Chebyshev series, but the workflow can be extended to any basis set of choice. The formalism introduced here is (theoretically) more convenient for both performing the fit and for creating the interface with MCTDH, because the creation of the *operator file* is in this case straightforward (*vide infra*).

However, the interconversion between both representations is prone of numerical instabilities due to the limitations floating point arithmetics. Indeed, the coefficients included in the original SOP expansion cover a wide numerical interval, and adding numbers of very different magnitudes (or subtracting numbers of similar magnitudes) can lead to severe loss of precision due to the roundoff error. The above can be improved if one uses a smarter sum algorithm instead of the *naïve* sequential sum, e.g. the Kahan algorithm [235]. Additionally, the condition number of the Chebyshev series coefficients is high, so the physics of the problem is better grasped in the original *ansatz* than in the transformed version, in which the information gets "diluted" on the products of the individual Chebyshev polynomials.

3.1.1.1 An algebraic approach

We will start our analysis with the standard Tucker decomposition of a generic multivariate function V into some basis functions $\{\Phi\}$:

$$V(q_1,\ldots,q_f)=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\prod_{\kappa=1}^{f}\Phi_{j_\kappa}^{(\kappa)}(q_\kappa) \tag{3.1}$$

where the $C_{j_1\cdots j_f}$ are some expansion coefficients.

The choice of the basis functions $\{\Phi\}$ may be highly dependent on the problem at hand. In this work we have used orthogonal polynomial series (particularly Chebyshev polynomials) because (i) every single polynomial can be represented as a linear combination of orthogonal polynomials (ii) orthogonal polynomials are convenient for curve fitting because they can reduce the multicolinearity of the variances of the parameters in the fit[2]. In most of the cases, this choice is adequate. According to our reasoning, we can

---

[2] There might be significant drawbacks in the use of polynomials of high order, like Runge's phenomenon in the edges of the fitting interval.

transform Equation 3.1 as follows:

$$V(q_1,\ldots,q_f)=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\prod_{\kappa=1}^{f}\sum_{\nu_\kappa=1}^{t_k}B^{(\kappa)}_{\nu_\kappa j_\kappa}T_{\nu_\kappa}(q_\kappa) \tag{3.2}$$

where the we have introduced the series of Chebyshev polynomials of first kind $T_{\nu_\kappa}(q_\kappa)$ of order t_κ and coefficients $B^{(\kappa)}_{\nu_\kappa j_\kappa}$ as basis functions:

$$\Phi_{j_\kappa}(q_\kappa)=\sum_{\nu_\kappa=1}^{t_k}B^{(\kappa)}_{\nu_\kappa j_\kappa}T_{\nu_\kappa}(q_\kappa) \tag{3.3}$$

Equation 3.2 can be further simplified if we make use of the relationship between product of sums (POS) and sums of products (SOP):

$$V(q_1,\ldots,q_f)=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\sum_{\nu_1=1}^{t_1}\cdots\sum_{\nu_f=1}^{t_f}\prod_{\kappa=1}^{f}B^{(\kappa)}_{\nu_\kappa j_\kappa}T_{\nu_\kappa}(q_\kappa) \tag{3.4}$$

Then we can apply the associative property of multiplication to the product in Equation 3.4 to write it in a more convenient way:

$$V(q_1,\ldots,q_f)=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\sum_{\nu_1=1}^{t_1}\cdots\sum_{\nu_f=1}^{t_f}\prod_{\kappa=1}^{f}B^{(\kappa)}_{\nu_\kappa j_\kappa}\prod_{\kappa=1}^{f}T_{\nu_\kappa}(q_\kappa) \tag{3.5}$$

By commutativity and associativity of the sums, we can rewrite Equation 3.5 as:

$$V(q_1,\ldots,q_f)=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}\sum_{\nu_1=1}^{t_1}\cdots\sum_{\nu_f=1}^{t_f}C_{j_1\cdots j_f}\cdot\prod_{\kappa=1}^{f}B^{(\kappa)}_{\nu_\kappa j_\kappa}\prod_{\kappa=1}^{f}T_{\nu_\kappa}(q_\kappa) \tag{3.6}$$

Or more conveniently:

$$V(q_1,\ldots,q_f)=\sum_{\nu_1=1}^{t_1}\cdots\sum_{\nu_f=1}^{t_f}\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\cdot\prod_{\kappa=1}^{f}B^{(\kappa)}_{\nu_\kappa j_\kappa}\prod_{\kappa=1}^{f}T_{\nu_\kappa}(q_\kappa) \tag{3.7}$$

Now, if we take a closer look to the inner sums and products in Equation 3.7, we find out that it has a very familiar structure, in fact the expression

$$A_{\nu_1\cdots\nu_f}=\sum_{j_1=1}^{m_1}\cdots\sum_{j_f=1}^{m_f}C_{j_1\cdots j_f}\prod_{\kappa=1}^{f}B^{(\kappa)}_{\nu_\kappa j_\kappa} \tag{3.8}$$

is simply the Tucker decomposition of a new tensor $A_{\nu_1\cdots\nu_f}$ (not to be confused with the MCTDH A- tensor, Section 2.3.2.1) into the same initial core tensor $C_{j_1\cdots j_f}$ and factor matrices $B^{(\kappa)}_{\nu_\kappa j_\kappa}$! Introducing this new quantity we arrive to our desired simplified expression:

$$V(q_1, \ldots, q_f) = \sum_{\nu_1=1}^{t_1} \cdots \sum_{\nu_f=1}^{t_f} A_{\nu_1 \cdots \nu_f} \prod_{\kappa=1}^{f} T_{\nu_\kappa}(q_\kappa) \tag{3.9}$$

3.1.1.2 A tensor approach

A completely analogous result to the one in the previous section can be found if we express our equation in tensor form or matrix form. In this case, we could write Equation 3.1 in tensor form as:

$$\mathcal{V} = \mathcal{C} \times_1 \boldsymbol{\Phi^{(1)}} \cdots \times_f \boldsymbol{\Phi^{(f)}} \tag{3.10}$$

where we have introduced the factor matrices $\boldsymbol{\Phi^{(\kappa)}}$ with elements $\Phi^{(\kappa)}_{j_\kappa}$. The previous expression can be rewritten in matrix form as:

$$\mathbf{V} = \boldsymbol{\Omega}\mathbf{C} \tag{3.11}$$

where:

$$\begin{aligned} \mathbf{V} &= \mathrm{vec}(\mathcal{V}) \\ \mathbf{C} &= \mathrm{vec}(\mathcal{C}) \\ \boldsymbol{\Omega} &= \bigotimes_{\kappa=1}^{f} \boldsymbol{\Phi^{(\kappa)}} \end{aligned} \tag{3.12}$$

The expressions 3.12 define the important relationships that allow for vectorizing the tensor n-mode product representation of the problem at hand (3.10).

With the collected information we can proceed to transform Equation 3.10. We start by defining the following vectors:

$$\mathbf{T}^{(\kappa)} = (T_0(q_\kappa), \cdots, T_{t_\kappa}(q_\kappa)) \tag{3.13}$$

Then, from Equation 3.3 we have:

$$\boldsymbol{\Phi^{(\kappa)}} = \mathbf{T}^{(\kappa)^\dagger} \mathbf{B}^{(\kappa)} \tag{3.14}$$

where we have introduced the matrices $\mathbf{B}^{(\kappa)}$ of elements $B_{\nu_\kappa j_\kappa}$. From this if follows that:

$$\mathcal{V} = \mathcal{C} \times_1 (\mathbf{T^{(1)}}^\dagger \mathbf{B^{(1)}}) \cdots \times_f (\mathbf{T^{(f)}}^\dagger \mathbf{B^{(f)}}) \tag{3.15}$$

Using the associativity property of the tensor n-mode product (see Section 2.4.1):

$$\mathcal{V} = \mathcal{C} \times_1 \mathbf{B^{(1)}} \times_1 \mathbf{T^{(1)}}^\dagger \cdots \times_f \mathbf{B^{(f)}} \times_f \mathbf{T^{(f)}}^\dagger \tag{3.16}$$

If we now employ the commutativity property of the tensor n-mode product on different modes:

$$\mathcal{V} = \mathcal{C} \times_1 \mathbf{B^{(1)}} \cdots \times_f \mathbf{B^{(f)}} \times_1 \mathbf{T^{(1)}}^{\dagger} \cdots \times_f \mathbf{T^{(f)}}^{\dagger} \tag{3.17}$$

Or simply:

$$\mathcal{V} = \mathcal{A} \times_1 \mathbf{T^{(1)}}^{\dagger} \cdots \times_f \mathbf{T^{(f)}}^{\dagger} \tag{3.18}$$

where we have set:

$$\mathcal{A} = \mathcal{C} \times_1 \mathbf{B^{(1)}} \cdots \times_f \mathbf{B^{(f)}} \tag{3.19}$$

We can see that Equation 3.18 is completely analogous to Equation 3.9. If we would like to express the former in matrix form we could make use of expressions 3.12 to finally get:

$$\mathbf{V} = \mathbf{\Omega A} \tag{3.20}$$

with

$$\mathbf{\Omega} = \bigotimes_{\kappa=1}^{f} \mathbf{T}^{(\kappa)^{\dagger}} \tag{3.21}$$

Equation 3.20 is of especial importance because the addends of this matrix product can be directly written into an MCTDH PES section of the *operator file*, needed to perform nuclear quantum dynamics calculations. Specifically, the section would look like this:

```
...
1.02404335833489E-02 | cheb0 | cheb1 | cheb2
-3.81637065306385E-04 | cheb0 | cheb1 | cheb3
2.95338724217441E-01 | cheb0 | cheb2 | cheb0
5.09858364484404E-02 | cheb0 | cheb2 | cheb1
1.25544332850752E-03 | cheb0 | cheb2 | cheb2
...
```

where the first column represent the elements of the tensor $\mathcal{A}$ and the labels **cheb**(ν_κ) represent[3] the element $T_{\nu_\kappa}(q_\kappa)$.

3.2 Publications

The next pages contain the original text of the articles *Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF): Automatic Generation of Sum-of-Products Form Potential Energy Surfaces for Quantum Dynamical Calculations* [7] and *Low-Rank Sum-of-Products*

---

[3] This labels have been introduced in the Heidelberg MCTDH software package source code by the authors

Finite-Basis-Representation (SOP-FBR) of Potential Energy Surfaces [11]. In both cases the formating imposed by the publishers is fully respected. As it was mentioned before, the first of them is already published in the journal Frontiers in Chemistry, whereas the other one is in the reviewing process at The Journal of Chemical Physics. The latter article is an invited contribution to the Special Topic "Quantum Dynamics with ab Initio Potentials ".

frontiers
in Chemistry

ORIGINAL RESEARCH
published: 14 August 2019
doi: 10.3389/fchem.2019.00576

Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF): Automatic Generation of Sum-of-Products Form Potential Energy Surfaces for Quantum Dynamical Calculations

Ramón L. Panadés-Barrueta[1], *Emilio Martínez-Núñez*[2] *and Daniel Peláez*[1]*

[1] Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM), Université de Lille, Villeneuve-d'Ascq, France,
[2] Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

OPEN ACCESS

Edited by:
Nino Russo,
University of Calabria, Italy

Reviewed by:
Hans-Dieter Meyer,
Heidelberg University, Germany
Riccardo Conte,
University of Milan, Italy

***Correspondence:**
Daniel Peláez
daniel.pelaez-ruiz@univ-lille.fr

Specialty section:
This article was submitted to Theoretical and Computational Chemistry,
a section of the journal Frontiers in Chemistry

Received: *04 June 2019*
Accepted: *30 July 2019*
Published: *14 August 2019*

Citation:
Panadés-Barrueta RL, Martínez-Núñez E and Peláez D (2019) Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF): Automatic Generation of Sum-of-Products Form Potential Energy Surfaces for Quantum Dynamical Calculations. Front. Chem. 7:576. doi: 10.3389/fchem.2019.00576

We present Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF), an automated methodology for the generation of global potential energy surfaces (PES), molecular properties surfaces, e.g., dipole, polarizabilities, etc. using a single random geometry as input. The SRP-MGPF workflow integrates: (i) a fully automated procedure for the global topographical characterization of a (intermolecular) PES based on the Transition State Search Using Chemical Dynamical Simulations (TSSCDS) family of methods;i (ii) the global optimization of the parameters of a semiempirical Hamiltonian in order to reproduce a given level of electronic structure theory; and (iii) a tensor decomposition algorithm which turns the resulting SRP-PES into sum of products (Tucker) form with the Multigrid POTFIT algorithm. The latter is necessary for quantum dynamical studies within the Multiconfiguration Time-Dependent Hartree (MCTDH) quantum dynamics method. To demonstrate our approach, we have applied our methodology to the *cis-trans* isomerization reaction in HONO in full dimensionality (6D). The resulting SRP-PES has been validated through the computation of classical on-the-fly dynamical calculations as well as calculations of the lowest vibrational eigenstates of HONO as well as high-energy wavepacket propagations.

Keywords: PES, sums-of-products, tensor-decomposition, quantum dynamics, reparametrized semiempirical, TSSCDS, global optimization

1. INTRODUCTION

A detailed knowledge of the topography of a Potential Energy Surface (PES) is a highly desirable prerequisite for the simulation of any dynamical process. Topography on its own, however, does not fully determine the behavior of a system and dynamics calculations become mandatory (Tuckerman et al., 2002; Peláez et al., 2014). Furthermore, for an accurate theoretical description of molecular processes (spectroscopy, reactivity), one should, if possible, resort to nuclear quantum dynamics calculations (Gatti, 2014). In the specific case of vibrational problems, powerful methods based on the resolution of the time-independent Schrödinger equation exist such as vibrational self-consistent field/vibrational configuration interaction (VSCF/VCI) (Rauhut, 2007; Neff and Rauhut, 2009), vibrational second-order perturbation theory (VPT2) (Barone, 2005) and vibrational coupled-cluster theory (Christiansen, 2004). For an extensive and recent review of some of them, the reader is referred to a recent publication (Puzzarini et al., 2019). However, owing to our interest in describing chemical processes, we shall turn our attention toward methods able to describe wave packet propagations. In this context, within the last few years, we have experienced a boost in dynamical methodologies capable of describing the dynamics of molecular systems up to medium-large size, ranging from semiclassical (Levine et al., 2008; Shalashilin, 2010) to fully quantal (Gatti, 2014). With respect to the latter, by far, the most popular approaches nowadays are those based on, or related to, the grid-based Multiconfiguration Time-Dependent Hartree (MCTDH) algorithm (Beck et al., 2000). In MCTDH, a molecular wavefunction (WF) is expanded in a basis of time-dependent nuclear orbitals. Taken MCTDH as reference, two powerful multiconfigurational methods exist. On the one hand, the partially grid-based G-MCTDH method in which some of the time-dependent basis functions are substituted by (typically frozen) Gaussians functions (G) (Burghardt et al., 2008), and the Variational Multiconfigurational Gaussian (vMCG) method (Richings et al., 2015) (and its direct-dynamics (DD) extension) which are grid-free and use Gaussian functions only. For the sake of completeness, one should mention the recent and promising direct-dynamics approach of MCTDH by Richings and Habershon (2018).

It should be evident that the quality of the results of any dynamical calculation is limited by the accuracy and efficiency of the underlying electronic structure method used to represent the PES, either globally (as in grid-based methods) or locally (on-the-fly approaches). When expressed *globally* on a grid, formally as a multidimensional tensor, the limitation lies on the number of dynamical degrees of freedom and the possibility of fitting the PES to an appropriate functional form. In the case of on-the-fly methods, on the other hand, the number of degrees of freedom (DOF) it is not the main limiting factor but the *number of electrons*, in other words, the level of theory and its performance in the form of electronic structure software *calls* (energies, gradients, Hessians) at each time-step. This fact constrains on-the-fly approaches to modest levels of theory.

Obtaining a fit for a high-dimensional PES is a complex and tedious task. Whatever the approach, any fitting procedure requires a more or less large set of reference values (molecular energies and/or gradients and, possibly, properties such as dipoles) which will constitute the data to which an algorithm will try to fit a given function. *Ad hoc* analytical functions are usually added to the resulting fit in order to ensure a correct physical behavior, for instance in the asymptotic regions, or to guarantee a correct periodicity of the potential as in the case of rotors. Focusing on the fitting methods typically used in combination with nuclear quantum dynamical approaches, many techniques have been proposed. To name but a few, popular methods include the permutationally invariant polynomials (Braams and Bowman, 2009), the interpolating moving least-squares (Dawes et al., 2007), the triatomics-in-molecules approximation (Sanz-Sanz et al., 2013), Shephard interpolation schemes (Frankcombe and Collins, 2011). Moreover, for more than a decade now, Neural Network (NN) approaches have (re)gained preeminence being triggered by the pioneering work of Manzhos and Carrington (2006) and, very recently, their application to MCTDH by Pradhan and Brown (2017). In this line, Jiang and Guo have gone a step further and have developed a NN approach with implicit nuclear permutational symmetry (Jiang and Guo, 2014). For the sake of completeness, one should mention the works of Rauhut (2004) and Sparta et al. (2010) in which PESs for vibrational calculations are generated in an automated and adaptive fashion. Powerful and accurate as these methods are, a high degree of expertise is still required to master and to apply these techniques, particularly for medium-large systems ($\geq$6D), thus preventing them from a wider-spread use. Furthermore, in studies where external fields (e.g., a laser) are needed, surfaces of molecular properties are also required and, as a consequence, extra fits are necessary.

In this work, we present Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF), a method which provides a well-balanced solution to the aforementioned issues. SRP-MGPF is able to generate a chemically-accurate PES as well as the same-level-of-theory molecular properties surfaces, starting from a single input geometry and requiring minimal intervention of the user. In this sense, we can safely affirm that the procedure is *quasi* black-box in nature. SRP-MGPF relies on three main steps: (i) generation of a set of reference geometries (energies and properties); (ii) reparametrization of a semiempirical Hamiltonian (Specific Reaction Parameter Hamiltonian, SRP) based on the previous information; and (iii) tensor-decomposing the SRP with MGPF. We shall focus on the *standard* MCTDH method for which a global PES needs to be fitted into some kind of functional form and, typically, refitted to a grid. Furthermore, our results can also be directly applied to any on-the-fly methodology. It should be highlighted at this point that reparametrized semiempirical Hamiltonians have been typically used in direct dynamics studies as well as in kinetic studies (Rossi and Truhlar, 1995; Troya, 2005; Rodríguez-Fernández et al., 2017). Moreover, semiempirical Hamiltonians have been successfully used in describing dynamics on electronically excited states (Toniolo et al., 2003; Silva-Junior and Thiel, 2010). It should be stressed that SRP methods qualify as quantum chemical ones. As such an SRP does not include, necessarily, any fitting functions. Hence, the SRP parameters obtained through our fitting process will define a level of electronic structure close to a high-level reference one.

In our approach, as generator set for the reference fitting points, we employ the so-called Reaction Network (RXN) (Martínez-Núñez, 2015b), i.e., the *complete* set of stationary points (minima, transition states,…) of a PES. The RXN captures the main topographical (even topological) features of the target PES and thus constitutes a sensible choice for the reference set. Characterization of the topography of a PES is, however, not an evident task. To this end, we make use of the recently developed Transition State Search Using Chemical Dynamics Simulations

(TSSCDS) (Martínez-Núñez, 2015a,b) method which relies on the efficient sampling of configuration space combined with a graph-theory based identification of transition state (TS) structures, which are finally optimized and the corresponding Minimum Energy Paths obtained with standard methods. The TSSCDS approach has been recently extended to specifically study van der Waals complexes (vdW) or, more generally, non-covalently bound systems (vdW-TSSCDS) (Kopec et al., 2019).

A set of *optimal* semiempirical Hamiltonian parameters is then obtained by global minimization of the Root-Mean Square Error (RMSE) between a set of reference *ab initio* energies, for instance, (on the RXN-derived geometries) and the corresponding SRP ones. The SRP approach to PESs presents interesting features that make it very appealing when compared to formally higher-level methods (Density Functional Theory, DFT, or *ab initio*). First, SRPs are fast-computing parametrized electronic structure methods, some of the integrals are neglected while the remaining are parametrized to reproduce high-level results. As such, they typically exhibit a *correct* physical behavior. Second, in contrast to other *fitting* procedures (for instance based on any kind of polynomial expansions or neural networks), SRPs exhibit a correct behavior outside the fitting boundaries, *if* the SRP parameters remain somewhat physical (*small* variation with respect to their reference values). Third, by varying the SRP parameters we can simultaneously fit both energies and the molecular properties accessible to the semiempirical software. It should be highlighted that in the usual approach energies and properties (e.g., dipole) are computed at a set of reference geometries and then need to be independently fitted to either potential energy surfaces or property surfaces (x-dipole, y-dipole, etc.). In contrast, in our method a single optimization process suffices to yield a simultaneous fit of all properties simultaneously, provided that information on the desired properties is included in the reference data. Last, but not least, the number of parameters is *independent* of the number of atoms. They only depend on the number of different atoms (and possibly on their chemical function) and, as such, it is in a sense not affected by the curse of dimensionality. In our specific approach, we have used as base model chemistry the Parametric Method 7 (PM7) method as implemented in the OpenMOPAC software package (Stewart, 2016). This choice is justified by the quality of the obtained results as well as its efficiency in terms of computational time (PM7 is orders of magnitude faster than *ab initio* methods) (Stewart, 2013).

The final step, specific for grid-based methods, is the tensor-decomposition of the SRP-PES into an appropriate form. To this end, we utilize the Multigrid POTFIT (MGPF) algorithm (Peláez and Meyer, 2013), succinctly described in section 2.3. MGPF has been successfully applied to the computation of vibrational eigenstates (Peláez et al., 2014), infrared (IR) spectra (Peláez and Meyer, 2017), and electron dynamics including continuum (Haller et al., 2019) With SRP-MGPF, owing to the extreme efficiency of the semiempirical calculations, we can directly generate the SRP-PES on a grid.

This manuscript is structured as follows. In section 2 we provide a succinct introduction to the methods employed in our workflow. In section 3, which presents the application of our novel methodology to the HONO molecule in full-dimensionality, we carefully discuss all specific aspects related to the actual calculations. Section 4 concludes the paper and gives some hints on future developments and possible applications of the method.

2. THEORY AND COMPUTATIONAL DETAILS

Our automated methodology for computing a global PES consists of three steps: (1) automatic and global determination of stationary points (minima and transition states), as well as the corresponding Intrinsic Reaction Coordinate paths (IRCs), the so-called Reaction Network (RXN); (2) reparametrization of a semiempirical Hamiltonian (SRP) to reproduce a desired level of electronic structure theory (e.g., *ab initio*) using the RXN and neighboring points; and (3) tensor-decomposition of the SRP Hamiltonian with the MGPF algorithm. It should be noted that after stage (2), we already have a global PES which can be used in conjunction with any type of *on-the-fly* dynamics scheme. We shall describe in the following each of the above mentioned stages. First of all, we shall discuss our specific procedure for the reparametrization of semiempirical Hamiltonians. Then, we shall present our way of generating a set of reference points based on the RXN obtained using the (vdW-)TSSCDS method (Martínez-Núñez, 2015a,b). Subsequently, we shall discuss how we integrate this information in combination with the NLOpt (Johnson, 2011) library and the openMOPAC software (Stewart, 2016) to produce an *optimal* set of SRP parameters. The resulting SRP-PES is then *interfaced* with MCTDH through the Multigrid POTFIT program (Peláez and Meyer, 2013) thus generating a SRP-MGPF PES on the grid and in sums-of-product (SOP) form.

Finally, it should be highlighted that, for the graphical representations, we have made extensive use of the SciPy scientific tools by Jones et al. (2001).

2.1. Global Optimization of Semi-empirical Hamiltonians Parameters

Semiempirical potentials can be seen as parametrized Hartree-Fock methods in which some of the electronic integrals are either neglected or replaced by parameters obtained as fitting constants using large sets of reference data (high-level *ab initio* calculations and/or experimental data) (Stewart, 2013; Thiel, 2014). In this sense, semiempirical methods lie somewhere between force fields and *ab initio* methods (Stewart, 2013). Owing to the lower amount of integral calculations, semiempirical methods are orders of magnitude faster than *ab initio* methods and, hence, they are routinely used in the study of large systems (Christensen et al., 2016). In addition to this, with a suitable configuration interaction formalism, semiempirical methods can also be used for the study of excited states (Toniolo et al., 2003; Silva-Junior and Thiel, 2010). A milestone in the usage of semiempiricals was achieved by Rossi and Truhlar (1995) who introduced the idea of reparametrizing a semiempirical Hamiltonian in order to reproduce a given high-level *ab initio* level of theory for a *specific* chemical reaction (or family thereof), hence the name of

Specific Reaction Parameter (SRP) Hamiltonians. Since then, this technique has been successfully applied to the study of chemical reactions of large-dimensional systems using classical dynamics (Layfield et al., 2008) as well as to kinetic studies (Rodríguez-Fernández et al., 2017). In the present work, we go a step further and will use the SRP approach for the generation of a PES suitable for quantum dynamical studies. To this end, we used the publicly available non-linear global optimization library NLOpt (Johnson, 2011) to reparametrize the PM7 semiempirical model (Stewart, 2013) as implemented in openMOPAC (Stewart, 2016). The choice of PM7 responds not only to its proven accuracy but also to the fact that it includes diatomic parameters in addition to the standard atomic ones, thus providing extra flexibility to the optimization process (Stewart, 2013). Hereafter, we shall refer to the set of SRP parameters as $\{\zeta_i\}_{i=1}^{D}$, being D the total number of parameters. It is important to notice that the latter depends on the number of *atom types* and *not* on the dimensionality of the system. It should be stressed that we are dealing with a *fitting function* which has an implicit physical character (HF-like) and, as such, it is expected to yield a *global* qualitatively-correct behavior and to require less fitting points than other traditional fitting approaches.

The problem that concerns us is thus the global optimization of a deterministic non-linear objective function $\chi(\boldsymbol{\zeta})$: $\mathbb{R}^D \rightarrow \mathbb{R}$, Equation (1), with a bounded parameter space ($\zeta_i \in [\zeta_i^{min}, \zeta_i^{max}], i = 1, \ldots, D$). In our specific case, we do not make use of the derivatives of this target function since: (i) the analytical expressions are unavailable; (ii) their numerical determination would be expensive and, more importantly, complicated due to the highly-corrugated character of the RMSE landscape (see **Figure 1**). We shall consider then a *derivative-free optimization* algorithm (Rios and Sahinidis, 2013). As general expression of the objective function (χ) we have considered a *rms-like* function (see Equation 1) composed by two terms: (i) a first one accounting for the error in the energies and (ii) a the second one corresponding to the error in the harmonic frequencies of the stationary points of the PES, with respect to our reference calculations. We have observed that the inclusion of the latter helps to preserve the correct topography of the PES, for instance the first order saddle point character of transition states.

$$\chi_0(\boldsymbol{\zeta}) = \sqrt{\sum_{i=1}^{n} \frac{\omega_E(E_i^{ab}) \cdot [E_i^{ab} - E_i^{srp}(\boldsymbol{\zeta})]^2}{n} + \sum_{j=1}^{m} \frac{\omega_F(\Delta F_j) \cdot [F_j^{ab} - F_j^{srp}(\boldsymbol{\zeta})]^2}{m}} \tag{1}$$

where $\boldsymbol{\zeta}$ is a vector containing the semiempirical parameters and n, m represent the number of (relative) energy data points ($E^{ab/srp}$) and harmonic frequencies ($F^{ab/srp}$), respectively, the labels referring to *ab initio* (*ab*) and semiempirical (*srp*) data. The weighting functions $\omega_E(E_i^{ab})$ and $\omega_F(\Delta F_j)$ (with $\Delta F_j = F_j^{ab} - F_j^{srp}$) have been defined as exponential step functions:

$$f(x) = \begin{cases} 1 & x \leq \alpha \\ e^{\beta(x-\alpha)} & x > \alpha \end{cases} \tag{2}$$

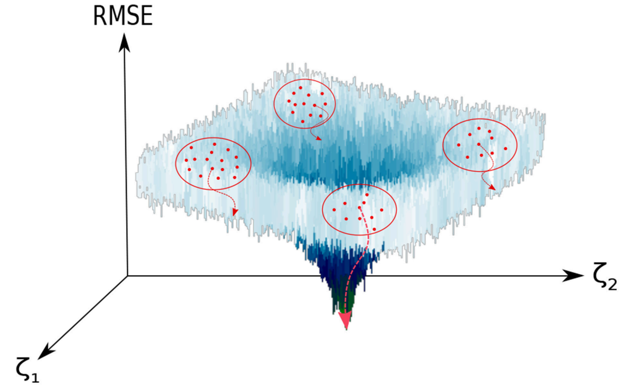

FIGURE 1 | Graphical representation of the optimization process of the set of SRP parameters ($\{\boldsymbol{\zeta}\}$). The vertical axis displays the RMSE between our reference data and our target function (see Equation 3), which in the figure depends just on two parameters (ζ_1, ζ_2). Non-overlapping clusters (red dots enclosed in a red circle) of walkers (red dots) are generated. In each cluster, the *optimal* solution is locally minimized (red dotted curved arrows) and compared to the rest of solutions. For a large enough number of clusters, convergence to the global minimum is guaranteed. In this representation, we have used a modified Ackley function (Ackley, 1987).

where α, β are parameters adjusted *a priori* and x corresponds to the selected argument ($E_i^{ab}, \Delta F_j$). However, in practice, we have obtained satisfactory results with a much simpler expression:

$$\chi_1(\boldsymbol{\zeta}) = \sqrt{\sum_{i=1}^{n+m} \frac{\omega_G(G_i^{ab}) \cdot [G_i^{ab} - G_i^{srp}(\boldsymbol{\zeta})]^2}{n+m}} \tag{3}$$

where $G_i = E_i || F_i$ are the components of a vector constructed by concatenating the vectors of energies and harmonic frequencies, respectively. As strategy, we have performed a global optimization step followed by local optimizations in order to refine the results. For the former, we used the Multi-Level Single-Linkage (MLSL) algorithm (Kan and Timmer, 1987) and for the latter we used the Bound Optimization BY Quadratic Approximation (BOBYQA) (Powell, 2009).

2.2. Automated Generation of the Set of Reference Points

In the following, we shall describe our automated methodology for the generation of a set of fitting points for the reparametrization of a semiempirical Hamiltonian. In brief, we propose the use of the *whole* set of stationary points of a given PES, the so-called RXN (Martínez-Núñez, 2015a,b; Kopec et al., 2019), as initial set from which neighboring geometries spanning the region of configuration space of interest will be generated. The main advantage of our method is that starting from a *single* initial input geometry, a *global* Potential Energy Surface is generated.

We propose as first step the determination of the ensemble of stationary points (RXN) on a given PES which will be used as seed for the subsequent generation of the remaining

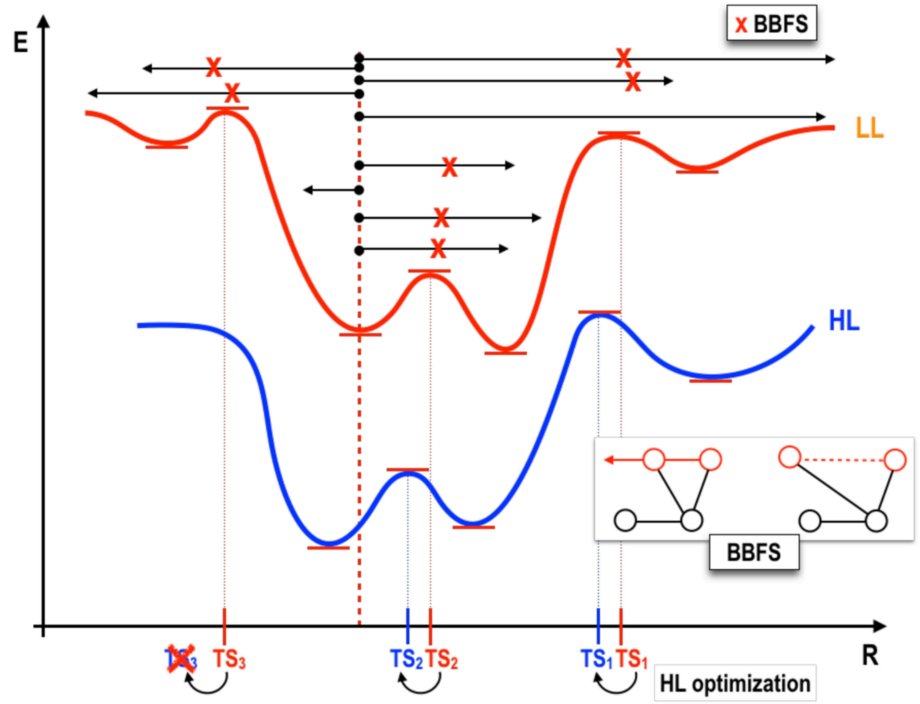

FIGURE 2 | One-dimensional representation of the TSSCDS procedure. A low level (LL) PES (upper energy curve, in red) is sampled starting from a given minimum (geometry indicated by a red dotted line). Classical random trajectories (black arrows) in combination with a graph theory based method (Bond Breaking/Formation Search, BBFS Martínez-Núñez, 2015a,b) lead to the determination of TS candidate structures (marked as x in red bold font), compatible with the total energy of the trajectories, from which LL optimizations are started. Subsequent optimization at the desired high-level (HL) are performed using the LL TS as guess structures.

fitting points. Indeed, the stationary points correspond to the molecular configurations which carry the most relevant topographical information of a given PES and, as such, make ideal candidates for fitting purposes. Finding stationary points, however, is a very tedious task which heavily relies on large amounts of chemical intuition. Fortunately, a family of methods for the automated determination of the RXN has been recently proposed, the so-called Transition State Search Using Chemical Dynamics (TSSCDS) (Martínez-Núñez, 2015a,b) as well as its generalization, vdW-TSSCDS (Kopec et al., 2019). The former is optimal for the study of unimolecular processes whereas the latter has been specifically designed to study non-covalently bound systems. The workflow in both cases is analogous (see **Figure 2**) and the difference lies in the way transition states (TSs) are characterized. Starting from an initial random geometry (or small set thereof), a large number of high-energy classical trajectories is run using a low-level (LL) of electronic structure theory (semiempirical in our case, other methods are also possible) to compute the forces. The geometries along these trajectories are analyzed by a graph-theory based algorithm (Bond Breaking/Formation Search, BBFS Martínez-Núñez, 2015a,b; Kopec et al., 2019) which detects conformations in which bonds are broken and/or formed. It should be highlighted that this step is precisely what determines the difference between TSSCDS and vdW-TSSCDS. In the former, a square connectivity matrix based on covalent distances is defined, whereas in the latter this matrix takes block-diagonal form and includes both covalent and non-covalent (van der Waals) distances, thus allowing for the determination of non-covalent saddle points. The so-determined structures, candidates to TSs, are optimized at the LL and subsequently reoptimized at an appropriate higher level of theory, say, *ab initio* or DFT. Obviously, this process can be continued by further refinements. From this set of final high-level TSs, IRC calculations connecting minima are performed. And, as a result of this, the so-called Reaction Network (RXN) is obtained, that is, all stationary structures together with their connectivities compatible with a given total energy (that of the initial classical trajectories). For further details on the method, the interested reader is referred to the original publications (Martínez-Núñez, 2015a,b; Kopec et al., 2019). As indicated, the RXN will serve us as initial set from which the full set of fitting points will be generated. The total number of stationary points (N_{RXN}) is:

$$N_{RXN} = n_{min} + n_{TS} + n_{asymp} + \ldots, \qquad (4)$$

where n_X, (with X=min, TS, asymp,. . .) is the number of minima, transition states (TS), asymptotic products, respectively. This initial set will be extended by systematically adding a set of *neighboring* geometries. This can be achieved in different ways. In our case, we have chosen to distort each of the N_{RXN} points following an n-body type of scheme inspired by a previous work (Pradhan and Brown, 2017). The novelty of our procedure lies in the fact that we observe convergence in the RMSE at each order of the expansion. As it will be clear later, this convergence provides us with an efficient error control and allows to determine a minimal number of fitting points necessary to achieve a given RMSE. The total number of fitting points (N_{ref}) can be calculated as:

$$N_{ref} = N_{RXN} \cdot \Big[\sum_{i \in 1D}^{f} N_i^{(1D)} + \sum_{i \in 2D}^{f} N_i^{(2D)} + \dots \Big] + rnd(fD) + \sum_{i}^{n_{TS}} N_i^{IRC} + \sum_{i}^{n_{asymp}} N^{(asymp)} + \dots \qquad (5)$$

where f is the number of degrees of freedom of the molecular system, N is the number of generated reference geometries of a given type, for instance, N^{nD} are grid points from a n-dimensional (D) grid and N_i^{IRC} are the IRC points stemming from TS_i, $rnd(fD)$ are random geometries in the full-D configuration space, n is the number of stationary points of a kind. Considering, for instance, a normal mode or internal coordinate local representation, 1D would refer to displacements along each mode/coordinate (leaving the remaining coordinates fixed at their equilibrium values) and nD refers to grids of points generated through simultaneous displacement along n modes/coordinates, leaving the remaining fixed as before.

Our goal is now to determine the minimum number of fitting points leading to the smallest possible RMSE (defined as the difference between reference PES and SRP-PES), or, in other words, the *optimal* set of SRP parameters ($\{\zeta_{opt}\}$). It should be emphasized that we are dealing with moderate-size configuration spaces, in our specific case HONO (6D), the parameter space is 34-dimensional. Hence, in order to systematically search for the global minimum in SRP parameter space ($\{\zeta\}$), we increase the number of reference points in a controlled way according to the following prescription. Starting with the PM7 parameters ($\{\zeta_{PM7}\}$) as initial guess, the RMSE($\boldsymbol{\zeta}$) landscape is explored in a first stage using a small number of ab initio reference data and a big number of iterations (typically of the order of 10^5) of the non-linear optimization algorithm (MLSL in our case). This allows to locate the most-likely candidate parameter set to global minimum. The latter is used as a guess in subsequent local optimization stages (BOBYQA). At each of these, extended sets of points are generated in the form of *nD* distortions. At each level (1D, 2D, etc.) and for each set, we carry out local optimizations, compare the resulting RMSEs and take as *optimal* the number of points (set) that leads to a satisfactory value of RMSE, in the form of convergence, thus guaranteeing the condition of minimum number of points.

2.3. Generation of the SRP-MGPF Potential Energy Surface

As any other grid-based method, MCTDH quantum dynamics relies on a discretization of the configuration space known as *primitive* grid (Kosloff, 1988). In an f-dimensional molecular system (typically f = 3N-6, with N being the number of atoms), a set of $i_\kappa = 1, \dots, N_\kappa$ grids points is defined for the κ-th DOF with $\kappa = 1, \dots, f$. In other words, a given grid point $I \equiv (i_1, \dots, i_f)$ has an associated molecular configuration ($Q \equiv (q_i, \dots, q_f)$). The wavefunction in MCTDH is expressed in a two-layer scheme, a first one in terms of time-dependent single-particle basis functions (SPFs, $\{\boldsymbol{\varphi}^{(\kappa)}\}$):

$$\Psi(q_1, \dots, q_f, t) = \sum_{j_1} \dots \sum_{j_f} A_{j_1 \cdots j_f}(t) \prod_{\kappa=1}^{f} \varphi_j^{(\kappa)}(q_\kappa, t) \qquad (6)$$

and a second in which each SPF is, in turn, expressed in a time-independent basis set ($\{\boldsymbol{\chi}^{(\kappa)}(q_\kappa)\}$):

$$\varphi_{j_\kappa}^{(\kappa)}(q_\kappa, t) = \sum_{i_\kappa=1}^{N_\kappa} c_{j_\kappa i_\kappa}^{(\kappa)}(t) \chi_{i_\kappa}^{(\kappa)}(q_\kappa) \qquad (7)$$

the latter, typically, Discrete Variable Representation (DVR) functions (Beck et al., 2000; Light and Carrington, 2000). In this frame, each grid point i_κ (κ-th DOF, $q^{(\kappa)}$) is associated to a localized time-independent basis function ($\chi^{(\kappa)}(q^{(\kappa)})$). Obviously, a minimum number of basis functions, or conversely grid points must exist to achieve the numerical convergence of a given calculation. Such grid representations imply that quantities, particularly the PES, are represented by f-dimensional *tensors*, where f is the number of DOF. If each DOF is represented by 10 grid points, a tensor of 10^f grid points would be necessary to represent the PES. It should be clear at this point that that generation of such a high-dimensional PES tensor directly from electronic structure (i.e., quantum chemistry) codes is, nowadays, a prohibitively-long process.

Apart from diminishing the computational time associated to each quantum chemical calculation, solutions to this issue must imply a reduction in the number of grid points necessary to achieve an accurate grid representation of the PES. This can be achieved in two ways. When considering a more or less localized region of the PES (i.e., centered around a given minimum), local approaches such as the Quartic Force Field representation (QFF) can be used. This is the case when computing vibrational eigenenergies and/or eigenstates (Barone, 2005; Ávila and Carrington, 2009; Neff and Rauhut, 2009). On the other hand, when more global representations are needed (e.g., spectroscopy in multi-well problems, reactivity, etc.) one has to resort to more elaborated forms such as tensor-decomposition algorithms (Kolda and Bader, 2009) or Neural Networks (NN) representations (Manzhos et al., 2006). Two examples of this have been recently proposed for a 6D problem (HONO). With respect to the former, Baranov and Oseledets have used a Tensor-Train tensor-decomposition approach (Baranov and Oseledets, 2015) and Pradhan and Brown have illustrated the use of an exponential NN *ansatz* to represent the same PES (Pradhan and Brown, 2017). In both cases, the number of data-points (i.e., high-level *ab initio* calls) needed to perform the fit was of the order of $\sim 10^4$. Upon an increase of the dimensionality of the problem, this last figure is expected to increase, at least, polynomically, hence preventing the use of these techniques for larger systems.

Our method deals with the aforementioned issues by combining an extremely efficient level of electronic structure, a reparametrized semiempirical Hamiltonian, with an efficient

TABLE 1 | Number and description of the fitting points used in each SRP-optimization stage and the algorithm used in the process.

No. points	Class of points	Type of optimization
53	*core*	Global/Local
367	1D + *core*	Local
546	1D + 2D + *core*	Local
648	1D + 2D + rnd(6D) + *core*	Local
954	1D + 2D + rnd(6D) + LIIC-IRC + *core*	Local
1084	1D + 2D + rnd(6D) + LIIC-IRC + rnd(LIIC) + *core*	Local

Structures have been generated within a set of fixed boundaries defined in ***Table 3****. The initial set of geometries (labeled core) consists on 53 points, namely: MIN1, MIN2, TS1, 1D- and 2D-distorted structures using the latter as reference geometries [26 1D, 14 2D], and 10 6D-randomly distorted (rnd(6D)) geometries. The number of points at each new set is cumulative. It includes nD-distorted geometries (n = 1, 2, 6), LIIC structures and distortions thereof [noted as rnd(LIIC)]. The algorithm for global optimization is MLSL and for local is BOBYQA (see section 2.1). The number of iterations in the global step has been set up to 100,000 and in the local one to 2,000.*

and accurate tensor decomposition scheme, Multigrid POTFIT (MGPF) (Peláez and Meyer, 2013). This tensor decomposition algorithm transforms a multidimensional function (e.g., PES) into Tucker product form (Equation 8) in an *quasi* black-box manner. MGPF, implemented in the MCTDH software package (Worth et al., 2016), avoids running over the full (primitive) MCTDH grid and, instead, uses a series of coarser (nested) grids using a number of PES data-points comparable to the aforementioned methods. However, the big difference is that in our case we shall perform SRP calls, in other words, our *ab initio* method will have the computational cost of a semiempirical one. In fact, as shown by our results (see **Table 1** in section 3.1), we need no more than hundreds of high-level *electronic structure* calls in comparison to the tenths of thousands points required by previous methods. This, obviously, leads to a (small) error inherent to the SRP approximation, but in contrast permits the extension of our approach toward higher-dimensional systems with a little more effort. In the following lines, we shall describe the actual MGPF approach that we have used.

In MGPF, we use a sum-of-products or Tucker expansion for the PES:

$$V = \sum_{j_1,\ldots,j_f}^{[m_1,\ldots,m_f]} C_{j_1,\ldots,j_f} \prod_{\kappa=1}^{f} v_j^{(\kappa)} \tag{8}$$

which, in tensor notation, can be written as: Kolda and Bader (2009)

$$\boldsymbol{V} = \mathcal{C} \times_1 \boldsymbol{v}^{(1)^T} \times_2 \boldsymbol{v}^{(2)^T} \cdots \times_f \boldsymbol{v}^{(f)^T} \tag{9}$$

There $\mathcal{C}$ is the so-called *core* tensor and $\boldsymbol{v}^{(\kappa)}$ are the expansion basis sets for the κ-th DOF. The reader is referred to the original article for a full description of the method and its capabilities (Peláez and Meyer, 2013). More specifically, our current application uses a bottom-up approach to MGPF (Peláez and Meyer, in preparation). The MGPF basis sets ($\{\tilde{\boldsymbol{v}}^{(\kappa)}\}$) can be expressed as:

$$\tilde{\boldsymbol{v}}^{(\kappa)} = \boldsymbol{\rho}^{(\kappa)'} \boldsymbol{\rho}^{(\kappa)^{-1}} \boldsymbol{v}^{(\kappa)} \,. \tag{10}$$

There we have introduced potential density matrices of the form: Peláez and Meyer (2013)

$$\rho_{kk'}^{(\kappa)} := \sum_{I^\kappa} V_{I_k^\kappa} \; V_{I_{k'}^\kappa} \qquad \kappa = 1, \ldots, f \,. \tag{11}$$

where the first index (k) runs along the primitive grid in $\boldsymbol{\rho}^{(\kappa)'}$ and along the coarse one in $\boldsymbol{\rho}^{(\kappa)}$. The transpose of these basis sets reads then:

$$\tilde{\boldsymbol{v}}^{(\kappa)^T} = \boldsymbol{v}^{(\kappa)^T} (\boldsymbol{\rho}^{(\kappa)'} \boldsymbol{\rho}^{(\kappa)-1})^T \tag{12}$$

Substituting in the MGPF expansion V^{MGPF} of the form Equation (9), we unitarily transform both the MGPF basis set ($\tilde{\boldsymbol{v}}$) and the MGPF *core* tensor ($\mathcal{C}$) using the complete basis $\boldsymbol{v}$: Peláez and Meyer (in preparation)

$$\tilde{V}^{\mathrm{MGPF}} = \mathcal{C} \times_1 (\boldsymbol{v}^{(1)^T} \boldsymbol{v}^{(1)}) \tilde{\boldsymbol{v}}^{(1)^T} \times_2 (\boldsymbol{v}^{(2)^T} \boldsymbol{v}^{(2)}) \tilde{\boldsymbol{v}}^{(2)} \cdots \times_f (\boldsymbol{v}^{(f)^T} \boldsymbol{v}^{(f)}) \tilde{\boldsymbol{v}}^{(f)} \tag{13}$$

It should be noted that this transformation does not change the representation. Then one obtains:

$$\tilde{\boldsymbol{V}}^{\mathrm{MGPF}} = \mathcal{V} \times_1 \tilde{\boldsymbol{\gamma}}^{(1)^T} \times_2 \tilde{\boldsymbol{\gamma}}^{(2)^T} \cdots \times_f \tilde{\boldsymbol{\gamma}}^{(f)^T} \tag{14}$$

where $\mathcal{V}$ is the tensor of the energies on the coarse grid and $\tilde{\boldsymbol{\gamma}}^{(\kappa)} = \boldsymbol{\rho}^{(\kappa)'} \boldsymbol{\rho}^{(\kappa)-1}$ is the new MGPF basis set. Both quantities, *core* tensor (V) and potential density matrices are directly computed by interfacing the MGPF routine of MCTDH to the openMOPAC software package.

2.4. Calculation of Vibrational Properties: Eigenenergies and Eigenstates

To provide a stringent test to the quality of our series of *chemically accurate* SRP-PES, in addition to RMSEs we have also computed ground and vibrationally excited eigenstates and compared them to those of the reference PES (Richter et al., 2004). These vibrational calculations have been computed using the Heidelberg version of the MCTDH software package (Worth et al., 2016) using our SRP-MGPF PES, as described above. It should be highlighted that the problem we are considering (HONO) features a double well and, consequently, single-reference approaches (e.g., QFF) are not well-suited to its study.

The calculation of the vibrational eigenstates and eigenenergies has been performed by propagating a guess WF in negative imaginary time using the so-called Improved Relaxation method (Meyer and Worth, 2003; Meyer et al., 2006). The MCTDH equations of motion (EOM) are here obtained through a time-independent variational principle. As a result, the propagated configuration interaction coefficients

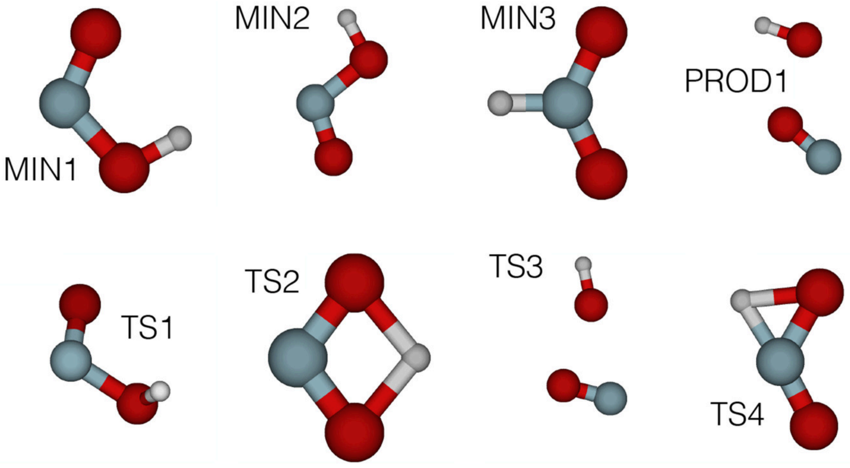

FIGURE 3 | MP2/cc-pVDZ (intermediate HL) structures of HONO automatically obtained using the TSSCDS algorithm on a PM7 (LL) PES. Target geometries in the *cis-trans* isomerization region (MIN1, MIN2, TS1) were subsequently reoptimized at the CCSD(T)/cc-pVQZ (final HL) level of theory.

(A, see Equation 6) are obtained through diagonalization of the Hamiltonian in the basis of the configurations:

$$\sum_L \langle \Phi_J | H | \Phi_L \rangle \, A_L = E \, A_J \,, \qquad (15)$$

and the single-particle basis functions (SPFs) are evolved in imaginary time using the *standard* MCTDH EOM (Beck et al., 2000). This iterative procedure is repeated until convergence in the energy. Moreover, a block version of this algorithm, the so-called Block Improved Relaxation, can be used to converge several eigenstates simultaneously, thus leading to the determination of a set of vibrationally excited states.

3. RESULTS AND DISCUSSION

In this section, we present the application of the SRP-MGPF methodology to the actual computation of the HONO (6D) PES for the *cis-trans* isomerization region, which has become a benchmark for this type of studies (Baranov and Oseledets, 2015; Pradhan and Brown, 2017). In the following subsections, we shall discuss the details on the generation of the fitting reference set of points, the reparametrization of the semiempirical Hamiltonian (SRP), and the technical details concerning the direct MGPF tensor decomposition of the SRP-PES into Tucker form. It should be stressed that the novelty and robustness of our approach resides in the fact that requires a minimum intervention of the user, thus qualifying as a *quasi*-black box approach. For the time being, we have interfaced the software openMOPAC to the MCTDH software package through the use of the MGPF tensor decomposition algorithm (Peláez and Meyer, 2013), hence allowing quantum dynamical simulations on a SRP-MGPF PES.

3.1. Computation of the SRP-MGPF PES for the *cis-trans* Isomerization Region in the HONO System (6D)

The first stage in our automated fitting procedure has been the determination of the stationary points of HONO, accomplished through the use of the TSSCDS package (Barnes et al., 2019), as described in section 2.2. Starting from a single random input geometry, LL guess structures have been obtained (see Martínez-Núñez, 2015a,b for a detailed discussion). **Figure 3** presents the corresponding MP2/cc-pvDZ structures. The relevant geometries for our study *cis* (MIN1), *trans* (MIN2) as well as the TS connecting them (TS1) have been reoptimized at the CCSD(T)/cc-pVQZ level of theory. Their geometrical parameters and harmonic frequencies are presented in **Tables S10–S13**. The reason behind the choice of this level of theory is that we have taken as model chemistry the CCSD(T)/cc-pVQZ quality analytical PES of Richter et al. (2004)

The generation of the remaining reference geometries and corresponding energies has been done according to our heuristic approach described in section 2.2. A set of geometries in the form of nD-product grids (n=1, 2) and 6D-random structures have been generated using the three lowest energy stationary points of HONO as pivotal geometries, namely: *cis*, *trans*-conformers and the corresponding TS (see **Figure 3**: MIN1, MIN2, and TS1, respectively). Moreover, the reaction path among them has been taken into account through a piecewise Linear Interpolation in Internal Coordinates (LIIC) (Soto et al., 2006) between the *cis*-TS and TS-*trans* pairs of stationary points (see **Figure S1**) as well as a cloud of distorted structures around them. To ensure that the latter remain close to the *reaction path* (LIIC), each i-th geometry along the LIIC has been generated by distorting along a set of directions resulting from the linear combination of the normal modes of the end structures according to:

$$\Delta\vec{Q}_i = (1 - X_i) \cdot \vec{Q}_{init} + X_i \cdot \vec{Q}_{fin} \qquad \vec{Q} \in \mathbb{R}^{3N-7}$$

where $\vec{Q}_{fin}$ =TS1, $\vec{Q}_{init}$ =MIN1/MIN2. X_i is a number that depends on the *distance* to the end structure. The closer to $\vec{Q}_{fin}$ the more $\Delta\vec{Q}$ resembles the normal modes of the end structure (TS1). Each of our LIIC consists of 50 points and the aforementioned *distance* is simply taken as the ordinal *i* within the LIIC. It should be noted that the torsion mode has not been included ($3N - 7$ modes in total), since it approximately corresponds to the reaction coordinate. Finally, for a given displacement ($\Delta\vec{Q}$), the geometries around the *i*-th geometry along the LIIC have been computed as:

$$\vec{R}_i = \vec{R}_i^{(0)} + \sum_{j=1}^{3N-7} f_j \cdot \Delta\vec{Q}_{i,j}$$

where $\vec{R}_i^{(0)}$ is the original geometry of the *i*-th point of the LIIC, f_j is a small random factor, and $\Delta\vec{Q}_{i,j}$ is the *j*-th component of $\Delta\vec{Q}_i$.

This systematic manner of generating reference points serves us to control the convergence of the RMSE error at each expansion order, in other words, how insensitive the RMSE is to an increase in the density of points in specific directions (or combinations thereof). This, in turn, provides us with a good estimate of the *lowest possible* number of reference geometries at each stage. In **Table 1**, we present the different convergence stages in terms of number of fitting points used together with the associated optimization algorithm. As it can be observed, at each specific stage, we either increase the density of points in the indicated directions (*modes/coordinates*) or add a new class of points in the form of a LIIC, for instance.

The first stage consists on a global optimization (MLSL) followed by a local one (BOBYQA) using a small number of judiciously chosen points: the RXN and a cloud of random geometries around them, adding up to a total of 53 points. This has enabled a very large number of iterations (10^5). The underlying hypothesis behind this calculation is that a reasonable and cheap estimate of the *global minimum* (set of SRP parameters yielding the minimum RMSE) can be obtained. Our best set of parameters at this stage ($\boldsymbol{\zeta}_{53}$, where 53 is the number of fitting points) yielded an initial RMSE of 806.8 cm^{-1} (**Table S1**). In the subsequent stages, we have performed local optimizations (BOBYQA) with 2,000 iterations. Before proceeding any further, we would like to justify the use of a global algorithm exclusively at the first stage, in other words, ζ_{53} must indeed correspond to a set near the *global* minimum or a local deep minimum. First, from a computational perspective, it should be noted that a small number of fitting points is ideally suited for this task. Second, we have performed calculations justifying this fact. **Table S2** (column 2) presents the BOBYQA variation of the RMSE for an increasing number of 1D-sets of fitting points. It can be observed that upon increase of this number, from 192 until 2088 fitting points, the RMSE monotonically decreases from 482.13 cm^{-1} till 365.13 cm^{-1}. According to our reasonings above, one should take the SRP parameters of the last set of points ($\boldsymbol{\zeta}_{1542}$ or $\boldsymbol{\zeta}_{2088}$) corresponding to the best RMSE of the 1D-series. For the sake of efficiency, we considered the $\boldsymbol{\zeta}_{1542}$. With this set of SRP parameters, we recomputed the whole series of RMSEs for the different sets of 1D-points and we observed a very close agreement with the BOBYQA values, except for the 192 set. This shows that indeed all sets of parameters of this series (from $\boldsymbol{\zeta}_{367}$ on) lie within the *same* RMSE landscape region (see **Figure 4**) and, in turn, validates our initial approach with a small number of *representative* points. One can then safely conclude that just 367 fitting points are necessary to improve the SRP-fitting at the 1D-level. Hence, subsequent 2D optimizations will start with the ($\boldsymbol{\zeta}_{367}$) set. A detailed description of all stages and RMSE values is presented in **Tables S1–S9**. A somewhat more complete information can be obtained through the cumulative error computed by addition of the RMSEs resulting form the configurations up to a certain energy value (see **Figure 5**). It can observed that for all sets of parameters, with the exception of $\boldsymbol{\zeta}_{53}$, the RMSEs remain below the limit of chemical accuracy (1 kcal/mol$\approx$ 350 cm^{-1}) within the targeted PES region (*cis-trans* isomerization). Moreover, in the last stage we have removed all structures with energies above 5000 cm^{-1} (above the classical barrier) and included an extra set of random points around the stationary points. This new set of points has been used to BOBYQA reoptimize the SRP. We observe a clear improvement of the RMSE in such a way that, up to 8000 cm^{-1}, the RMSE is inferior to the chemical accuracy level. The correctness of these results has been supported by a calculation using a validation set consisting of 1200 6D random points with energies below 12000 cm^{-1} for which the same pattern is obtained. We have also compared the geometries and harmonic frequencies of all stationary points at the reference *ab initio* level of theory and at the SRP level for each stage. Geometries are displayed in **Tables S10–S12** and harmonic frequencies are shown in **Table 2**. As it can be observed, SRP does indeed improve, in terms of both geometrical parameters and harmonic frequencies, with respect to the original PM7 and, furthermore, we obtain a very good agreement with the reference *ab initio* data. This is particularly true for the last stage ($\boldsymbol{\zeta}_{1084}$).

To finalize this section, we present in **Figure 6** a comparison of 2D projections of the *cis-trans* isomerization regions for: (i) the reference surface, (ii) the SRP-PES($\boldsymbol{\zeta}_{1084}$); and (iii) the PM7 semiempirical Hamiltonian. These contour plots have been obtained through orthogonalization of the two LIIC vectors used in **Figure 6**. The positive effect of the reparametrization can be clearly observed: while PM7 provides a *blurred* description of the TS region, the SRP-PES reproduces it correctly.

3.1.1. Classical Molecular Dynamics on the SRP-PES

As a first test of the quality of the SRP-PES, we have carried out classical molecular dynamics simulations for the HONO system in full dimensionality using the VENUS96 software package (Hu et al., 1991). Classical trajectories have been run using the reference PES (Richter et al., 2004). The energies of the so-obtained geometries have been subsequently computed at the SRP-PES level and compared to the original calculation. Starting from the equilibrium geometries of the *cis* and *trans* isomers, we have propagated for 1 *ps* each trajectory with a time-step of 5*fs*. The vibrational energy of each starting geometry was classically distributed in a random way between all normal modes using the option *normal mode sampling* of the VENUS

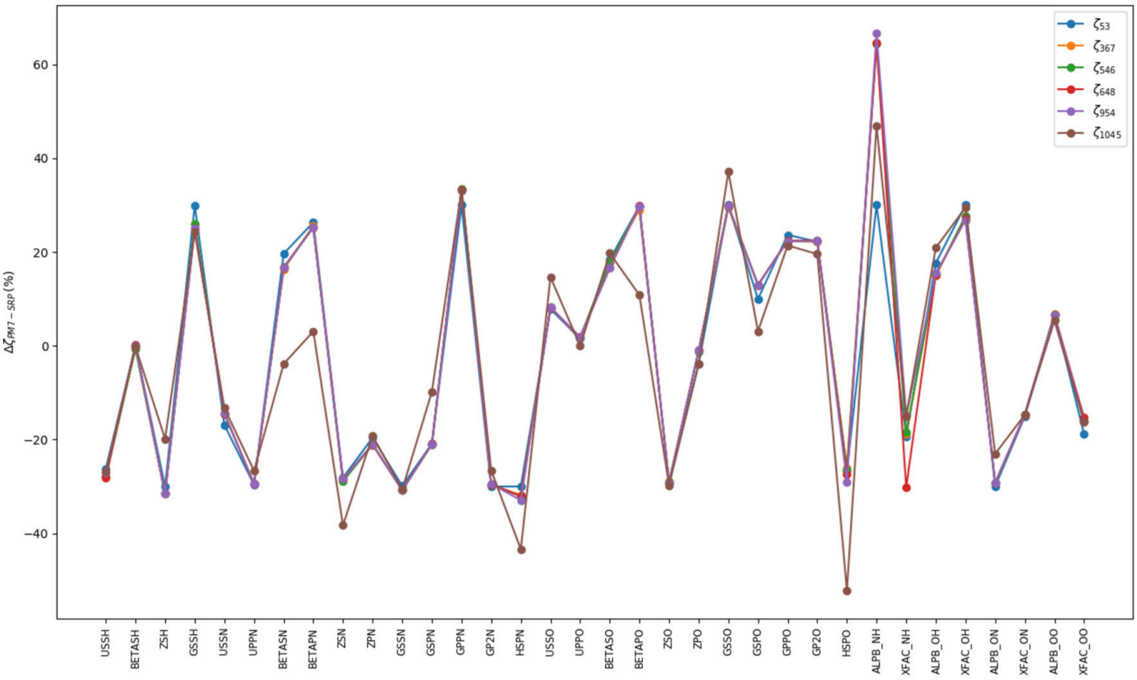

FIGURE 4 | Percentage of variation of the SRP parameters with respect to the original PM7 ones. Each fitting stage is represented by its *optimal* parameters, ζ_N, where N is the number of points used in the process (see **Table 1**). On abscissas we present the label of semiempirical parameters for the different type of atoms in HONO. Standard semiempirical parameter labeling has been used (Stewart, 2013). Parameters from USSH until HSPO correspond to a single type of atom whereas parameters labeled ALPB$_{XY}$ and XFAC$_{XY}$ correspond to two-atom ones (atom X and atom Y).

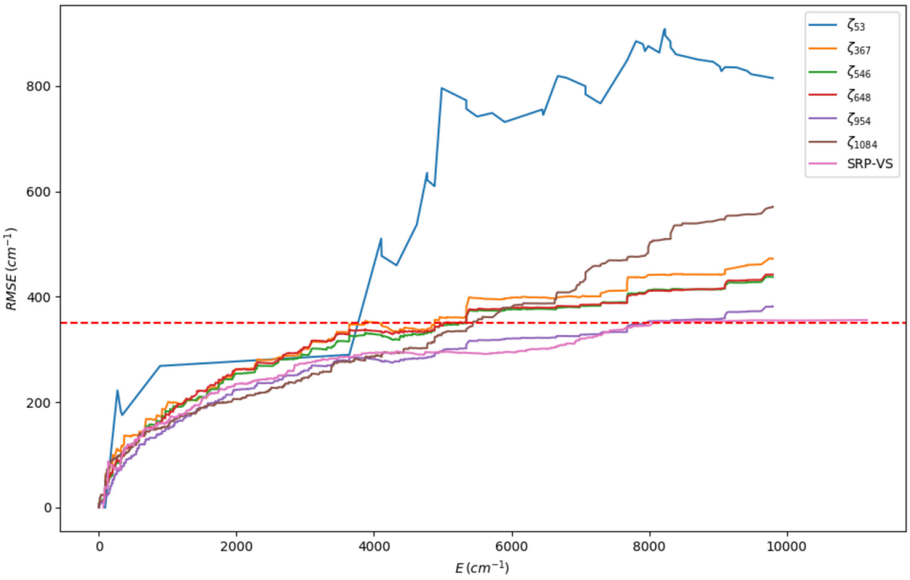

FIGURE 5 | Cumulative RMSE for each SRP-fit labeled by its set of parameters, ζ_N, where N is the number of points used in the fit (see **Table 1**). The last set (ζ_{VS}) corresponds to the validation set. The red dotted horizontal line represents the value of the chemical accuracy (1 kcal/mol≈ 350 cm^{-1}).

software. We have computed 10 trajectories per isomer, each isomer having 4 different vibrational energies (5, 10, 15, and 20 kcal/mol) thus making a total of 80 trajectories and 16,080 geometries. In **Figure 7**, we present a comparison of the variation of the potential energies along two of these trajectories. As it can be observed, the PM7 largely deviates from the reference calculation both in their relative values and the phase, whereas SRP-PES follows closely the *ab initio* values. In particular, it

TABLE 2 | Harmonic frequencies of the normal modes of each stationary point at the CCSD(T)/cc-pVQZ *ab initio* level of theory and corresponding values for the PM7 method and the SRPs in the different stages of the optimization.

	Harmonic frequencies (cm^{-1})							
	Ab initio	ζ_{PM7}	ζ_{53}	ζ_{367}	ζ_{546}	ζ_{648}	ζ_{954}	ζ_{1084}
TS	-599.2	-553.6	-581.0	-565.1	-568.7	-570.3	-573.3	-606.8
	559.1	621.7	512.2	467.1	465.5	463.6	463.8	597.2
	791.2	1021.3	654.7	649.5	652.7	653.6	654.9	738.4
	1122.3	1175.3	1174.3	1092.3	1099.8	1095.8	1106.8	1195.4
	1728.0	1839.5	1763.8	1705.8	1709.4	1710.5	1711.0	1737.9
	3785.3	2801.7	3747.9	3568.9	3585.3	3585.0	3586.4	3736.2
cis	648.7	589.0	615.4	613.0	616.1	618.2	619.3	622.5
	687.9	629.2	724.9	712.0	718.4	716.0	718.4	698.9
	901.9	1084.8	745.3	715.4	721.0	721.7	728.4	854.2
	1350.9	1346.0	1316.4	1252.5	1255.9	1253.7	1262.3	1369.5
	1675.5	1823.5	1725.5	1693.2	1696.6	1698.3	1701.6	1719.1
	3632.1	2802.9	3668.9	3504.6	3519.8	3520.0	3521.4	3667.3
trans	574.8	455.9	517.1	515.2	517.1	518.1	521.6	540.5
	633.1	609.8	533.3	515.2	519.1	523.7	528.3	602.5
	839.6	1096.0	730.5	736.6	741.7	744.7	748.9	835.1
	1319.3	1308.8	1232.9	1130.0	1136.9	1131.6	1148.4	1264.6
	1732.6	1826.5	1715.9	1666.7	1670.2	1671.9	1674.8	1704.7
	3790.8	2828.3	3815.8	3662.7	3678.9	3680.9	3682.9	3796.1

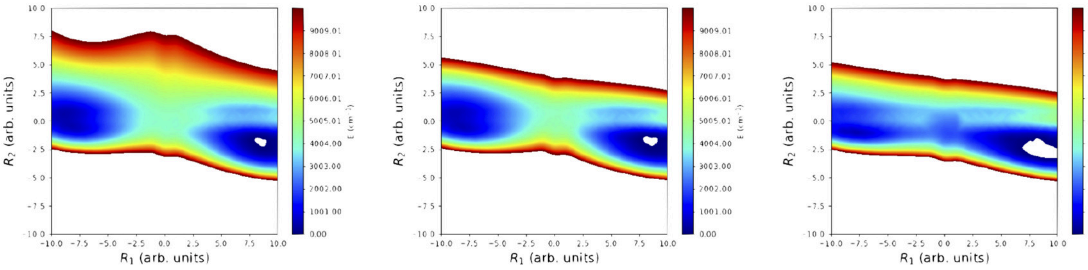

FIGURE 6 | Comparison of the 2D projections of the cis-trans isomerization region for: (i) reference PES (Richter et al., 2004) (left panel); (ii) SRP-PES (ζ_{1084}) (middle panel); and (iii) PM7-PES (right panel). These projections have been obtained by orthonormalization of two linear interpolation (LIIC) vectors as described in Soto et al. (2006).

is remarkable the fact that for low energies PM7 presents a large amount of structures with energies below the value of the global minimum, the *trans* conformer. To finalize this subsection, we would like to provide some performance features of the SRP-PES which directly show the efficiency of the underlying openMOPAC software. In the case of the HONO, from an average of the order of $\sim 10^4$ points, we have obtained a mean CPU-time of 10^{-2} s per single-point energy. Moreover, Hessians are computed in less than a second. This properties make SRP approaches suitable for any on-the-fly type of calculation. In particular, we are currently exploring their use with non-grid based quantum dynamical methods such as the Direct-Dynamics Variational Multiconfigurational Gaussian (DD-vMCG) method (Richings et al., 2015).

3.2. Full Quantum Analysis of the Vibrational Properties of the SRP-PES for the *cis-trans* HONO System (6D)

To further assess the quality of our SRP-PES we have computed vibrational properties by means of MCTDH quantum dynamical calculations and the results have been compared to the ones from the reference PES (Richter et al., 2004). More specifically, ground and excited vibrational states as well as vibrational spectra, in the form of Fourier transforms of autocorrelation functions. At this point, it should be recalled that our main goal is not to achieve spectroscopical accuracy but to provide PESs, in a fully automated fashion, accurate enough to disentangle chemical processes.

3.2.1. MGPF Tensor Decomposition of the HONO 6D PES

To *interface* the SRP-PES with the MCTDH quantum dynamics software package, we have used the Multigrid POTFIT tensor decomposition algorithm (Peláez and Meyer, 2013). More specifically, all PES *calls* within the MGPF workflow have been addressed directly to the openMOPAC software package using an external set of *optimal* SRP parameters. In other words, at each grid point, i.e., configuration, a SCF process is performed. Of course, this is only possible due to the high efficiency of the underlying PM7 frame. This fact, precisely, has allowed us to circumvent the issues encountered in previous studies in which the *ab initio* energies were generated directly from a quantum chemical calculation thus severely limiting the level of theory which could be applied.

We have carried out *bottom-up* MGPF calculations Peláez and Meyer (2013) to the different SRP-PESs at different parameter optimization stages. In **Table S14**, we present a comparison in terms of CPU time and memory needs for a reference exact Tucker decomposition (using POTFIT, PF) (Jäckle and Meyer, 1996) and the different MGPF tensor decomposition levels that we have used in this work. The full primitive grid, needed in PF, consists of $2.804 \cdot 10^7$ points. In contrast, the coarse grids in MGPFs include every third, fourth, or fifth fine grid point for each DOF. These coarse grids have been labeled *ev*3, *ev*4, and *ev*5 and consist of 172,800, 51,200, and 18,432 coarse grid points, respectively. The MGPF partial grids increase these figures by a factor <10. This is due to the fact that the contracted mode lies fully in the fine grid (see section IIIB in Peláez and Meyer, 2013). Hence, as expected, MGPF is orders of magnitude less demanding that an exact decomposition. The global RMSE values show that MGPF PES are accurate, cheap and, more importantly, add a very small (global, full grid) error to the PES. Finally, it should also be highlighted that none of our SRP-PES present energies below the global minimum (*trans* conformer), whereas the PM7 does. In other words, PM7 presents artificial PES structure when compared to the reference one. We have observed that even the simplest SRP optimization corrects this wrong behavior.

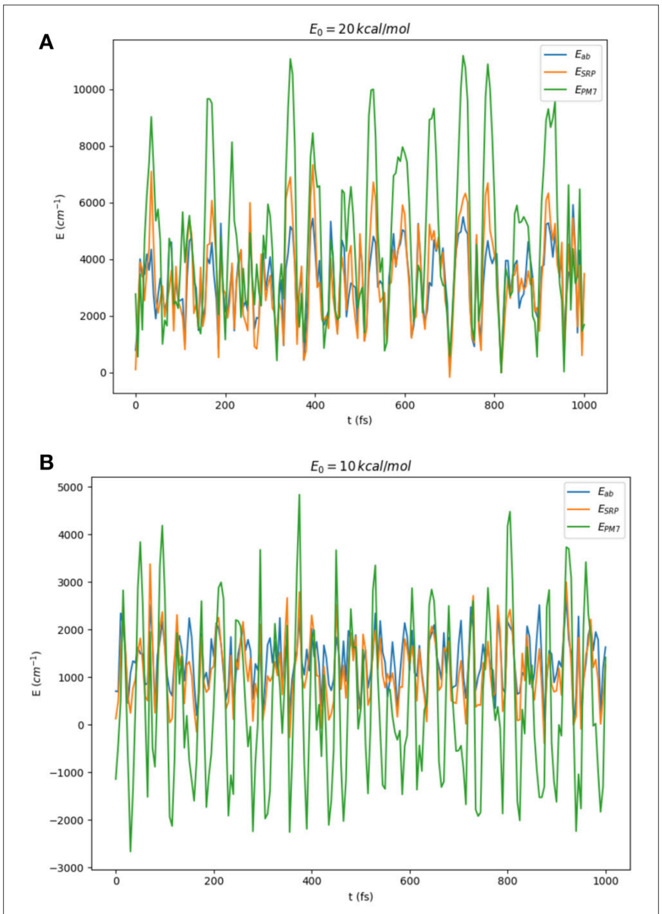

FIGURE 7 | Comparison of ab initio (blue line), PM7 (green line), and SRP-PES (ζ_{1084}) (orange line) energies for the geometries generated in classical on-the-fly trajectories of HONO(6D) with total energies (randomly distributed among all modes) of 10 and 20 kcal/mol starting at: **(A)** the *trans*-conformer and **(B)** the *cis*-conformer.

3.2.2. MCTDH Quantum Molecular Dynamics on the SRP-MGPF

As discussed in section 2.3, MCTDH requires the discretisation of the configuration space. The HONO (6D) molecule has been represented in internal coordinates (see **Figure 8**) as in previous works (Peláez and Meyer, 2013; Pradhan and Brown, 2017), and a Discrete Variable Representation (DVR) grid has been defined accordingly (see **Table 3**). We have performed ground and excited eigenstate vibrational calculations for the reference PES, the PM7-MGPF PES as well as for selected SRP-MGPF PES using the Improved Relaxation algorithm and its Block version, as implemented in the Heidelberg version of MCTDH (Meyer et al., 2006). We have combined the physical modes into logical particles as follows: [ϕ=15], [d_{OH}=10] [u_2, d_{ON}=25], [u_1, d_{NO}=25], where the number represents the number of single-particle functions (SPFs) and $u_i = \cos\theta_i$ (see **Figure 8**). In all cases, the initial wave packet has been propagated in *negative imaginary time* (see section 2.4) during 500 fs.

With respect to ground state energies, the reference PES yields a value of 4367.7 cm^{-1} for the Zero Point Energy (ZPE) and the PM7-MGPF PES a value of 3221.3 cm^{-1}, well off the analytical one. We attribute this discrepancy to the artificial structure of

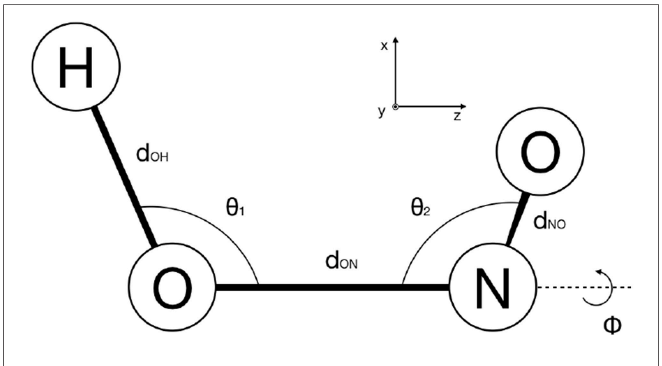

FIGURE 8 | Definition of the internal coordinates of HONO used in this work.

the PES revealed by the presence of *negative* energies (geometries with energies below the global minimum, *trans* conformer) as discussed in section 3.2.1) and clearly illustrated in **Figure 7**. On the other hand, concerning the SRP-MGPF PESs, a nice convergence can be observed upon increase of the number of fitting points, toward a final value of 4332.8 cm^{-1} which compares well with the analytical one. It is also remarkable that a simple fit using only 53 fitting points already leads to a qualitative improvement with respect to PM7. Moreover, our results show that the ZPE values are somewhat insensitive to the size of the coarse grid (cf. last three rows of **Table 4**). Consequently, we shall use hereafter the *ev*5 SRP-MGPF scheme.

We have also computed the 20 lowest-lying vibrational eigenstates of HONO (**Table 5**). It should be noted that this energy interval spans all HONO fundamentals except the OH stretching mode. For this, we have considered four different PES, namely: (i) PM7-MGPF, SRP-MGPF with $\boldsymbol{\zeta}_{53}$ and $\boldsymbol{\zeta}_{1084}$, as well as the reference (exact) PES. The first remark to be done is that the original PM7-MGPF PES fails to predict the initial vibrational state corresponding to the *ground state* of the *cis* conformer (Richter et al., 2004). In contrast, even at the minimum level of reparametrization ($\boldsymbol{\zeta}_{53}$), this eigenstate is obtained. Furthermore, this incorrect behavior worsens upon increase of the energy. In fact, eigenenergies are off by several hundreds of cm^{-1} in almost the its whole range. This can be readily understood by simple observation of the 2D contour plots of the *cis-trans* region of the PES (see **Figure 6**). In contrast, both SRP-MGPFs nicely follow the reference values and, what is more important, the discrepancies (of the order of tens of cm^{-1}) do not increase but remain, in average, constant.

Finally, to take into account higher excited vibrational states, we have computed a vibrational spectrum by Fourier transform

TABLE 3 | Definition of the MCTDH primitive grid: HO denotes a harmonic oscillator (Hermite) and cos a cosine Discrete Variable Representation (DVR) basis functions.

DOF	DVR	N	Range
d_{OH}	HO	18	[1.30, 2.45]
d_{NO}	HO	13	[1.90, 2.60]
u_2	HO	13	[-0.65, -0.10]
d_{ON}	HO	16	[2.10, 3.25]
u_1	HO	18	[-0.65, 0.25]
ϕ	cos	32	[0, $2\pi/2$]

*N is the number of primitive (fine) grid points. The range represents the first and last grid points in atomic units for the distances and ϕ is the torsion angle in radians. Cosines of the valence angles have been used: $u_i = \cos\theta_i$. See **Figure 8** for the definition. Physical degrees of freedom have been combined into logical modes or* particles *according to the following scheme: $[\phi]$, $[d_{OH}]$ $[u_2, d_{ON}]$, $[u_1, d_{NO}]$. The first particle (ϕ) has been contracted in MGPF (see section IIIB in Peláez and Meyer, 2013).*

TABLE 4 | Ground state energies of HONO using PESs of different quality.

Set	MGPF	ZPE (cm^{-1})
$\boldsymbol{\zeta}_{PM7}$	*ev*4	3221.3
$\boldsymbol{\zeta}_{53}$	*ev*4	4070.7
$\boldsymbol{\zeta}_{648}$	*ev*4	4095.0
$\boldsymbol{\zeta}_{1084}$	*ev*4	4332.8
$\boldsymbol{\zeta}_{1084}$	*ev*5	4330.8
$\boldsymbol{\zeta}_{1084}$	*ev*3	4332.9

*The first column indicates the set of SRP parameters used, labeled by its set of parameters, ζ_N, where N is the number of points used in the fit (see **Table 1**). The second column presents the size of the MGPF coarse grid: evn indicates a coarse grid in which every (ev) n-th fine grid point has been considered (see section 3.2.1). The final column presents the Zero Point Energies (ZPE) for each of the previous PES.*

TABLE 5 | Comparison of the 20 lowest vibrational eigenvalues of HONO for different PESs denoted by its set of parameters, $\boldsymbol{\zeta}_N$, where N is the number of points used in the fit (see **Table 1**).

	Vibrational eigenenergies (cm^{-1})			
	$\boldsymbol{\zeta}_{PM7}$	$\boldsymbol{\zeta}_{53}$	$\boldsymbol{\zeta}_{1084}$	**Analytical**
	0.0	0.0	0.0	0.0
	593.6	163.0	88.5	94.1
	794.3	604.7	597.1	600.8
	1070.6	693.2	703.9	710.7
	1151.5	706.9	822.3	795.9
	1186.3	888.9	917.9	944.1
	1365.9	1134.3	1012.5	1055.4
	1403.1	1204.8	1189.7	1188.1
	1641.3	1221.6	1234.7	1264.9
	1659.6	1263.0	1317.9	1306.6
	1751.1	1308.9	1363.5	1312.8
	1773.1	1361.6	1417.2	1385.3
	1811.5	1395.7	1451.1	1404.8
	1869.9	1424.9	1530.5	1547.9
	1968.7	1426.3	1607.7	1574.9
	2011.4	1612.4	1633.9	1640.9
	2060.3	1656.9	1690.9	1689.9
	2118.1	1698.3	1743.0	1726.0
	2136.5	1748.6	1778.7	1762.4
	2226.5	1842.0	1785.8	1779.7
	2253.3	1853.0	1807.3	1829.0
RMSE	360.2	58.4	24.5	–
	N/A	–	[42.0]	–
MAD	53.7	38.3	23.7	–
	N/A	–	[25.5]	–

Energies have been computed by MCTDH Block Improved Relaxation (see section 2.4). All PESs have been MGPFitted *using a coarse grid consisting on 18,432 points, the so-called ev5 (see section 3.2.1). The first column presents the PM7-MGPF values (PM7), second and third correspond to SRP-MGPF with $\boldsymbol{\zeta}_{53}$ and $\boldsymbol{\zeta}_{1084}$, respectively. The last column presents the corresponding eigenenergies obtained using the analytical surface by Richter et al. (2004). The last four rows present the RMSE and the mean-absolute deviation (MAD) of each set of eigenvalues with respect to the analytical ones. The values in square brackets indicate the RMSE and MAD values taking into account the corresponding OH stretching anharmonic frequencies. The latter have been obtained through Fourier transform of an autocorrelation function (see **Figure 9**): (i) Analytical: 3533.8 cm^{-1} and (ii) $\boldsymbol{\zeta}_{1084}$: 3695.7 cm^{-1}. It should be noted that the PM7 values could not be determined (indicated by N/A) owing to a wrong behavior of the PM7-PES at this energy range (see **Figure 9**).*

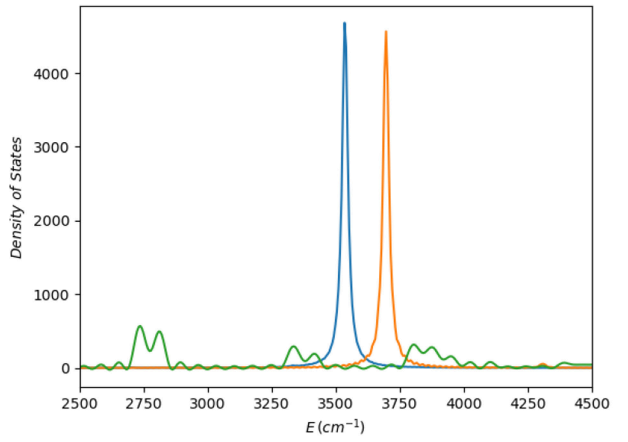

FIGURE 9 | Vibrational spectra computed as the Fourier transform of the autocorrelation function obtained after excitation of one quantum in the OH stretching vibration centered the *cis* conformer region: (i) green line corresponds to the PM7-MGPF PES; (ii) orange line to the SRP-MGPF (ζ_{1084}) PES; and (iii) blue line, the reference PES (*ab initio*) (Richter et al., 2004).

of the autocorrelation function corresponding to the dynamics of a wave packet generated by excitation of a quantum of energy in the OH stretching mode in the *cis* region of the potential. As observed (**Figure 9**), the PM7-MGPF spectrum is radically different to that of the reference PES, whereas the SRP-MGPF one shows the correct behavior. Apart from the, certainly not unexpected, shift in energy, both reference PES and SRP-MGPF reveal that the OH mode is practically uncoupled from the rest.

4. CONCLUSIONS AND FUTURE PROSPECTS

We have introduced Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF) a methodology which permits the generation of global chemically accurate Potential Energy Surfaces in sums-of-products (Tucker) form in a *quasi* black-box manner starting from a random input geometry. The SRP-MGPF workflow combines: (i) the automated determination of stationary points of a Potential Energy Surface (PES); (ii) the reparametrization of a Semiempirical Hamiltonian (SRP) using high-level *ab initio* data; and (iii) direct tensor-decomposition of the resulting SRP-PES with the Multigrid POTFIT (MGPF) algorithm. The resulting surface can be used with any on-the-fly dynamical software or, after MGPF, with grid-based quantum dynamical method, in particular the Multiconfiguration Time-Dependent Hartree (MCTDH) method. We have proven the validity of this method by fitting the SRP-MGPF PES for the HONO system in full dimensionality (6D) and reproducing, to a good agreement, the vibrational properties of a surface of CCSD(T)/cc-pVQZ quality. Current work deals with the extension of the method to treat coupled electronic excited states. To finalize, it should be highlighted that SRP-MGPF provides an inexpensive and accurate enough means of performing full-dimensional chemically meaningful quantum or classical simulations.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript/**Supplementary Files**.

AUTHOR CONTRIBUTIONS

DP conceived the original idea, managed the project, and wrote the first draft of the manuscript. RP-B and DP were responsible for its implementation. RP-B was main responsible for the development of the necessary software. EM-N has been responsible for the TSSCDS calculations and classical trajectories. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

RP-B thanks the Région Hauts-de-France and the Université de Lille for a Ph.D. Fellowship. RP-B and DP thank the Région Hauts-de-France and the Ministére de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development for their financial support. This work has been partially supported by the Agence Nationale de la Recherche through the Labex CaPPA (ANR-11-LABX-0005-01). EM-N thanks Xunta de Galicia (Research Grant No. ED431C 2017/17). EM-N is very grateful to l'Institut de Physique for a CNRS visiting professorship to the Laboratoire PhLAM of the University of Lille.

ACKNOWLEDGMENTS

The authors thank Dr. James Stewart for helpful and lively discussions as well as for having granted us permission for using the openMOPAC package as default in SRP-MGPF. DP thanks Prof. Dr. Diego Troya for discussions on reparametrizations of semiempricial Hamiltonians. DP would like to thank Prof. Dr. Hans-Dieter Meyer for encouraging discussions. The authors thank Dr. Alejandro Rivero Santamaría for discussions on the NLOpt libraries and Prof. Dr. Maurice Monnerville for fruitful scientific exchanges. The authors thank Dr. Sabine Kopec for providing the initial LL-TSSCDS calculations and Hugo Geindre for the use of his script for the generation of random geometries based on normal modes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2019.00576/full#supplementary-material

REFERENCES

Ackley, D. H. (1987). *A Connectionist Machine for Genetic Hillclimbing*. Norwell, MA: Kluwer Academic Publishers.

Ávila, G., and Carrington, T. Jr. (2009). Nonproduct quadrature grids for solving the vibrational Schrödinger equation. *J. Chem. Phys.* 131:174103. doi: 10.1063/1.3246593

Baranov, V., and Oseledets, I. (2015). Fitting high-dimensional potential energy surface using active subspace and tensor train (AS+TT) method. *J. Chem. Phys.* 143:174107. doi: 10.1063/1.4935017

Barnes, G. L., Kopec, S., Peláez, D., Rodríguez, A., Rodríguez-Fernández, R., J. J. P., Stewart, P. T., et al. (2019). Available online at: https://rxnkin.usc.es/index.php/AutoMeKin

Barone, V. (2005). Anharmonic vibrational properties by a fully automated second-order perturbative aproach. *J. Chem. Phys.* 122:014108. doi: 10.1063/1.1824881

Beck, M. H., Jäckle, A., Worth, G. A., and Meyer, H.-D. (2000). The multi-configuration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wave packets. *Phys. Rep.* 324, 1–105. doi: 10.1016/S0370-1573(99)00047-2

Braams, B. J., and Bowman, J. M. (2009). Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* 28, 577–606. doi: 10.1080/01442350903234923

Burghardt, I., Giri, K., and Worth, G. A. (2008). Multi-mode quantum dynamics: the G-MCTDH method applied to the absorption spectrum of pyrazine. *J. Chem. Phys.* 129:174104. doi: 10.1063/1.2996349

Christensen, A. S., Kubar, T., Cui, Q., and Elstner, M. (2016). Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chem. Res.* 116, 5301–5337. doi: 10.1021/acs.chemrev.5b00584

Christiansen, O. (2004). Vibrational coupled cluster theory. *J. Chem. Phys.* 120, 2149–2159. doi: 10.1063/1.1637579

Dawes, R., Thompson, D. L., Guo, Y., Wagner, A. F., and Minkoff, M. (2007). Interpolating moving least-squares methods for fitting potential energy surfaces: computing high-density potential energy surface data from low-density ab initio data points. *J. Chem. Phys.* 126:184108. doi: 10.1063/1.2730798

Frankcombe, T. J., and Collins, M. A. (2011). Potential energy surfaces for gas-surface reactions. *Phys. Chem. Chem. Phys.* 13, 8379–8391. doi: 10.1039/c0cp01843k

Gatti, F. (ed.). (2014). Molecular Quantum Dynamics. Heidelberg: Springer.

Haller, A., Peláez, D., and Bande, A. (2019). Inter-coulombic decay in laterally-arranged quantum dots controlled by polarized lasers. *J. Phys. Chem. C* 123, 14754–14765. doi: 10.1021/acs.jpcc.9b01250

Hu, X., Hase, W. L., and Pirraglia, Y. (1991). Vectorization of the general monte carlo classical trajectory program venus. *J. Comput. Chem.* 12, 1014–1024. doi: 10.1002/jcc.540120814

Jäckle, A., and Meyer, H.-D. (1996). Product representation of potential energy surfaces. *J. Chem. Phys.* 109:3772. doi: 10.1063/1.471513

Jiang, B., and Guo, H. (2014). Permutation invariant polynomial neural network approach to fitting potential energy surfaces. III. Molecule-surface interactions. *J. Chem. Phys.* 141:034109. doi: 10.1063/1.4887363

Johnson, S. G. (2011). *The NLopt Non-linear-optimization Package*. Available online at: http://ab-initio.mit.edu/nlopt

Jones, E., Oliphant, T., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python*. Available online at: http://www.scipy.org/

Kan, A. H. G., and Timmer, G. T. (1987). Stochastic global optimization methods part II: multi level methods. *Math. Program.* 39, 57–78. doi: 10.1007/BF02592071

Kolda, T. G., and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* 51, 455–500. doi: 10.1137/07070111X

Kopec, S., Martínez-Núñez, E., Soto, J., and Peláez, D. (2019). vdW-TSSCDS - an automated and global procedure for the computation of stationary points on intermolecular potential energy surfaces. *Int. J. Quant. Chem.* 2019:e26008. doi: 10.1002/qua.26008

Kosloff, R. (1988). Time-dependent quantum-mechanical methods for molecular dynamics. *J. Phys. Chem.* 92:2087. doi: 10.1021/j100319a003

Layfield, J. P., Owens, M. D., and Troya, D. (2008). Theoretical study of the dynamics of the H+CH_4 and H+C_2H_6 reactions using a specific-reaction-parameter semiempirical hamiltonian. *J. Phys. Chem.* 128:194302. doi: 10.1063/1.2918358

Levine, B. G., Coe, J. D., Virshup, A. M., and Martínez, T. J. (2008). Implementation of *ab initio* multiple spawning in the Molpro quantum chemistry package. *Chem. Phys.* 347, 3–16. doi: 10.1016/j.chemphys.2008.01.014

Light, J. C., and Carrington, T. Jr. (2000). Discrete variable representations and their utilization. *Adv. Chem. Phys.* 114, 263–310. doi: 10.1002/9780470141731.ch4

Manzhos, S., and Carrington, T. Jr. (2006). Using neural networks to represent potential surfaces as sum of products. *J. Chem. Phys.* 125:084109. doi: 10.1063/1.2387950

Manzhos, S., Wang, X., Dawes, R., and Carrington, T. Jr. (2006). A nested molecule-independent neural network approach for high-quality potential fits. *J. Phys. Chem. A* 110, 5295–5304. doi: 10.1021/jp055253z

Martínez-Núñez, E. (2015a). An automated method to find transition states using chemical dynamics simulations. *J. Comp. Chem.* 36:222. doi: 10.1002/jcc.23790

Martínez-Núñez, E. (2015b). An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.* 17:14912. doi: 10.1039/C5CP02175H

Meyer, H.-D., Le Quéré, F., Léonard, C., and Gatti, F. (2006). Calculation and selective population of vibrational levels with the Multiconfiguration Time-Dependent Hartree (MCTDH) algorithm. *Chem. Phys.* 329, 179–192. doi: 10.1016/j.chemphys.2006.06.002

Meyer, H.-D., and Worth, G. A. (2003). Quantum molecular dynamics: propagating wavepackets and density operators using the multiconfiguration time-dependent Hartree (MCTDH) method. *Theor. Chem. Acc.* 109, 251–267. doi: 10.1007/s00214-003-0439-1

Neff, M., and Rauhut, G. (2009). Toward large scale vibrational configuration interaction calculations. *J. Chem. Phys.* 131:124129. doi: 10.1063/1.3243862

Peláez, D., and Meyer, H.-D. (2013). The multigrid POTFIT (MGPF) method: grid representations of potentials for quantum dynamics of large systems. *J. Chem. Phys.* 138:014108. doi: 10.1063/1.4773021

Peláez, D., and Meyer, H.-D. (2017). On the infrared absorption spectrum of the hydrated hydroxide ($H_3O_2^-$) cluster anion. *Chem. Phys.* 482:100. doi: 10.1016/j.chemphys.2016.08.025

Peláez, D., Sadri, K., and Meyer, H.-D. (2014). Full-dimensional MCTDH/MGPF study of the ground and lowest lying vibrational states of the bihydroxide $H_3O_2^-$ complex. *Spectrochim. Acta A* 119, 42–51. doi: 10.1016/j.saa.2013.05.008

Powell, M. J. D. (2009). *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives*. Cambridge NA Report NA2009/06. Cambridge: University of Cambridge, 26–46.

Pradhan, E., and Brown, A. (2017). A ground state potential energy surface for HONO based on a neural network with exponential fitting functions. *Phys. Chem. Chem. Phys.* 19:22272. doi: 10.1039/C7CP04010E

Puzzarini, C., Bloino, J., Tasinato, N., and Barone, V. (2019). Accuracy and interpretability: the devil and the holy grail. new routes across old boundaries in computational spectroscopy. *Chem. Res.* 119, 8131–8191. doi: 10.1021/acs.chemrev.9b00007

Rauhut, G. (2004). Efficient calculation of potential energy surfaces for the generation of vibrational wave functions. *J. Chem. Phys.* 121:9313. doi: 10.1063/1.1804174

Rauhut, G. (2007). Configuration selection as a route towards efficient vibrational configuration interaction calculations. *J. Chem. Phys.* 127:184109. doi: 10.1063/1.2790016

Richings, G. W., and Habershon, S. (2018). MCTDH on-the-fly: efficient grid-based quantum dynamics without pre-computedpotential energy surfaces. *J. Chem. Phys.* 148:134116. doi: 10.1063/1.5024869

Richings, G. W., Polyak, I., Spinlove, K. E., Worth, G. A., Burghardt, I., and Lasorne, B. (2015). Quantum dynamics simulations using gaussian wavepackets: the vmcg method. *Int. Rev. Phys. Chem.* 34:269. doi: 10.1080/0144235X.2015.1051354

Richter, F., Hochlaf, M., Rosmus, P., Gatti, F., and Meyer, H.-D. (2004). A study of mode–selective trans–cis isomerisation in HONO using ab initio methodology. *J. Chem. Phys.* 120, 1306–1317. doi: 10.1063/1.1632471

Rios, L. M, and Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *J. Global Optim.* 56, 1247–1293. doi: 10.1007/s10898-012-9951-y

Rodríguez-Fernández, R., Pereira, F. B., Marques, J. M., Martínez-Núñez, E., and Vázquez, S. A. (2017). Gafit: a general-purpose, user-friendly program for fitting potential energy surfaces. *Comput. Phys. Comm.* 217:89. doi: 10.1016/j.cpc.2017.02.008

Rossi, I., and Truhlar, D. G. (1995). Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.* 233, 231–236. doi: 10.1016/0009-2614(94)01450-A

Sanz-Sanz, C., Roncero, O., Paniagua, M., and Aguado, A. (2013). Full dimensional potential energy surface for the ground state of $H_4(+)$ system based on triatomic-in-molecules formalism. *J. Chem. Phys.* 139:184302. doi: 10.1063/1.4827640

Shalashilin, D. (2010). Nonadiabatic dynamics with the help of multiconfigurational Ehrenfest method: improved theory and fully quantum 24D simulation of pyrazine. *J. Chem. Phys.* 132:244111. doi: 10.1063/1.3442747

Silva-Junior, M. R., and Thiel, W. (2010). Benchmark of electronically excited states for semiempirical methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.* 6:1546. doi: 10.1021/ct100030j

Soto, J., Arenas, J. F., Otero, J. C., and Peláez, D. (2006). Effect of an S_1/S_0 conical intersection on the chemistry of nitramide in its ground state. a comparative CASPT2 study of the nitro-nitrite isomerization reactions in nitramide and nitromethane. *J. Phys. Chem. A* 110:8221. doi: 10.1021/jp0617219

Sparta, M., Hansen, M. B., Matito, E., Toffoli, D., and Christiansen, O. (2010). Using electronic energy derivative information in automated potential energy surface construction for vibrational calculations. *J. Chem. Theory Comput.* 6:3162. doi: 10.1021/ct100229f

Stewart, J. J. (2013). Optimization of parameters for semiempirical methods vi: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* 19:1. doi: 10.1007/s00894-012-1667-x

Stewart, J. J. P. (2016). *Mopac2016, Stewart Computational Chemistry*. Colorado Springs, CO. Available online at: http://OpenMOPAC.net

Thiel, W. (2014). Semiempirical quantum-chemical methods. *WIREs Comput. Mol. Sci.* 4:145. doi: 10.1002/wcms.1161

Toniolo, A., Granucci, G., and Martínez, T. J. (2003). Conical intersections in solution: a QM/MM study using floating occupation semiempirical configuration interaction wave functions. *J. Phys. Chem. A* 107:3822. doi: 10.1021/jp022468p

Troya, D. (2005). Ab initio and direct quasiclassical-trajectory study of the F + CH_4 → HF + CH_3 reaction. *J. Chem. Phys.* 123:214305. doi: 10.1063/1.2126972

Tuckerman, M. E., Marx, D., and Parrinello, M. (2002). The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature* 417, 925–929. doi: 10.1038/nature00797

Worth, G. A., Beck, M. H., Jäckle, A., and Meyer, H.-D. (2016). *The MCTDH Package, H.-D. Meyer, Version 8.4.12.* Available online at: http://mctdh.uni-hd.de/

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Low-Rank Sum-of-Products Finite-Basis-Representation (SOP-FBR) of Potential Energy Surfaces

Ramón L. Panadés-Barrueta[1] and Daniel Peláez[2,*]

[1]*Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM),*
Université Lille 1, Villeneuve d'Ascq cedex, France.
[2]*Institut des Sciences Moléculaires d'Orsay (ISMO) - UMR 8214. Bât. 520,*
Université Paris-Saclay, 91405 Orsay Cedex

(Dated: November 17, 2020)

Abstract

The SOP-FBR approach for the automated multidimensional fit of potential energy surfaces (PES) is presented. In its current implementation the method yields a PES in the so-called Tucker sum-of-products form but it is not restricted to this specific *ansatz*. The novelty of our algorithm lies in the fact that the fit is performed in terms of a direct product of *Schmidt* basis, also known as *natural potentials*. These encode in a non-trivial way all the physics of the problem and, hence, circumvent the usual extra *ad hoc* and *a posteriori* adjustments (*e.g.* damping functions) of the fitted PES. Moreover, we avoid the intermediate *refitting* stage common to other tensor-decomposition methods typically used in the context of nuclear quantum dynamics. The resulting SOP-FBR PES is analytical and differentiable *ad inifinitum*. Our *ansatz* is fully general and can be used in combination with most (molecular) dynamics codes. In particular, it has been interfaced and extensively tested with the Heidelberg implementation of the Multiconfiguration Time-Dependent Hartree (MCTDH) quantum dynamical software package.

---

[*] Electronic address: `Daniel.Pelaez-Ruiz@universite-paris-saclay.fr`

I. INTRODUCTION

Major advances in the field of quantum dynamics have been connected to the introduction of novel *ansätze* for the efficient representation of multidimensional objects, the wavefunction (WF) being the most distinguished member of this category. Indeed, a major breakthrough took place with the introduction of the Multiconfiguration Time-Dependent Hartree (MCTDH) method[1]. In MCTDH the WF is represented by an efficient two-layer adaptive scheme consisting on a time-dependent nuclear basis set expressed in terms of a time-independent one. This allowed, for the first time, to overcome the traditional *4-atom limit* of the so-called *standard* method when considering quantum dynamical problems using non-analytical potentials. Perhaps, the most relevant change of paradigm arrived with the adoption of a multilayer (tree) formalism epitomized by the Multilayer MCTDH (ML-MCTDH)[2–4] which was subsequently adopted by related methods such as the Gaussian variants of MCTDH.[5,6] These compact and highly flexible multilayer-tree structures have allowed the implementation of very efficient iterative[7] and recursive algorithms.[4,8] More recently, a new generation of promising methods in which the object structures belong to the Tensor-Train family have appeared.[9–12]

A common feature to all these methods is that they are particularly suited or rely exclusively on Hamiltonian operators expressed in a separable form, the so-called sum-of-products (SOP) form, such as the Vibronic Coupling model Hamiltonian. Considering the Kinetic Energy Operator, either it is already in this form or it can be expressed either analytically[13] or numerically[14] in SOP form in a *quasi* automated fashion. On the other hand, the computation of potential energy surfaces (PESs) still poses major challenges. For decades now, major efforts have been devoted to the development of efficient tensor decomposition algorithms to turn these *a priori* unstructured objects (multidimensional PES) into separable form.[15–20] A common feature to all these methods is that they rely on a two-step *decoupled* scheme. After the multidimensional fit of a set of *ab initio* energies, the resulting PES is tensor decomposed to achieve the sought separable form. An exception to this can, in a certain sense, be found in the so-called *on-the-fly* methods in which a more or less local (in configuration space) PES is obtained directly from electronic structure calculations.[21]

In this work, we shall introduce a scheme for the computation of a *global* PESs directly from electronic structure calculations in separable form, more specifically, but by no means restricted to, the Tucker SOP form. The computation of a more or less *global* high-dimensional PESs is a laborious task involving the calculation of energies and possibly properties in a large set of reference configurations, followed by the fit into an analytic expression. The latter will be turned into a *hopefully* efficient code which will be interfaced with or directly implemented in a dynamics software package. Assuming the possibility of numerically converged dynamical calculations, one can conclude that two major factors will contribute to the quality of a simulation: (i) the accuracy and adequacy of the underlying electronic structure calculations (*i.e.* combination of level of theory and electronic basis set); (ii) the errors associated to the fitting procedure which, in turn, depend on the adequacy of the distribution of the reference geometries. In addition to this, as already mentioned above, the use of PES in combination with grid-based quantum dynamical methods, like MCTDH,[1] does greatly benefit from PESs expressed in SOP form. Consequently, an extra tensor decomposition stage with its intrinsic error (see for instance Ref.[16]) should be added to this workflow.

There is a plethora of fitting procedures for PESs available, so we refer the reader to (part of) the original and extensive literature.[22–27] We will, however, turn our attention into methods which, by *ansatz*, directly provide a SOP form (Eq.1). In this category, we can distinguish three types of methods: (i) those providing a local (Taylor) expansion, typically Quartic Force Fields;[28–30] (ii) those using an appropriate Neural Network (NN) approach (epitomized by the seminal work of Manzhos and Carrington[31] and recent applications thereof)[32,33]; and (iii) those directly fitting into a separable form (see for instance,[9,20,34–38]) including specific NN *ansätze* which have been shown to be directly expressible in separable form.[31,32] It is worth highlighting the efficient combination of a Smolyak quadrature with non-direct product basis achieved by Ávila and Carrington.[37] Perhaps, a more particular case of this last category is given by our recent method for global multidimensional fit of PES[20] in which an intermediate and inexpensive (yet chemically accurate) representation of the electronic Hamiltonian (a reparametrised semiempirical Hamiltonian) can be directly interfaced to a tensor decomposition method such as the Multigrid POTFIT algorithm,[16] hence the name of the method Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF).[20] In this regard, the physically meaningful representation of the SRP Hamiltonian allows for the use a fairly small amount of reference points while, at the same time, it provides a physically correct PES.[20] The quality of this SRP PES has been assessed through the determination of accurate vibrational eigenstate and the propagation of a wavepacket using the MCTDH software.

As briefly sketched above, in (molecular) quantum dynamics (*aka* multidimensional wavepacket propagations) arguably the main bottleneck is the computation and representation the potential energy surface (PES) in a suitable form.[16,20] Turning our attention to MCTDH, a potentially numerically exact method, an already existing PES (in the form of a subroutine or an appropriate analytical function) has to be preprocessed (tensor decomposed) in order to achieve the so-called sum-of-products (SOP) form representation:

$$V = \sum_r c_r \prod_\kappa v_r^{(\kappa)} \tag{1}$$

where the $v^{(\kappa)}$ are mono- (or low-dimensional) basis functions expressed on a grid (see Eq.3). There are plenty different types of SOP forms (Tucker form, CP format, Tensor-Train decomposition or even the conceptually simple Taylor expansion, among others). In this work, we shall focus on the so-called Tucker form (Eq.3) which is the most commonly used in MCTDH. The reader is referred to the abundant literature concerning the advantages of using SOP in general (curse of dimensionality, quadrature) and in MCTDH in particular (see for instance the *classical* MCTDH Review).[1] It should be stressed that the subsequent discussion is fully general and thus applicable to any type of multidimensional functions. However, without any loss of generality, we shall make reference to aspects related to MCTDH quantum dynamics.

The different *variational* Tucker decomposition schemes proposed over the years[15–17] have been shown to be efficient, accurate, and, most relevantly, they have been shown to allow a systematic control of the error owing to their variational character. In all of these methods, the main goal is to compute an accurate representation of (or an approximation to) potential density matrices from which grid-based *Schmidt* basis functions (*aka* single-particle potentials, natural potentials) can be extracted as eigenvectors. The main difference in these methods is simply how they arrive to these density matrices. For instance, in the pioneering POTFIT algorithm,[15] this was achieved by computing the potential on the whole grid (thence its power but limited applicability). On the other hand, in Multigrid POTFIT,[16] this was achieved using a series of nested grids, whereas in the equivalent Monte-Carlo POTFIT,[17] the potential density matrices were stochastically approximated. For the sake of completeness, we should at this point mention the powerful strategies proposed by Otto[18,19] and Schröder[39] in which the sought separable form is obtained through a series of (recursive) High-Order Singular Value Decompositions[40] and a CANDECOMP[41,42] formalism, respectively. In any case, the common requirement of all these approaches is the *preexistence* of a PES in analytical form. Consequently, they can be seen as *refitting* methods.

In this work, we tackle this very problem under the assumption that it is possible to achieve a SOP form directly from high-level electronic structure calculations. As it will be shown, this makes the computation of a global PES more efficient and, what is more relevant, less prone to errors since this scheme reduces the necessary number of stages to this end. As such, our sum-of-products Finite-Basis-Representation (SOP-FBR) approach, constitutes an alternative to some NN-based methods.[31,32] The idea behind SOP-FBR is simple. Given a low-rank Tucker expansion on the grid, we replace the grid-based basis functions (*i.e.* natural potentials) by an expansion in terms a set of functions. As a result, a global SOP-FBR PES is simply an analytical expression in Tucker form (but not restricted to this specific form). The SOP-FBR parameters are obtained by optimisation using an appropriate number of reference *ab initio* energies. A subtle difference between SOP-FBR and any traditional multivariate regression method is the fact that we do not directly fit our PES to, say, a set of physically-oriented polynomials (*one-layer*). In SOP-FBR, we optimize a certain number of configuration interaction coefficients together with the expansion coefficients necessary to expand optimal *Schmidt* basis (*i.e.* natural potentials) in terms of a given set of functions (*e.g.* orthonormal polynomials). Consequently, SOP-FBR relies on a *two-layer* scheme, analogous (no time-dependence though) to the MCTDH one in which the WF is represented by a time-dependent basis (upper layer) and an underlying time-independent one (lower layer). This scheme provides us with a highly flexible model relying on the *optimal* properties of a multidimensional Schmidt decomposition (see Appendix D in[1] and Appendix B in[16]).

II. THEORY

The *traditional recipe* for the preparation of a PES suited for quantum dynamical calculations typically implies three/four steps:

1.- the judicious choice of reference geometries;

2.- the computation of accurate reference energies;

3.- an analytical fit; and

4.- (in certain cases) a tensor decomposition.

With respect to step 1, we have recently shown that it is possible to automatically obtain an *optimal* set of pivotal structures (based on the stationary points) from which systematically generate a set of reference geometries.[20] On the other hand, from step 2, all stages imply a direct loss of quality in the resulting PES owing to: (i) the intrinsic accuracy of the selected level of theory; (ii) the accuracy of the fit (dramatically dependent on the dimension of the system); and (iii) the quality of the decomposition, which is directly related to the size of the resulting expansion and the quality of the basis functions. To understand the last point, one should simply consider the vast configuration space associated to the parametrical representation of the chosen tensor structure. In this respect, multiple minima are to be expected owing to the usual non-unique character of a, for instance, Tucker decomposition.[43] Since our approach involves (but it is by no means limited to) Tucker expansions, in the following we shall use the terms SOP and Tucker exchangeably.

We propose to combine stages 3 and 4 in a single step thus, potentially (and effectively) reducing the number of steps involved in a PES fit together with the concomitant errors. Obviously, electronic structure calculations are unavoidable, so far, but owing to the computer power available (*e.g.* supercomputer centers), this is not anymore a *bottleneck*. The basic idea is that it is possible to arrive to an *optimal* SOP (Tucker) expression (*i.e.* to the optimal parameters defining it) either stochastically or by using some kind of, possibly global, optimization algorithm. In other words, a guess of initial SOP parameters will be optimised according to a suitably defined target function. In our case, this will be the root-mean square error (RMSE, $\chi(\mathbf{Q})$) between the reference function (*exact*) and our current approximation to it (*app*):

$$\chi(\mathbf{Q}) = \sqrt{\frac{\sum_i^N \Delta V_i^2(Q_i)}{N}} \tag{2}$$

where $\Delta V_i = V^{exact}(Q_i) - V^{app}(Q_i)$ is the difference between the exact potential and the approximated one at the i-th reference geometry. Ideally, the approximation should be such that when computing the RMSE using another (possibly denser) set of geometries ($\{\mathbf{Q}'\}$, validation set) different from the reference one, the RMSE remains reasonably close to the reference one: $|\chi(\mathbf{Q}) - \chi(\mathbf{Q}')| < \epsilon$ with ϵ small.

At this point, it should be emphasised that in contrast to grid-based tensor decomposition methods, in SOP-FBR we do not necessarily rely on a grid. The so-called single-particle (or natural) potentials of POTFIT (or related methods) are replaced by analytical functions which we refer to as *Schmidt* functions. The latter are expanded in a given polynomial series (Chebyshev in our case). This results in a set of parameters (*core* tensor and expansion coefficients) which are optimised in order to reproduce a given level of theory. To obtain a suitable guess for the optimisation process, we employ an initial (grid-based) Tucker decomposition using, for instance, the Higher Order Orthogonal Iteration (HOOI) algorithm as implemented in the `TensorLy` library. It should be stressed that this initial decomposition cannot be directly performed on the high-level PES (e.g. high-level ab initio) owing to the huge number of grid points needed and the necessary CPU time for this. This has been extensively discussed in our previous work on SRP-MGPF.[20]

A low-rank f-dimensional grid-based expansion of the PES in Tucker form reads:

$$V^{app}_{i_1,i_2,\ldots,i_f} = \sum_{j_1}^{m_1} \cdots \sum_{j_f}^{m_f} C_{j_1,\ldots,j_f} \prod_{\kappa}^{f} v^{(\kappa)}_{i_\kappa j_\kappa} \tag{3}$$

where $C_{j_1,\ldots,j_f}$ is a coefficient belonging to the so-called *core* tensor and $v^{(\kappa)}$ are the *Schmidt* basis functions (factor matrices in the tensor community[44]) expressed on a grid (note the two indices).[15] The latter are referred to as single-particle potentials (SPPs) in the MCTDH literature. This is the *ansatz* commonly used in POTFIT, MGPF or MCPF to represent the PES in SOP form. In this expression, the PES value at a given geometry (V($\mathbf{Q}$), with $\mathbf{Q} \equiv \left(q_{i_1}, \ldots, q_{i_f}\right)$) specified by its grid indices ($(i_1, i_2, \ldots, i_f) \leftrightarrow \left(q_{i_1}, \ldots, q_{i_f}\right)$) is expressed as a sum of products of weighing coefficients ($C_{j_1,\ldots,j_f}$) times a (Hartree) product of one- (or

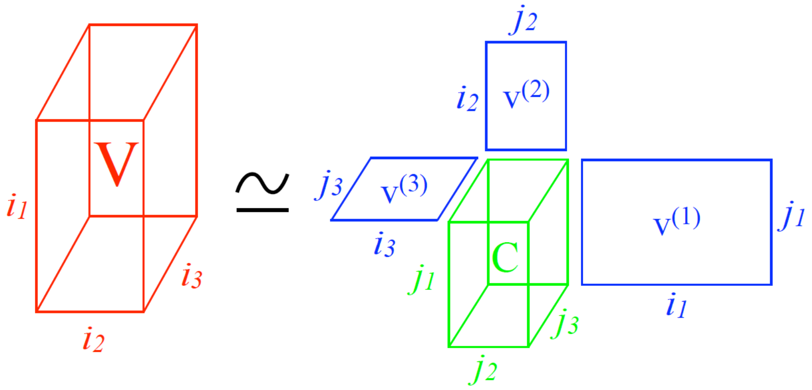

FIG. 1: Graphical representation of a 3D Tucker expansion. See main text for explanation.

low-) dimensional functions on a grid ($v^{(\kappa)}_{i_\kappa j_\kappa}$). Note that by κ we refer to a given degree of freedom (DOF). In Figure 1, we display the graphical representation of a 3-dimensional (3D) Tucker expansion. Note that the *core*-tensor has the same dimensionality than the initial reference tensor of PES on the grid and thus it is affected by the curse of dimensionality at the same rate. With respect to notation, the above Tucker expression can be more compactly written as (see Kolda and Bader's comprehensive review)[43]:

$$V^{app} = [\![\mathcal{C}, \mathbf{v}]\!] \tag{4}$$

To motivate our discussion, we should recall that in the original POTFIT (and related methods), the SPPs are grid-based basis functions adopting the form of matrices. Hence the SPPs merely contain the collection of SPP values at the grid points. Consequently, the existence of implicit *underlying SPP/Schmidt functions* can be exploited in order to increase the density of the grid representation of PES. In MCTDH, this is the so-called `chnpot` procedure in which some kind of interpolating procedure (*e.g.* cubic splines) is used to approximate SPPs values in the regions in between the original grid points.[15] The SPPs or Schmidt functions are smooth except perhaps for the large j ones. To illustrate this, we present in Figure 2 the first 10 SPPs of the water molecule (see details of the PES later in text) on a grid (dots) together with interpolating functions (lines). The SPPs have been obtained through the Higher Order Orthogonal Iteration (HOOI) method as implemented in the `TensorLy` Python library.[45]

As it can be observed, there is a remarkable smoothness in the lower-j basis functions whereas the higher-j SPPs exhibit a highly oscillating (somewhat noisy) and, hence, unde-

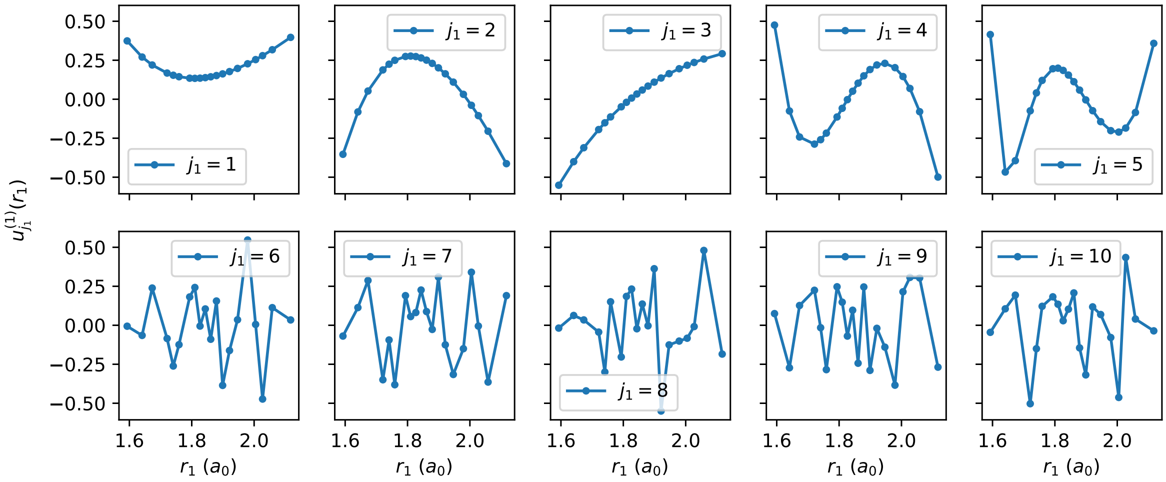

FIG. 2: Representation of the first 10 natural potentials of the PJT2-PES of the H_2O molecule. The dots represent the values of the natural potentials at the DVR grid points.

sirable trend for fitting. At this point, it should be further emphasised that these lower-j Schmidt basis carry the most relevant information of the PES as proven by the fact that grid-based methods such as MGPF or MCPF only require such a small amount of SPPs. The behaviour of the SPPs was already discussed in the seminal POTFIT paper by Jäckle and Meyer[15] as well as by Manzhos and Carrington.[46] Interestingly, this important property has not been fully exploited and lies at the core of our approach. Based on this simple observation, we propose a new low-rank *ansatz* in which we substitute the SOP grid-based expression (Eq.3) by a SOP Finite Basis Representation (SOP-FBR) in terms of some basis functions ($T(q)$):

$$V(q_1,\ldots,q_f) = \sum_{j_1}^{m_1}\cdots\sum_{j_f}^{m_f} C_{j_1,\ldots,j_f}\prod_{\kappa}^{f}(\sum_{\mu}^{t_\kappa} c_{\mu,j_\kappa}^{(\kappa)} T_\mu(q_\kappa)) \tag{5}$$

where the the grid indices (i_κ) in Eq.3 have been replaced by the value of the actual coordinate (q_κ). This expression can be used in any grid since the grid-based SPPs ($v^{(\kappa)}$) are now interpolated through the use of a weighted sum of t_κ T-basis functions:

$$v_j^{(\kappa)}(q_\kappa) = \sum_{\mu}^{t_\kappa} c_{\mu,j_\kappa}^{(\kappa)} T_\mu(q_\kappa) \tag{6}$$

with T assumed to be appropriated for the specific problem. The case we discuss is that of a general molecular PES (multiwell bound region) for which the typical behaviour is that of the SPPs in Figure 2. However, we have observed that the higher-j SPPs for the Coulomb potential present a convoluted sinus type behaviour (work in progress). Throughout this work we have expressed T with Chebyshev polynomials but any other functions are also possible. Consequently, this change in the representation, from grid to FBR, leads to a full analytical global PES, grid-independent and differentiable *ad infinitum.* Indeed, one can immediately realise that n-th order derivatives can be obtained in a straightforward manner. By direct differentiation, one obtains:

$$\frac{\partial^n V(q_1, q_2, \ldots, q_f)}{\partial q_\kappa^n} = \sum_{j_1}^{m_1} \cdots \sum_{j_f}^{m_f} C_{j_1,\ldots,j_f} \left(\sum_{\mu}^{t_\kappa} c_{\mu,j_\kappa}^{(\kappa)} \frac{\partial^n T_\mu(q_\kappa)}{\partial q_\kappa^n} \right) \prod_{\nu \neq \kappa}^{f} \left(\sum_{\mu}^{t_\nu} c_{\mu,j_\nu}^{(\nu)} T_\mu(q_\nu) \right) \tag{7}$$

At this point it is convenient to highlight some analogies between low-rank grid based methods, such as grid-based MGPF and grid-free SOP-FBR. In the original publication[16] and subsequent applications,[47–49] it could be observed that MGPF is able to extract a very compact representation of the PES following a given prescription in the choice of the reference points (see discussion on the sparsity of the *core* tensor in Section V). For instance, in Ref.[48], the infrared spectrum of $H_3O_2^-$ (9D) was computed. The MGPF PES required 1.7 MB of storage. In contrast, the original PES which would have required 10^5 MB. The resulting MGPF PES based spectrum was in good agreement with the reference experimental Ar-vibrational predissociation one. In view of the nice agreement, discrepancies between both spectra could be attributed to the effect of the Ar atom in the vibrational structure of the $H_3O_2^-$ complex. In other words, considering the PES as a multidimensional image, it is possible to compress it using just a few basis functions. The difference in SOP-FBR is only that we achieve the same expansion through an extra parametrization of the SPP (Schmidt functions) and an optimisation. In summary, it is possible to obtain such an $\mathcal{L}^2$-optimal Tucker expansion using numerical methods such as stochastic, global, local search algorithms or a suitable combination thereof.

To finalize this Section, for the MCTDH oriented reader, we would like to point out that the *contracted* mode (ν):[15]

$$D^{(\nu)}_{j_1 \ldots i_\nu \ldots j_f} = \sum_{j_\nu} C_{j_1 \ldots j_\nu \ldots j_f} v^{(\nu)}_{i_\nu j_\nu} \tag{8}$$

can be readily obtained from a SOP-FBR expression through:

$$D^{(\nu)}_{j_1 \ldots i_\nu \ldots j_f} = \sum_{j_\nu} C_{j_1 \ldots j_\nu \ldots j_f} v^{(\nu)}_{j_\nu}(q^{(\nu)}_{i_\nu}) \tag{9}$$

provided that a large enough sampling has been considered for the contracted mode (*cf.* torsion mode in HONO in Section V).

III. RESULTS AND DISCUSSION

In this Section, we shall introduce and discuss the actual implementation of the method as well as the computational and numerical properties of the SOP-FBR PES. Hereafter, the following notation for our SOP-FBR expansions will be used: $[(m_1, t_1), (m_2, t_2), (m_3, t_3), ...]$ representing an f-dimensional *core* tensor possessing $\prod_{\kappa}^{f} m_{\kappa}$ elements. This also indicates that each κ DOF is given by a set of m_{κ} basis functions which, in turn, are expressed in terms of t_{κ} underlying basis, in the present case, Chebyshev polynomials. It should be further emphasized that there is no restriction in the type of basis sets employed, except that it should be adapted to the target potential. The total number of parameters will formally then be: $\prod_{\kappa}^{f} m_{\kappa} + \sum_{\kappa}^{f} m_{\kappa} \cdot t_{\kappa}$. As it will be discussed, we will profit from the huge range of values (over 12 orders of magnitude) among the *core* tensor elements to greatly reduce the number $\prod_{\kappa}^{f} m_{\kappa}$ terms. Finally, in those cases where d DOF are represented by the same amount of basis functions, we shall simply write $[d \times (m, t)]$. Note that the Chebyshev polynomials start at order 0, hence (m, t) implies the use of a Chebyshev polynomial of maximum order $(t - 1)$.

A. Proof of concept: SOP-FBR of a Henon-Heiles potential (3D and 9D)

In order to test our algorithm and assess its quality, we have studied the celebrated Henon-Heiles potential which constitutes a common potential benchmark[4,50] of low-rank character:

$$V = \frac{1}{2}\sum_{\kappa=1}^{f} q_{\kappa}^2 + \lambda \sum_{\kappa=1}^{f-1} \left(q_{\kappa}^2 q_{\kappa+1} - \frac{1}{3} q_{\kappa+1}^3 \right) \qquad \lambda{=}0.111803 \tag{10}$$

In our case, the choice of the anharmonic parameter (λ) has been based on purely *historic* reasons.[4,50] The potential boundaries are [-9, 7] length units.[4] The study of a low-rank expression serves as both a proof of concept for our algorithm as well as a validation for the numerical accuracy of our approach. Indeed, as shown in Reference[16] for the MGPF tensor decomposition algorithm, we should expect SOP-FBR to be able to exactly describe low-rank potentials. In other words, it should be possible to obtain, in a non-algebraic manner, an accurate approximation to the optimal *core* tensor coefficients as well as the values of the Schmidt factor matrices (*aka* single-particle potentials, SPPs) through a *walk* through the configuration space of the SOP-FBR parameters.

First of all, to illustrate the idea of sampling the Tucker configuration space, we have considered the stochastical optimisation of a SOP-FBR ($[3 \times (6, 5)]$) for a 3D Henon-Heiles potential (Eq.10 with $f = 3$) (boundaries as previously indicated). For this we have used the robust Basin-Hopping (BH) global optimisation algorithm[51] as implemented in the `SciPy` library.[52] The initial guess parameters have been obtained through random sampling in the intervals [0., 0.01] and [-1.0,1.0] for the *core* and Chebyshev polynomials coefficients, respectively. A systematic strategy for the computation of this guess will be discussed in Section IV. The obtained RMSE error has been of $2.076 \cdot 10^{-5}$

This result nicely illustrates the *possibility* of strolling along (or guiding our algorithm through) the Tucker configuration space to find minima. In the following, we shall introduce a more educated procedure to generate the set of initial guess parameters and to search for the RMSE minimum.

We continue now with same very same type of potential, Henon-Heiles for 3D (again $[3 \times (6, 5)]$) and for 9D ($[9 \times (6, 7)]$). In this case, we employ a more educated guess consisting in values randomly chosen from a Gaussian distribution around HOOI values (see discussion about the SOP-FBR guess in Section IV A) This time, however, we shall make use of a local optimiser, more specifically, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[53] as implemented in the `SciPy` library.[52]. We do this in a decoupled manner. We first optimize the basis functions (Chebyshev) coefficients (leaving the *core* tensor fixed) and then we optimise the *core* tensor elements (leaving the Chebyshev coefficients fixed). The simple idea behind this is the reduction of computational cost when dealing with large dimensional problems. In Table I, we present the RMSE errors obtained for these two optimisations. As it can be observed, in both cases the error can be assimilated to purely numerical noise thus validating our approach and, more relevantly at this stage, our algorithm.

To further analyse the characteristics of our SOP-FBRs, in Figure 3, we present contour plots (x=0) of the HH 3D reference potential, the SOP-FBR approximation, and the $\mathcal{L}_1$

TABLE I: RMSE of the SOP-FBR expansions of the fD Henon-Heiles potentials (f=3,9). First column displays the structure of the SOP-FBR and, implicitly, its dimensionality. Second and third columns present the length and corresponding energy intervals, respectively. Fourth column, presents the RMSE in potential units. Validation sets comprising 30^3 (3D) and 8^9 (9D) points have been used to compute the error.

SOP-FBR	**Boundaries**	**ΔE**	**RMSE**
$[3 \times (6,5)]$	[-9, 7]	$[-1.83, 165.67]$	$1.241 \cdot 10^{-13}$
$[9 \times (6,7)]$	[-9, 7]	$[-81.00, 466.07]$	$5.165 \cdot 10^{-12}$

difference between them. As it can be seen, the error (10^{-13}) is homogeneously distributed along all energy domains. This can also be observed for the 9D case in the graph of the energy correlation comparing the reference potential and our SOP-FBR approximation (Fig.4), in which latter follows closely the former without any appreciable divergence within the whole energy interval. The 3D(9D) SOP-FBR optimisation required 0.27(0.35) hours of CPU using 24 processors.

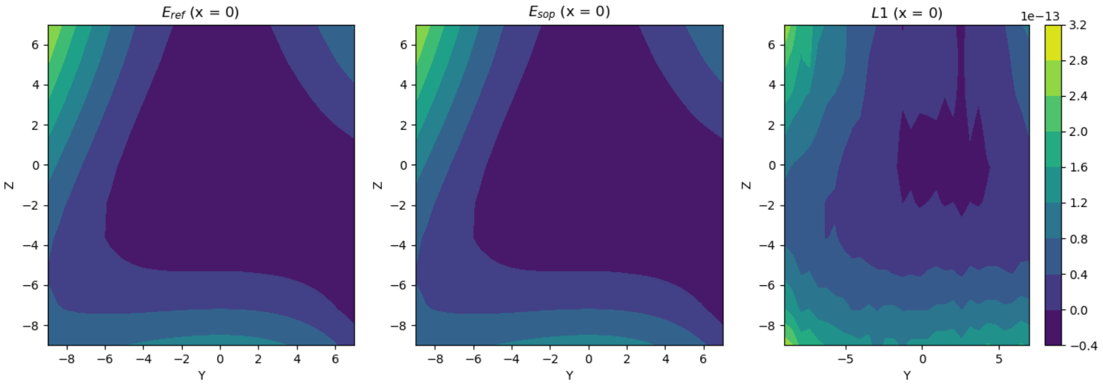

FIG. 3: Contour plots (for x=0) showing the reference potential and the SOP-FBR approximation for the 3D Henon-Heiles potential.

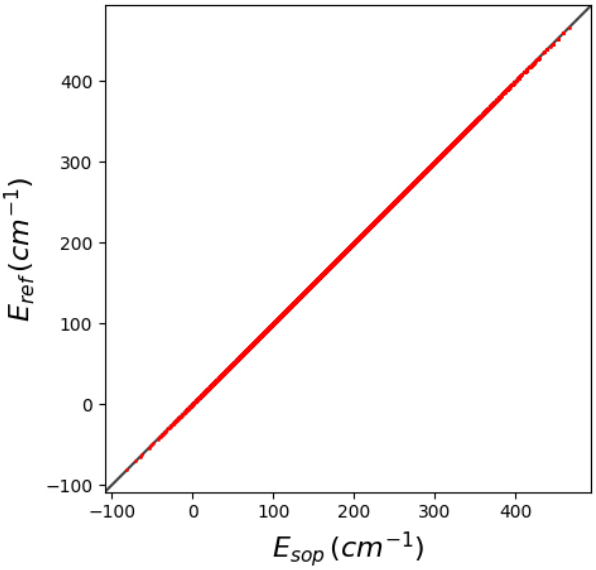

FIG. 4: Error correlation between the reference potential (y-axis) and the SOP-FBR approximation (x-axis) for the 9D Henon-Heiles potential.

IV. SOP-FBR OF A LOW-DIMENSIONALITY PES: H_2O (3D)

In the following section, we shall tackle a *realistic* 3D problem, the water (H_2O) molecule. We have used the Polyanski-Jensen-Tennyson $H_2^{16}O$ spectroscopically refined PES known as PJT2.[54]

A. Generation of the guess parameters

In contrast to our previous model case-study, the HH potential, when dealing with *ab initio* PES, it would be desirable, for the sake of efficiency, to start our calculation with a physically oriented guess. Our strategy for this has consisted in the use of an inexpensive SRP PES computed *a priori* following our recently published method[20] using our own software (SRPTucker[55]) which is interfaced with the MOPAC semiempirical software.[56] This SRP PES can be readily decomposed on a grid using the High Order Orthogonal Iteration (HOOI) algorithm[43] thus obtaining an initial *core* tensor and its associated Schmidt basis functions (SPPs). In this particular SRP PES, we utilised a grid of 5^3 reference geometries (a subset of the one that we use for the SOP-FBR optimisation) and made use of the

Parametric Model 3 (PM3) Hamiltonian.[57]

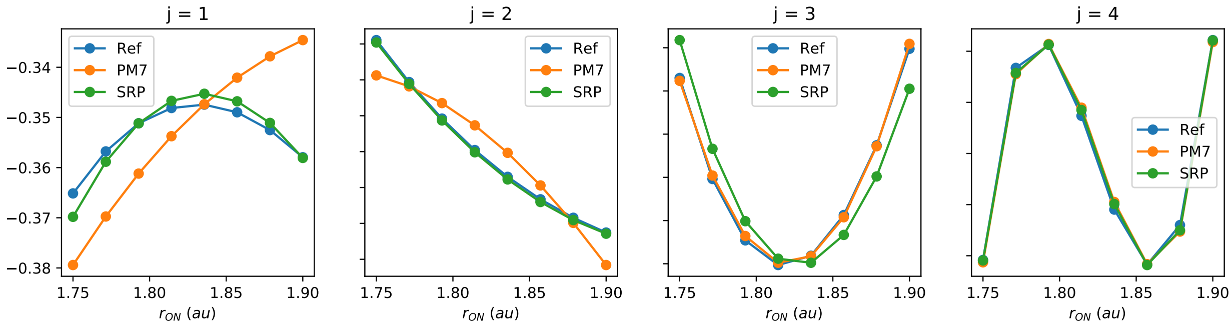

FIG. 5: First five SPPs of the angular DOF (θ_{HOH}) for the H_2O molecule obtained by HOOI Tucker decomposition using (i) the reference potential PJT2[54] (blue) (ii) the PM3 Hamiltonian (orange); and (iii) the resulting SRP (green).

Figure 5 displays a comparison of the first five SRP SPPs for the *ab initio* reference surface, the PM3 surface, and the resulting SRP. As it can be observed, the guess SRP SPPs are in much better agreement with the *exact* ones than the original PM3. From these plots, we notice that the SRP constitutes a better source of guess parameters than the PM3 method. Despite the fact that we are dealing with guess (initial) Schmidt functions (SPPs), any improvement in the shape of the semiempirical SPPs will be turned into a saving in computational effort when searching for the optimal SOP-FBR solution. On the other hand, the most relevant elements of the *core* tensor are very similar for the three methods (see Figure 6).

For the sake of efficiency, prior to any optimisation, we have imposed a symmetrization by *ansatz* to the SOP-FBR expression: (i) the SPPs for both OH stretching modes have been forced to be equal, and (ii) the *core* tensor elements satisfy the $C_{ijk} = C_{jik}$ condition, where ijk are indices referring to Schmidt basis functions (SPPs) and we have implicitly assumed the $[r_1 \equiv \nu_{OH1}, r_2 \equiv \nu_{OH2}, \theta_{HOH}]$ mode ordering. It should be noticed that if no such restriction is imposed, the obtained SPPs are nevertheless reasonably close to the symmetrized ones.

The SOP-FBR optimisation has been performed in a *decoupled* manner (see discussion in Section III A) using a combination of global and local algorithms, namely Multi-Level

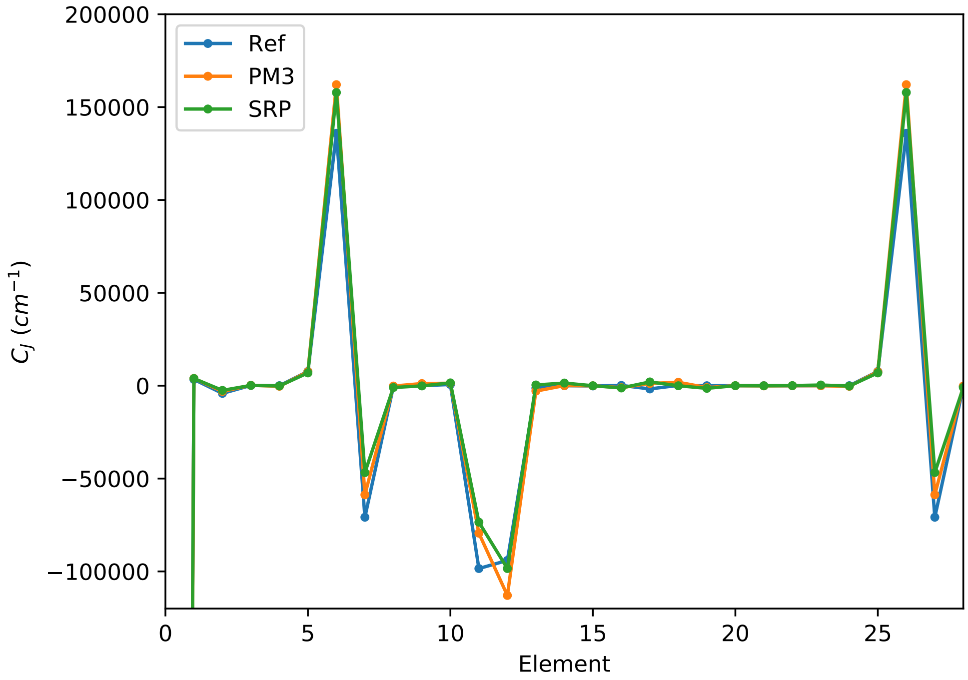

FIG. 6: Comparison of the first elements of the *core* tensor of the H_2O molecule obtained by HOOI Tucker decomposition of (i) the reference potential[54] (ii) the PM3 hamiltonian and (iii) the corresponding SRP. Since the *core* tensor happens to be very sparse, for higher index values the differences are even less appreciable.

Single-Linkage (MLSL)[58] for the former and Bound Optimization BY Quadratic Approximation (BOBYQA) for the latter.[59] This type of approaches have been shown to be very well adapted to the reparametrisation of semiempirical Hamitonians.[20] For this optimisation, we used a set (grid) of [15, 15, 15] structures (mode ordering as above) and the SOP-FBR structure was $[3 \times (5, 11)]$ thus generating a parameter space of size $5^3 + 3 \cdot 5 \cdot 11 = 290$. The reference points were generated through homogeneous sampling of normal distributions centred at the water molecule equilibrium parameters. We used for this the `truncnorm` method of SciPy.[52] The standard deviation was set up in order to cover the 1D intervals taken from the MCTDH model inputs, which are available in the Heidelberg MCTDH software package (see Table III). It should be stressed that equilibrium geometrical parameters can be computed through the use of automated methods (see for instance Ref.[60] and references therein).

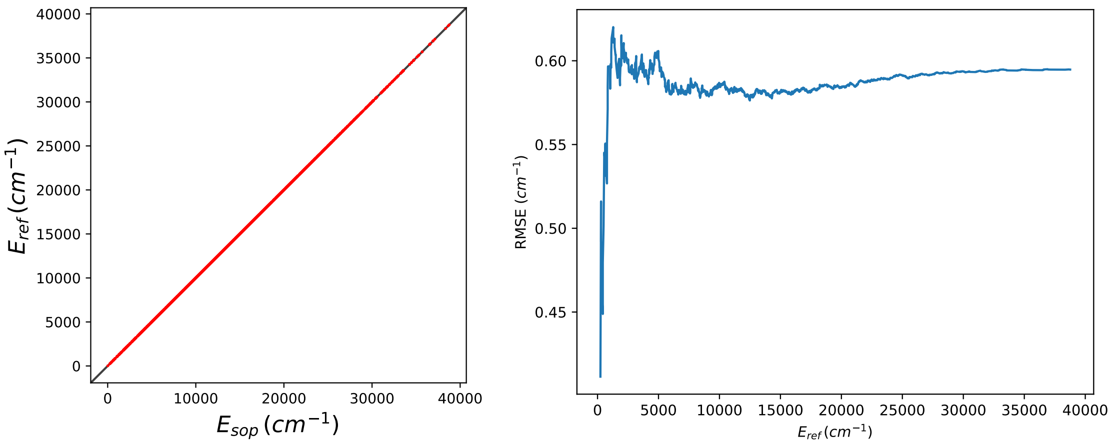

FIG. 7: Correlation plot between the reference *ab initio* energies and corresponding SOP-FBR energies, and cumulative RMSE values obtained with a reference tensor of 15 × 15 × 15 points.

The resulting SOP-FBR RMSE was of 0.686 cm^{-1}, a value which was nicely preserved (0.703 cm^{-1}) when considering a validation set of 10^6 points, different from the reference ones. It should be noted that similar errors have been obtained with larger sets, all converging to the reference value (HOOI decomposition) of RMSE $= 0.673\ cm^{-1}$. The quality of the resulting SOP-FBR PES has been assessed in several different ways. In Figure 7, we present a graphical analysis of its quality in the form of a correlation plot between the reference and SOP-FBR energies (left panel) and the cumulative RMSE (right panel). As it can be observed, the error is equally distributed along the whole energy interval [0, 40000] cm^{-1}. Moreover, we have computed the minimum of the PES using the steepest descent method. The SOP-FBR and reference values (see Table II) agree up to the fourth decimal place.

TABLE II: Minimum geometrical parameters for the reference PES and the SOP-FBR PES obtained with the steepest descent algorithm. Distances are in atomic units and angles are in radians.

PES	$\mathbf{r_1}$	$\mathbf{r_2}$	$\boldsymbol{\theta}$
Ref.	1.8102	1.8102	1.8238
SOP-FBR	1.8102	1.8102	1.8240

As final test of the quality of the SOP-FBR PES, we have computed the Zero Point

Energy (ZPE) and lowest vibrational eigenvalues with the MCTDH algorithm. For this problem, we have considered the Discrete Variable Representation (DVR) shown in Table III.

TABLE III: Definition of the MCTDH DVR primitive grid for the water molecule. Stretching motions (ν_{OH}) are represented by r_i ($i = 1, 2$) and the HOH bending by θ. The DVR basis functions are sinus (sin) for the former and Legendre (Leg) DVR for the latter. N is the number of primitive (DVR) grid points. The range represents the first and last grid points in atomic units for the distances and in radians for the angle. The contracted mode is the angle.[16]

DOF	**DVR**	**N**	**Range**
r_1	sin	34	[1.4500, 2.4500]
r_2	sin	34	[1.4500, 2.4500]
θ	Leg	50	[1.1400, 3.1416]

Our results are in very nice agreement with the reference PES. Our ZPE is of 4660.225 cm^{-1}, a value which nicely compares to the reference one of 4660.124 cm^{-1}. The remaining vibrational eigenvalues (relative to the ZPE) are presented in Table IV. In all cases, the differences between the SOP-FBR and the reference values is below 1 cm^{-1}.

V. SIX DIMENSIONAL CASE: HONO ISOMERIZATION

As final example, we study now the HONO (6D) PES in the *cis-trans* isomerization region which has become a benchmark for these type of methodological studies.[9,20,32] In this case the reference PES is given by the model CCSD(T)/cc-pVQZ-quality analytical PES of Richter *et al.*[61] The structure of our 6D SOP-FBR PES is given by $[5 \times (5, 7), (12, 12)]$, that is, 5 Schmidt basis (SPPs) represented by 7 Chebyshev polynomials each, for all DOF except the torsion which necessitates a somewhat larger basis (*i.e.* sampling) of 12 Schmidt basis (SPPs) represented by 12 Chebyshev polynomials each (up to 11th order).

For the SOP-FBR (decoupled) optimisation, we have used the BFGS algorithm in combination with a set of 37500 reference geometries, mapped onto a [5, 5, 5, 5, 5, 12] Cartesian product grid of reference energies. The DOF ordering is $[r_{OH}, r_{NO}, u_2, r_{ON}, u_1, \phi]$ (see Table

TABLE IV: Comparison of the 20 lowest vibrational eigenvalues of H_2O for the original PES and the SOP-FBR one.

State	**Eigenenergies ($\mathbf{cm^{-1}}$)**	
	SOP-FBR	**Reference**
0	0.	0.
1	1593.261	1593.624
2	3151.076	3151.074
3	3722.318	3722.315
4	3820.802	3820.786
5	4670.742	4670.669
6	5297.252	5297.631
7	5394.652	5395.063
8	6150.238	6150.351
9	6839.599	6839.602
10	6938.042	6938.046
11	7452.389	7452.381
12	7572.758	7572.731
13	7584.352	7584.555
14	7644.830	7644.798
15	8345.769	8345.688
16	8450.176	8450.059
17	8949.339	8949.394
18	9010.874	9011.259
19	9136.455	9136.860
20	9204.255	9204.683

V). The intervals of definition of the different DOF can be found in Table V). For the sake of comparison, we have defined $u_i = \cos\theta_i$ (i=1,2) as in previous studies.[20,61]

Our strategy to initialize our SOP-FBR optimisation has consisted again in generating a guess *core* tensor and the associated set of SPPs using the HOOI decomposition (same

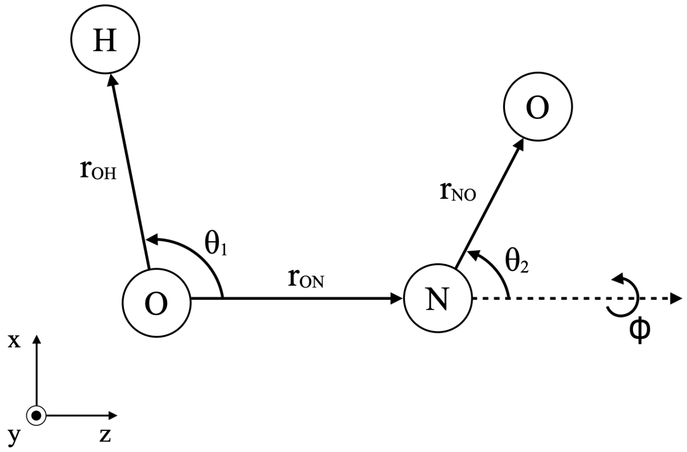

FIG. 8: Definition of the internal coordinates of HONO.

TABLE V: Definition of the MCTDH primitive grid for HONO (6D). Discrete Variable Representation (DVR) basis functions are: HO, harmonic oscillator (Hermite) and cos, cosine functions. N is the number of primitive (DVR) grid points. The range in each case is given the first and last grid points in atomic units for the distances and ϕ is the torsion angle in radians. Cosines of the valence angles have been defined: u_i = cos θ_i (i=1,2). The DOF representation is given in Figure 8. Physical modes have been combined into logical ones as follows: [ϕ], [r_{OH}] [u_2, r_{ON}], [u_1, r_{NO}]. The first mode (ϕ) has been contracted (see Section IIIB in Ref.[16]).

DOF	**DVR**	**N**	**Range**
r_{OH}	HO	18	[1.30, 2.45]
r_{NO}	HO	13	[1.90, 2.60]
u_2	HO	13	[-0.65, -0.10]
r_{ON}	HO	16	[2.10, 3.25]
u_1	HO	18	[-0.65, 0.25]
ϕ	cos	32	[0, $2\pi/2$]

structure of our target SOP-FBR) using a SRP PES. For the latter, we have chosen the so-called $\{\zeta_{1084}\}$ set of SRP-parameters which we generated for the title system and which showed an accurate vibrational structure (see Ref.[20] for more details). In Figure 9, we display a comparison of the first four SPPs for selected DOF (a distance, an angle, and the torsion), r_{NO}, u_1 and ϕ, respectively. As in the case of water, the SRP SPPs are closer to the reference ones, in turn, leading to a higher efficiency in the SOP-FBR optimisation.

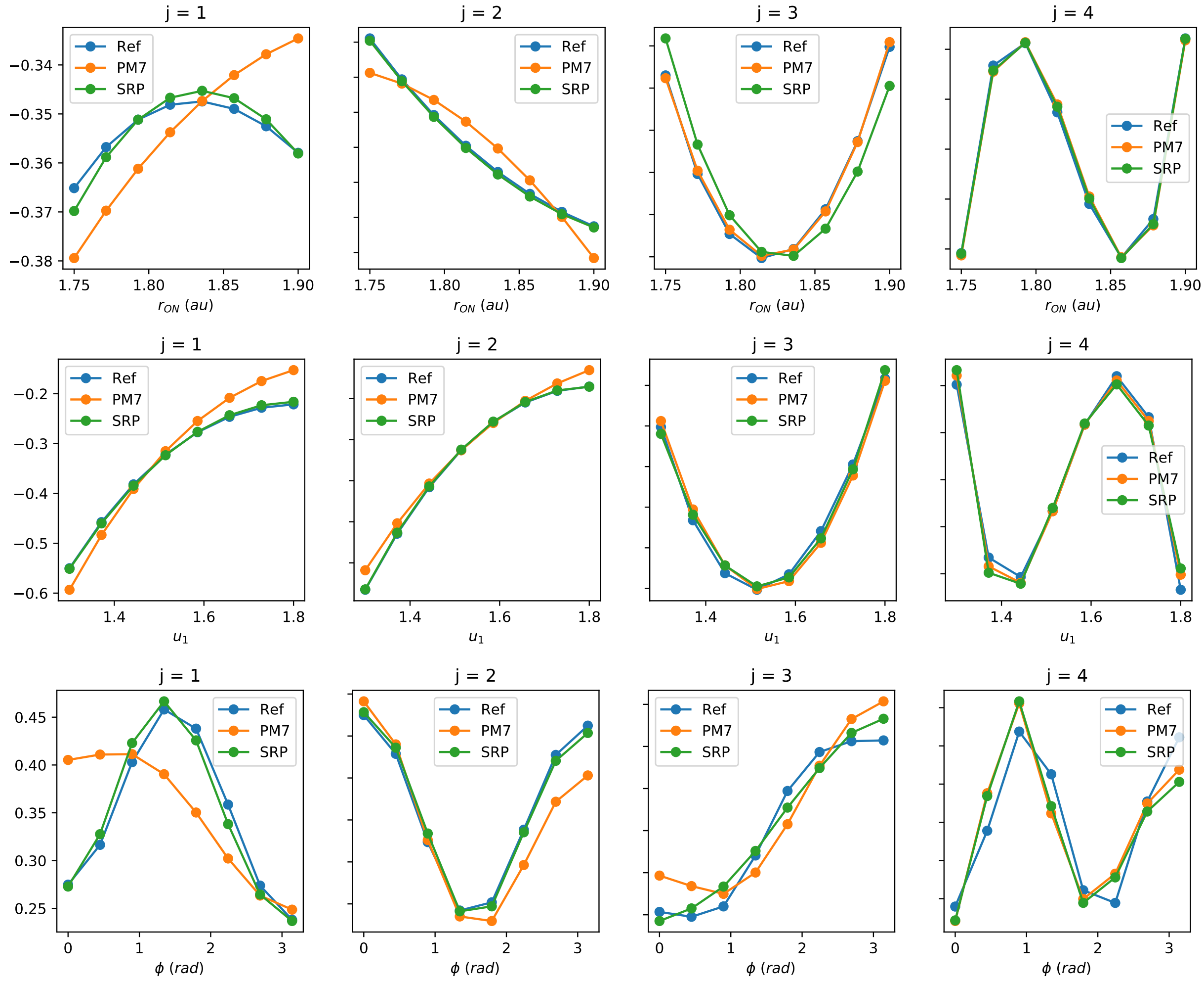

FIG. 9: First four SPPs for the r_{NO}, u_1, and ϕ DOF (see Figure 8 and Table V) in the HONO molecule obtained by HOOI Tucker decomposition using (i) the reference potential by Richter *et al.*[61] (ii) the Parametric Model 7 (PM7) Hamiltonian[57] and (iii) the resulting SRP.

At this point, we make use of the particular distribution of values within the *core* tensor. The latter could be understood to play a role analogous to that of configuration interaction coefficients in electronic structure. First of all, it should be noticed that these values span a huge energy interval from 10^7 down to 10^{-6} (values in cm−1). The maximum corresponds to 22646896 cm−1 and 6940 cm−1 is the mean value. Second, a large fraction of the total number of coefficients lies between $[10^{-2}, 10^{-3}]$ and almost the half of them lie below 10^{-2}. Considering this particular distribution and this huge span of values, we can safely consider consider the *core* tensor as being *effectively* sparse. In other words, the relative weight of a large fraction of the configurations is negligible. We can profit from this fact and assign zero value to all the elements below a certain threshold. This, in turn, implies a dramatic reduction in the dimension of the SOP-FBR configuration space, thus leading to a higher efficiency and a more compact SOP-FBR final expression.

TABLE VI: Analysis of the [5, 5, 5, 5, 5, 12] *core* tensor values for HONO (see main text for the discussion). The total number of tensor elements is N=37500, n is the number of *core* elements below a certain threshold value (Thrs) in cm^{-1}. In the last column, we define an *effective* sparsity as ((N-n)/N) in percentage.

Thrs	**n**	**N-n**	**(N-n)/N**
10^{+7}	37499	1	0.00
10^{+6}	37495	5	0.00
10^{+5}	37409	91	0.00
10^{+4}	37000	500	1.33
10^{+3}	36165	1335	3.56
10^{+2}	34834	2666	7.11
10^{+1}	33105	4395	11.72
10^{-0}	31225	6275	16.73
10^{-1}	28661	8839	23.57
10^{-2}	20041	17459	46.56
10^{-3}	3650	33850	90.27
10^{-4}	402	37098	98.93
10^{-5}	36	37464	99.90
10^{-6}	5	37495	99.99

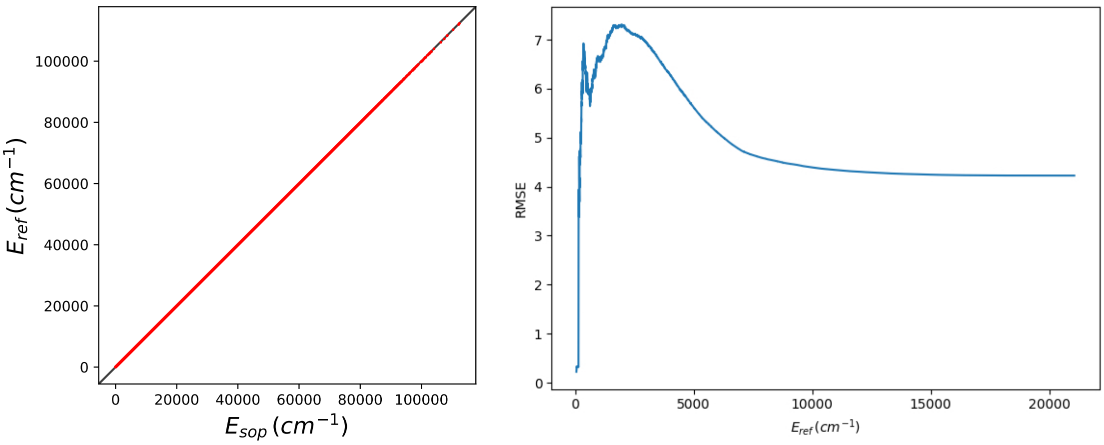

FIG. 10: Correlation plot between the reference *ab initio* energies and corresponding SOP-FBR energies, and cumulative RMSE values.

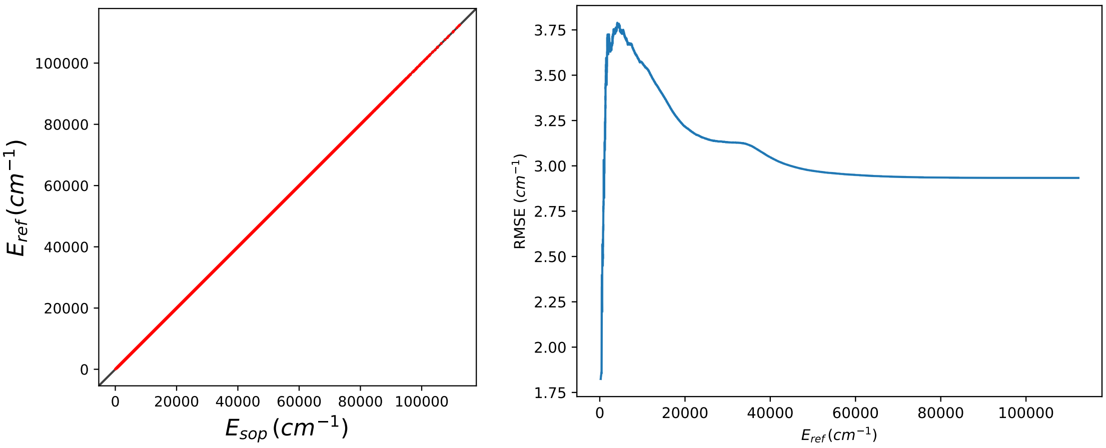

FIG. 11: Correlation plot between the reference *ab initio* energies and corresponding HOOI energies, and cumulative RMSE values.

To carefully assess the quality of our SOP-FBR PES, we have used a validation set consisting on 10^6 points generated in the form of a grid within the same boundaries as our reference but much denser. In addition to this, in order to assess/identify possible errors associated to the SOP-FBR form, we have performed an extra set of calculations using directly the HOOI decomposition of the original PES and having it transformed into SOP-FBR form (Chebyshev expansion of the basis).

The SOP-FBR RMSE was 7.22 cm^{-1}, a satisfactory value somewhat larger than the one obtained for HOOI (2.93 cm^{-1}). The former value should be understood in the context of a direct fit from *ab initio* points, that is, it should be compared to the RMSE obtained for the reference PES of about 5 cm^{-1}.[61] The reliability of the SOP-FBR has been also assessed with a validation set consisting in 10^6 points for which we obtained an RMSE of 11.85 cm^{-1}. Analogously, the same calculation for the HOOI yielded a value of 4.20 cm^{-1}. It should be noticed that the ratio between the RMSEs in the validation and reference sets is about 1.5 for both SOP-FBR and HOOI.

In Figure 10, we provide a graphical representation for these calculations. It can be observed that for the whole range of fitted energies [0, 100000] cm^{-1} a nice agreement between the reference and SOP-FBR values is obtained. At this point, it should highlighted that we have tackled a range of energies ten times higher than previous studies dealing with the same system.[32] The SOP-FBR optimisation required 0.37 hours of CPU using 12 processors.

As an extra test of quality, we have also characterised the stationary points of the SOP-FBR PES. For this, we have computed the *cis* and *trans* minima as well as the corresponding transition state using a steepest descent algorithm in reduced dimensionality: we have enforced the torsion angle at the corresponding reference value, optimise the remaining coordinates and assess the resulting gradient. In all cases, the maximum value of the gradient did not exceed 10^{-7} units. The resulting geometrical parameters are presented in Table VII. As in the case of the water molecule, we obtain values that agree up to the fourth decimal place for both distances and angles.

The reliability of our SOP-FBR can be further illustrated by noticing its *variational* behaviour. In Table VIII, we present the RMSE computed on the reference primitive grid (Table V) for trimmed SOP-FBR expansions (*cf.* td-MGPF[16]) As it can be observed, the results compare well (and even improve) those of POTFIT (from Table IV in Ref.[16]) thus confirming the possibility of numerically *strolling* towards different (and sometimes more favourable) minima than POTFIT (or HOOI) in Tucker configuration space.

TABLE VII: Comparison of the coordinate values of the stationary points (*trans*- and *cis*-minima and *cis-trans* transition state, TS) for the reference PES and the SOP-FBR PES. Geometries obtained with the steepest descent algorithm by imposing the reference torsion angle. In all cases, the maximum component of the gradient (in absolute value) lies below 10^{-7} units (all values are in fact between 10^{-7} and 10^{-10}). Distances are given in atomic units and angles in radians.

	Coordinate					
PES	$\mathbf{r_{NO}}$	$\mathbf{r_{ON}}$	$\mathbf{r_{OH}}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\phi}$
			trans			
Ref.	2.6965	1.8224	2.2129	1.7773	1.9312	3.1416
SOP-FBR	2.6967	1.8231	2.2133	1.7777	1.9315	3.1416
			cis			
Ref.	2.6314	1.8411	2.2369	1.8218	1.9746	0.
SOP-FBR	2.6305	1.8412	2.2374	1.8224	1.9753	0.
			TS			
Ref.	2.8475	1.8178	2.2008	1.7575	1.9290	1.5080
SOP-FBR	2.8475	1.8179	2.2009	1.7578	1.9288	1.5080

TABLE VIII: Analysis of the RMSE for trimmed SOP-FBR (FBR) expansions starting from the [5, 5, 5, 5, 5, 12] (see main text for the discussion). The RMSE units are meV for the sake of comparison with POTFIT (PF) results obtained from Ref.[16].

		Error	
SPPs	**Terms**	**FBR**	**PF**
$[3, 3, 3, 3, 3, c]$	243	20.84	131.68
$[4, 4, 3, 4, 4, c]$	768	5.05	34.06
$[4, 4, 4, 4, 4, c]$	1024	3.39	21.46
$[5, 5, 5, 5, 5, c]$	3125	1.93	0.75

As definitive test of the quality of our SOP-FBR PES, we have computed its ZPE and the 20 lowest vibrationally excited eigenvalues. With respect to the ZPE, our value is of

4366.7 cm^{-1} which nicely compares to the reference one of 4367.7 cm^{-1} and fully agrees with the HOOI one (4365.5 cm^{-1}). In Table IX, we provide a comparison of the results for SOP-FBR, HOOI and the reference PES (exact). As it can be observed, there is a nice agreement among the three sets. There is slight misestimate (common to both SOP-FBR and HOOI) of certain eigenvalues associated to the *cis* well,[61] whose origin can safely attributed to the sampling of this particular region.

TABLE IX: Comparison of the 20 lowest vibrational eigenvalues of HONO for the original PES, the HOOI, and the SOP-FBR one. The first column indicates the label of the state and its character: c, *cis*-well and t, *trans*-well.

State	**Eigenenergies ($\mathbf{cm^{-1}}$)**		
	SOP-FBR	**HOOI**	**Reference**
0	0.0	0.0	0.0
1c	91.2	98.5	94.1
2t	600.7	600.6	600.8
3c	707.8	715.0	710.7
4t	795.7	795.7	795.9
5c	941.2	948.5	944.1
6t	1055.3	1055.2	1055.4
7t	1187.8	1187.6	1188.1
8t	1264.5	1264.4	1264.9
9c	1300.1	1310.4	1306.6
10c	1309.8	1317.0	1312.8
11t	1385.1	1385.0	1385.3
12c	1401.4	1408.8	1404.8
13c	1544.8	1552.9	1547.9
14t	1574.5	1574.6	1574.9
15t	1640.7	1640.4	1640.9
16t	1689.7	1689.8	1689.9
17c	1723.0	1730.1	1726.0
18t	1762.4	1762.1	1762.4
19c	1776.6	1783.7	1779.7
20t	1828.7	1828.5	1829.0

VI. DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

VII. CONCLUSIONS

We have presented sum-of-products Finite Basis Representation (SOP-FBR), an algorithm that yields a PES in Tucker form directly from a set of reference *ab initio* points. In SOP-FBR, the Schmidt basis (*aka* SPPs in the MCTDH literature) are expressed in terms of a weighted sum of orthonormal polynomials. The Schmidt basis are universal for the same type of PES and enclose in a non-evident way the topography of the PES (correlation among different DOF). The SOP-FBR expression is analytical and differentiable *ad infinitum* and as such, SOP-FBR is ideally suited for its use in (classical or semi-classical) dynamical simulations. More specifically, SOP-FBR allows for a seamlessly interfacing with quantum dynamical codes such as MCTDH[62] or Quantics.[63] We have shown that the SOP-FBR can be straightforwardly symmetrized by *ansatz* thus diminishing the computational cost of the algorithm. In addition to this, we have also profited from the sparse character of the Tucker *core* tensors. The quality and smoothness of the SOP-FBR PES has been extensively benchmarked against reference potentials (both model and high-level *ab initio* PES) by carefully analysing the different errors (RMSE, correlation plots, cumulative RMSE) as well as by comparing the resulting vibrational eigenvalues of each PES. We have observed that SOP-FBR can lead to Tucker expansions which are *more optimal* than those obtained by an equivalent POTFIT (or related methods). The current (beta) version of the SOP-FBR code has been written in Python 3.8.5 and will be freely distributed upon request. The implementation has been interfaced with the Heidelberg MCTDH software package.[62]

VIII. SUPPLEMENTARY MATERIAL

The full list of coefficients corresponding to the SOP-FBR PES for H_2O and HONO are available in the Supplementary Material file. Energies are given in cm^{-1}. In each list, the first two lines present (1) the core dimensions (same as the number of DOF) and (2) the number of Chebyshev functions per dimension, respectively. The rest of the lines present the Chebyshev expansion coefficients and the core tensor elements in a column-major (colexicographic) ordering.

IX. ACKNOWLEDGEMENTS

The authors would like to thank Hans-Dieter Meyer and Markus Schröder (Heidelberg) for fruitful and sustained discussions. The authors would also like to thank Tucker Carrington (Queen's) for his careful reading of the original version of the manuscript and his thoughtful comments. RLPB thanks the Région Hauts-de-France and the Université de Lille for a Ph.D. Fellowship. The authors gratefully acknowledge the computational support of Valérie Vallet and Florent Réal (Lille) and Andrei Borissov (Paris-Saclay). The authors acknowledge the usefulness Reviewers' comments and suggestions that have helped improving this manuscript. Finally, RLPB would like to especially thank Jimmy Aguilar Mena (BSC-CNS) for fruitful discussions on Computer Science related aspects.

---

[1] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, Phys. Rep **324**, 1 (2000).

[2] H. Wang and M. Thoss, J. Chem. Phys. **119**, 1289 (2003).

[3] U. Manthe, J. Chem. Phys. **128**, 164116 (2008).

[4] O. Vendrell and H.-D. Meyer, J. Chem. Phys. **134**, 044135 (2011).

[5] S. Römer, M. Ruckenbauer, and I. Burghardt, J. Chem. Phys. **138**, 064106 (2013).

[6] G. W. Richings, I. Polyak, K. E. Spinlove, G. A. Worth, I. Burghardt, and B. Lasorne, Int. Rev. Phys. Chem. **34**, 269 (2015).

[7] P. S. Thomas, T. Carrington, J. Agarwal, and H. F. Schaefer, J. Chem. Phys. **149**, 064108 (2018).

[8] D. Mendive-Tapia, E. Mangaud, T. Firmino, A. de la Lande, M. Desouter-Lecomte, H.-D. Meyer, and F. Gatti, J. Phys. Chem. B **122**, 126 (2018).

[9] V. Baranov and I. Oseledets, J. Chem. Phys. **143**, 174107 (2015).

[10] H. R. Larsson, J. Chem. Phys. **151**, 204102 (2019).

[11] S. M. Greene and V. S. Batista, J. Chem. Theory Comput. **13**, 4034 (2017).

[12] N. DeGregorio and S. S. Iyengar, J. Chem. Theory Comput. **15**, 2780 (2019).

[13] F. Gatti, C. Iung, M. Menou, Y. Justum, A. Nauts, and X. Chapuisat, J. Chem. Phys. **108**, 8804 (1998).

[14] D. Lauvergnat and A. Nauts, J. Chem. Phys. **116**, 8560 (2002).

[15] A. Jäckle and H.-D. Meyer, J. Chem. Phys. **104**, 7974 (1996).

[16] D. Peláez and H.-D. Meyer, J. Chem. Phys. **138**, 014108 (2013).

[17] M. Schröder and H.-D. Meyer, J. Chem. Phys. **147**, 064105 (2017).

[18] F. Otto, J. Chem. Phys. **140**, 014106 (2014).

[19] F. Otto, Y.-C. Chiang, and D. Peláez, Chem. Phys. **509**, 116 (2018).

[20] R. L. Panadés-Barrueta, E. Martínez-Núñez, and D. Peláez, Front. Chem. **7**, 576 (2019).

[21] G. W. Richings and S. Habershon, J. Chem. Phys. **148**, 134116 (2018).

[22] B. J. Braams and J. M. Bowman, Int. Rev. Phys. Chem. **28**, 577 (2009).

[23] R. Dawes, D. L. Thompson, Y. Guo, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **126**, 184108 (2007).

[24] C. Sanz-Sanz, O. Roncero, M. Paniagua, and A. Aguado, J. Chem. Phys. **139**, 184302 (2013).

[25] T. Frankcombe and M. Collins, Phys. Chem. Chem. Phys. **13**, 8379 (2011).

[26] B. Jiang, J. Li, and H. Guo, International Reviews in Physical Chemistry **35**, 479 (2016).

[27] O. T. Unke, D. Koner, S. Patra, S. Käser, and M. Meuwly, Mach. Learn.: Sci. Technol. **1**, 013001 (2020).

[28] V. Barone, J. Chem. Phys. **122**, 014108 (2005).

[29] R. Ramakrishnan and G. Rauhut, J. Chem. Phys. **142**, 154118 (2015).

[30] L. Barnes, B. Schindler, I. Compagnon, and A.-R. Allouche (2017), 1704.02144.

[31] S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **125**, 084109 (2006).

[32] E. Pradhan and A. Brown, Phys. Chem. Chem. Phys. **19**, 22272 (2017).

[33] B. Jiang and H. Guo, J. Chem. Phys. **141**, 034109 (2014).

[34] F. Richter, P. Carbonniere, A. Dargelos, and C. Pouchand, J. Chem. Phys. **136**, 224105 (2012).

[35] P. S. Thomas, M. F. Somers, A. W. Hoekstra, and G. J. Kroes, Phys. Chem. Chem. Phys. **14**, 8628 (2012).

[36] F. Richter, P. Carbonniere, and C. Pouchan, J. Chem. Phys. **114**, 1401 (2014).

[37] G. Ávila and T. Carrington, J. Chem. Phys. **143**, 044106 (2015).

[38] A. Aerts, P. Carbonniere, F. Richter, and A. Brown, J. Chem. Phys. **152**, 024305 (2020).

[39] M. Schröder, J. Chem. Phys. **152**, 024108 (2020).

[40] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, SIAM Journal on Matrix Analysis and Applications **21**, 1253 (2000).

[41] J. D. Carroll and J. J. Chang, Psychometrika **35**, 283 (1970).

[42] R. A. Harshman, UCLA Working Papers in Phonetics **16**, 1 (1970).

[43] T. G. Kolda and B. W. Bader, SIAM Review **51**, 455 (2009).

[44] V. K. B. N. Khoromskij, Central European Journal of Mathematics **5**, 523 (2007).

[45] J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic, J. Mach. Learn. Res. **20**, 925 (2019), ISSN 1532-4435.

[46] S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **125**, 084109 (2006).

[47] D. Peláez, K. Sadri, and H.-D. Meyer, Spectrochimica Acta part A **119**, 42 (2014).

[48] D. Peláez and H.-D. Meyer, Chem. Phys. **482**, 100 (2017).

[49] A. Haller, D. Peláez, and A. Bande, J. Phys. Chem. C **123**, 14754 (2019).

[50] M. Nest and H.-D. Meyer, J. Chem. Phys. **117**, 10499 (2002).

[51] D. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).

[52] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., Nature Methods **17**, 261 (2020).

[53] J. Nocedal and S. J. Wright, *Numerical optimization* (Springer, Heidelberg, 2006).

[54] O. L. Polyansky, P. Jensen, and J. Tennyson, J. Chem. Phys. **105**, 6490 (1996), pES, water.

[55] R. L. Panadés-Barrueta and D. Peláez, *SRPTucker software package* (2019).

[56] J. J. P. Stewart, `HTTP://OpenMOPAC.net` (2016).

[57] J. J. P. Stewart, J. Mol. Model. **19**, 1 (2013).

[58] A. H. G. Kan and G. T. Timmer, Math. Program. **39**, 57 (1987).

[59] M. J. D. Powell, Cambridge NA Report NA2009/06, University of Cambridge pp. 26–46 (2009).

[60] S. Kopec, E. Martínez-Núñez, J. Soto, and D. Peláez, Int. J. Quant. Chem. p. e26008 (2019).

[61] F. Richter, M. Hochlaf, P. Rosmus, F. Gatti, and H.-D. Meyer, J. Chem. Phys. **120**, 1306 (2004).

[62] G. A. Worth, M. H. Beck, A. Jäckle, and H.-D. Meyer, The MCTDH Package, H.-D. Meyer, Version 8.4.12. See http://mctdh.uni-hd.de/ (2016).

[63] G. A. Worth, K. Giri, G. W. Richings, I. Burghardt, M. H. Beck, A. Jäckle, and H.-D. Meyer, The QUANTICS Package, Version 1.1, See https://www2.chem.ucl.ac.uk/worthgrp/quantics/doc/quantics/citation.html/ (2015).

Chapter 4

The Pyrene-NO_2 system

Les obstacles sont inventés pour être vaincus; quant aux dangers, qui peut se flatter de les fuir ?

Jules Verne, *Cinq Semaines en Ballon*

In this chapter we present the initial results of the application of the previously described methodologies to the Pyrene($C_{16}H_{10}$)-NO_2 system, which is depicted in Figure 4.1. The study of the aforementioned interaction (Pyr-NO_2 in what follows) has been performed using both the recently developed van der Waals extension of the Transition State Search by Chemical Dynamics Simulation method (vdW-TSSCDS)[4], as implemented in a beta version that is now included in the AutoMekin software package[236], the SRP-MGPF method [7]. The former (Section 2.5) allows for efficiently and automatically sample the molecular configuration space of a (non)-covalently bound system in order to obtain the *so-called* Reaction Network (RXN), i.e. all the stationary points of the system and their topological interconnections. The latter method (Chapter 3) generates a reparametrized semiempirical electronic Hamiltonian which is used to generate a PES expressed on the grid which is subsequently tensor decomposed by the MGPF [8] or the HOOI [188] algorithms in order to perform nuclear quantum dynamics calculations. The PESs generated with SRP-MGPF can be also employed, as is the case here, as a high quality guess for the SOP-FBR method [11]. To the best of our knowledge, the present study constitutes the first theoretical investigation of the system of interest in a systematic way. This chapter is divided in two main parts, the first one dealing with the topographical studies of the Pyr-NO2 intermolecular (I) PES, and the second one with the computation of the interaction potential.

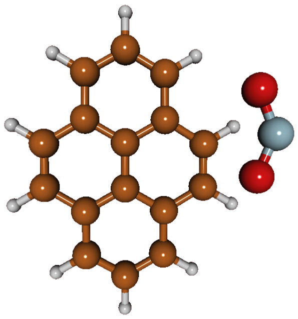

FIGURE 4.1: The Pyrene($C_{16}H_{10}$)-NO_2 system. Carbon atoms are represented in brown, Hydrogen atoms in white, Oxygen atoms in red, and the Nitrogen atom in light gray.

4.1 Topographical studies of intermolecular potentials

In this Section, we shall present a concise description of the fully automated algorithm employed for the determination of the relevant (non-covalently bound) stationary points for the target system and a description of the levels of electronic structure theory used to characterize its intermolecular Potential Energy Surface (PES). The obtained stationary points will be presented and from them the reaction network will be elucidated.

4.1.1 The (vdW-)TSSCDS method revisited

As discussed in Section 2.5, the Transition State Search using Chemical Dynamics Simulation (TSSCDS) method [5, 6, 205] in its freely distributed implementation AutoMekin [236] provides an elegant way of exploring PESs and simultaneously characterizing the stationary points in covalent [5, 6, 205] as well as organometallic compounds [237]. The TSSCDS algorithm has been recently generalized in our research group in collaboration with its core developer for the study of non-covalent compounds, the so-called van der Waals (vdW)-TSSCDS approach[4]. These modifications have been recently implemented in the AutoMeKin software[236]. Furthermore, we have recently shown that the (vdW-)TSSCDS approach constitutes an excellent means of generating a set of *highly optimal* reference points for the subsequent fitting of a global PES [7].

The basic idea behind the (vdW-)TSSCDS approach is to sample the configuration space of a molecular system making use of a large number of high-energy microcanonical (or canonical if desired) classical trajectories in which the forces are provided by inexpensive methods, e.g. semiempirical (low-level, LL) methods. The method is *quasi* black-box and requires a single initial input geometry to commence, although several geometries can be used in order to reduce computational time and ensure a more efficient sampling of the regions of interest in the configuration space. The LL search for transition states in vdW-TSSCDS is a metaheuristic process that loosely resembles the Iterated Local Search algorithm[238]. The BBFS-detected guess structures (see Section 2.5) are used as starting points for transition state (TS) optimizations at the low-level of theory. Due to the stochastic nature of the method, the resulting transition states must be screened in order to detect duplicates. The screening process is performed by comparison of one or several molecular parameters in the queried structures, for example the weighted connectivity matrices, a set of SPRINT (social permutation invariant) coordinates for each structure [213] or the combined set of energies and harmonic frequencies of the molecule. With the previous step we ensure that no duplicated structure is present in the set. Intrinsic reaction coordinate (IRC) [214] calculations are then performed for every transition state in the pruned set. The IRC calculations are followed by optimizations of the final step geometries in the forward and backward direction, which are then classified into minima and reaction products of the PES. This information allows one to construct the so-called reaction network (RXN) that summarizes which transition state connects which minima and products. At this point it should be noted that given a set of equivalent geometries, the algorithm will provide only one of these structures [4].

With the optimized LL transition states in hand, one can proceed to its reoptimization to an appropriate high-level (HL) *ab initio* or Density Functional Theory (DFT) electronic structure method. The low-level transition states serve therefore as guess structures for the subsequent high-level optimizations. As in the LL step, one can follow the intrinsic reaction coordinates for all HL transition states and identify the corresponding minima and products, thus allowing for a complete characterization of the reaction network at the HL. By combining the low-level and high-level calculations, the (vdW-)TSSCDS method becomes a powerful tool to obtain accurate information about reactivity and topography of the PES of a system at a low computational cost. The method has been successfully employed in a number of different applications ranging from combustion chemistry, [239, 240] through photolysis [241–243] and mass spectrometry [244] to organometallic catalysis [237].

In the vdW-TSSCDS formulation the information about intramolecular bounds is contained in the on-diagonal blocks of the connectivity matrix (Equation 2.179), whereas the non-covalent interactions are described by the off-diagonal blocks (see Section 2.2.5). The definition of the reference distance (Equation 2.178) consequently needs to be extended, being now the sum of covalent radii or the sum of van der Waals radii, respectively. This definition of the connectivity matrix has enabled the study of n-body problems within the rigid and semi-rigid monomer approaches, as well as in a fully flexible manner. [4].

4.1.2 vdW-TSSCDS calculations on the Pyr-NO2 system

As it was previously outlined, the vdW-TSSCDS method operates in two main stages: LL and HL. Even though the default set of LL and HL parameters are typically appropriate for *any* system (this has been the rule for our studies so far), it is of course possible to fine tune them according to the needs of the system in hand. One of the key aspects in this process is the selection of the adequate level of electronic structure, which will be discussed in the next subsection.

The low level stage of the method was initialized with a set of different geometries containing the NO_2 molecule judiciously placed according to the symmetry (D_{2h}) of the Pyrene molecule at its equilibrium geometry. Following the logic of vdW-TSSCDS, two main fragments were defined (the Pyrene and the NO_2 molecule) and the electronic structure level was specified (see Section 4.1.3). The exploration of the configuration space started in parallel from each of the initial geometries. A large number of batches of microcanonical trajectories with a maximum energy of 200 kcal/mol were propagated for 500 fs. The temporal evolution of the connectivity matrices was analyzed at a fixed frequency (every 5 fs) by the BBFS algorithm, which was in turn also parametrized to set thresholds for the detected values of the lower harmonic frequencies (see discussion in Section 3 in Reference [4]). The candidate structures to LL TS were then screened and the resulting unique[1] geometries were optimized with a large gradient threshold. The rationale of this choice is that the obtained transition states at this level don't need to be *exact* but rather *close* to the high level transition states. By increasing the level of accuracy of the low level calculations one can potentially lose valuable guess structures at the HL due to the different topographies of the PES at the different levels (see discussions in Sections 3 and 4.3 in Reference [4]). To finish at the LL stage one may optionally determine the RXN and isolate the corresponding minima and products.

The high level stage of the method receives as input the LL transition states and reoptimizes then at the new level of electronic structure. This constitutes the "topographic filter" of the method because some of the LL transition states might not exist at the HL. The *surviving* optimized transition states are then screened in search for duplicates, and the unique ones are used as initial structures in IRC calculations. The results of this IRC calculations allow for obtaining the corresponding minima and products, and with them the high level RXN.

It can be inferred from the previous description that the bottleneck of the algorithm (apart from the unavoidable HL optimizations and IRCs) is the computation and BBFS-identification of *all* the LL TSs. Despite the efficiency of the independent algorithms, the

---

[1] There are symmetry equivalent geometries (over and below the Pyrene cycle for instance) that must be generated in an additional step.

stochastic nature of the method might imply a large number of LL trajectories and the concomitant BBFS steps, thus rendering the whole process slow. In this respect, it has to be emphasized that new promising techniques of accelerated dynamics are currently being implemented in the AutoMeKin software [236]. The remaining steps of the algorithm are fairly straightforward with the exception of the high level transition state optimizations, which are far more complicated than the optimization of minima [245]. The IRC calculations can be also problematic if the region of the PES in consideration is flat (shallow potential valley) [246], as it is frequently the case in weakly bound complexes. As it should be clear, these are issues independent of the (vdW-)TSSCDS algorithm. For a more detailed description of the (vdw-)TSSCDS algorithm please consult Section 2.5.

4.1.3 Electronic Structure Calculations: Low-level and High-level

The (vdW-)TSSCDS method relies on the electronic structure machinery available for the system in hand. Our approach in this regard was to maximize the ratio between the quality of results and computational time. The AutoMekin software package [236] (former (vdW-)TSSCDS) has implemented the suitable interfaces to the corresponding *academic standard* electronic structure software packages. As LL calculations are usually carried out with semiempirical methods, we use MOPAC2016 [247] to perform them. All the HL calculations has been carried out by the Gaussian software package [248]. Concerning the low-level of theory, the PM7 method [106] is the default choice in the TSSCDS approach and it has proven its capabilities in chemical studies ranging from covalently bound systems [6] to organometallic ones [237]. More recently, it has shown its value in the case of non-covalently bound systems [4].

With respect to the high-level, all stationary points (minima and first-order saddle points) as well as points along the IRC have been optimized and characterized by their harmonic frequencies with the unrestricted version of the long-range and dispersion corrected functional ωB97X-D by Chai and Head-Gordon [249]. All geometries have been optimized at the tight criterion from Gaussian and the option SuperFineGrid has been used throughout in order to avoid spurious results in the DFT calculations [250]. A benchmark of several Dunning basis sets (see Section 2.2.2) was performed. The results indicated that the cc-pVDZ is one that offered the best compromise between accuracy and CPU time. The proposed approach has been shown to be adequate for the treatment of similar systems, where one needs an accurate description of the long-range dispersion forces [210, 251].

As was briefly mentioned in the previous subsection and extensively discussed in Reference [4], IRC calculations might be very challenging for very flat regions of the potential. This implies that to obtain a smooth profile (or even convergence) in the IRC curves the algorithm selection and their corresponding parameters needs to be craftily tinkered. There is a number of IRC algorithms available in the Gaussian software package [248]. The selection of the most adequate for the problem in hand depends on many factors like size and nature of the system and the underlaying electronic structure approach. Whereas the Hessian-based Predictor-Corrector integrator [252] (HPC) is probably the most accurate and usually a default choice algorithm, for a large system with very flat potential landscape like Pyr-NO_2 a simpler local quadratic approximation [253] (LQA) has proven to be a better choice, as has been established [254]. In addition to this choice, the step size of the calculations was also reduced with respect to the default value and the number of steps increased. This combination of options has allowed the computation of relatively smooth IRC energy profiles.

4.1.4 Low Level calculations: PM7 semiempirical guess PES

Any stochastic optimization algorithm must define a set suitable termination conditions in order to properly operate. In our case, we have used the maximum number of iterations[2] and an iteration threshold for the newly discovered transition states (LL-TS). In simple words, we increase progressively the number of microcanonical trajectories with random initial conditions (compatible with the total energy available), thus sampling different regions of configuration space and stop the search when the convergence curve (number of unique LL-TS versus the number of trajectories) reaches a *sufficiently large plateau* for a fixed number of trajectories.

Using the MOPAC2016 software package, a total of 41629 dynamic reaction coordinate trajectories[3] (DRC) recorded every femtosecond were launched at the PM7 level of theory. The calculations were distributed in 4120 batches, and the average CPU time was of 78 s. The convergence threshold was set to 40 % of the total number of trajectories. The trajectories were started from an small set of initial geometries chosen to cover key regions of configuration space according to the symmetry of the *substrate*, the Pyrene molecule. The number of unique transition states was monitored at a constant frequency by screening of the BBFS guess geometries. The convergence process can be better understood by analyzing Figure 4.2, in which the evolution of the amount of unique detected transition states is represented again the total number of iterations. Note how the final stabilization of the amount of transition states spans 40 % of the total number of trajectories. At the end of this stage, all unique transition states were fed to the HL section of the method. As it can be appreciated on the plot, a total of 23 unique transition states was found.

---

[2] Each iteration corresponds to a trajectory in our framework.

[3] The path followed by all atoms of a system while assuming constant energy. See the MOPAC manual for more details.

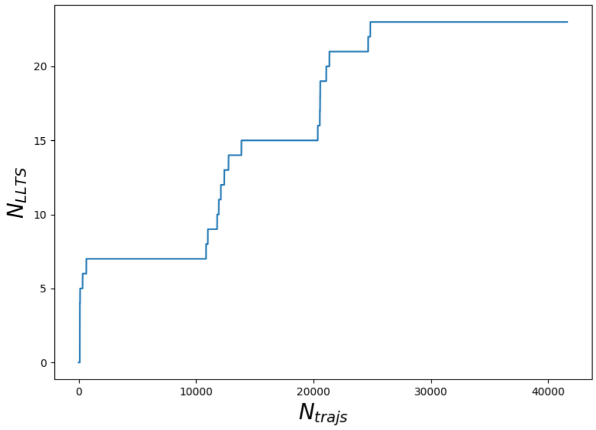

Figure 4.2: Convergence of the amount of unique LL transition states detected by the vdW-TSSCDS algorithm with respect to the total number of trajectories. The termination condition was set to no new structures in span of 40 % of the total number of trajectories, with a minimum of 30000 trajectories. Note the stabilization of the curve in the last segment.

4.1.5 High Level calculations

Once all transition states at the LL were located, we proceeded to perform the HL calculations. In this second stage the former structures were reoptimized at a higher level of theory, which was set to the unrestricted version of the long-range and dispersion corrected functional UωB97XD/cc-pVDZ (see Section 4.1.3). After the corresponding screening, IRC calculations were carried out on the surviving structures and the corresponding minima were optimized. Single-point BSSE calculations were also performed. Due to the size of the system of interest, nodes with octa-core processors and 36 gigabytes of RAM were used for each optimization.

4.1.5.1 Transition states

The 23 unique LL transition states were then reoptimized at full dimensionality (81D) using the Gaussian software package. The convergence criteria was set to *tight* and the default maximum number of cycles increased to 100. The average CPU time for the TS optimizations under these conditions was 4.5 hours. After optimization and screening, a total of 5 unique transition states were obtained. The corresponding structures are represented in Figure 4.3, where we have represented with green arrows the normal modes associated with the imaginary frequency.

As we can see from the frontal and transverse projections in Figure 4.3, in all structures but one (TS5) the projection of the center of masses of the NO_2 molecules lies near

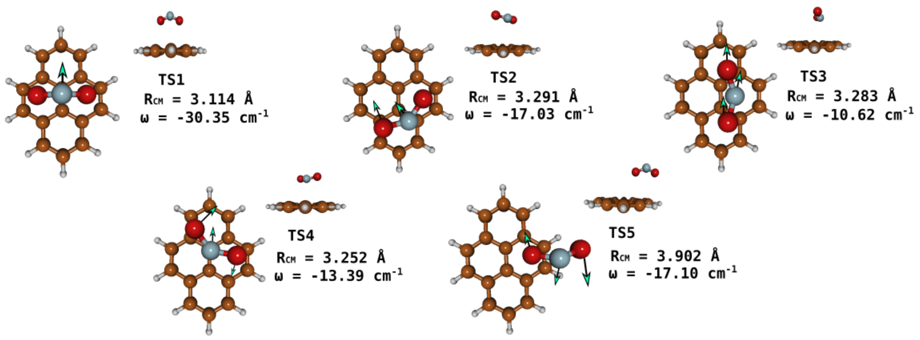

FIGURE 4.3: Frontal and transverse projections of the five transition states of the Pyr-NO_2 system obtained after optimization of the 23 LL transition states at the UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory, and subsequent screening. The normal mode associated with the imaginary frequency is represented and the value of the latter reported, as well as the distance between the centers of mass of both molecules R_{CM}.

the in the vicinity of the central C atoms of Pyrene. The average distance between the centers of mass between the molecules R_{CM} is 3.3682 Å. The orientation of the NO_2 molecule also covers a wide range of angles, with some structures having the oxygen atoms pointing “downwards” with respect to the Pyrene (TS1, TS5) and some “upwards” (TS2, TS3). In agreement with experimental findings [255] in which no reactivity between NO_2 and Pyrene was observed, we have only found physisorbed structures. Nevertheless, to further test this observation, we performed some additional HL optimizations by manually placing the NO_2 molecule at distances ranging from 3.0 up to 5.0 Å over different sites of the Pyrene molecule. Neither was found reactivity in this case.

4.1.5.2 IRC calculations

We have further characterized the HL TS by performing the corresponding IRC calculations in order to obtain the minima and products that will further conform the reaction network (RXN). We performed the IRC calculations using the LQA algorithm with reduced step size and augmented number of steps, keeping the same level of theory as for the transition states (UωB97XD/cc-pVDZ). The corresponding energy profiles are shown in Figure 4.4.

In the plots the geometry of the corresponding transition state has been set in the zero of energy (in cm^{-1}). The first observation here is that the curves are not smooth and the corresponding landscapes are very shallow in all cases, with a maximum barrier of only 200 cm^{-1} and a minimum of less than 10 cm^{-1}. The limitations of IRC calculations for this kind of topology was have been addressed before (see Section 4.1.3), and as expected for the Pyr-NO2$_2$ system, several difficulties were encountered. Despite our efforts to circumvent convergence issues that implied a fine tuning of each calculation individually, some of the IRC curves do not show a smooth profile.

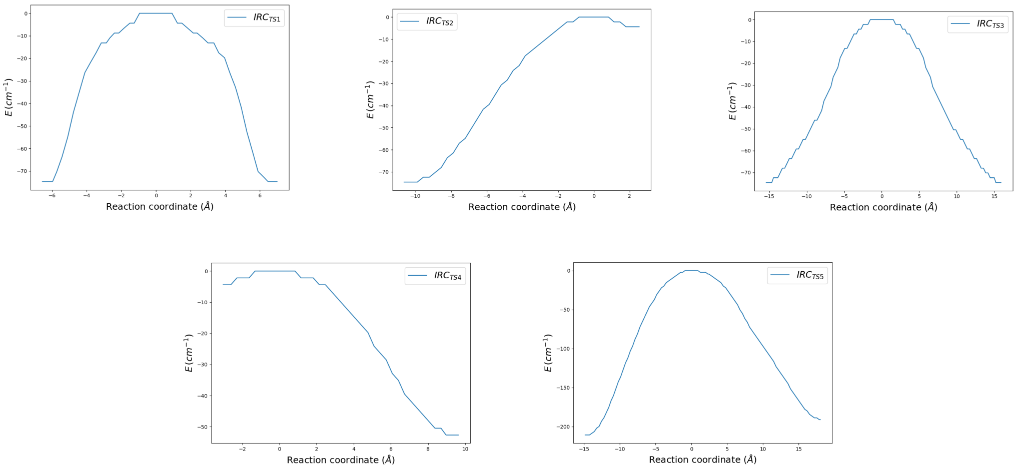

FIGURE 4.4: Intrinsic Reaction Coordinate energy profiles obtained with the LQA algorithm for each one of the five transition states of the Pyr-NO$_2$ system at the UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory. Energies are reported in cm^{-1}.

In addition to the LQA algorithm, all of the available algorithms in the Gaussian software packaged were tested. The presented results are the best the authors could converge at the selected level of theory, and the underlying difficulties will be considered in what follows as *state of the art* limitations of this particular kind of calculation [246]. The final structures in each of the two directions of the IRC pathways were further optimized in search for minima or asymptotic products, and together with the transition states and the corresponding IRCs, they constitute the RXN. In the next subsection the optimization of this structures is described in detail.

4.1.5.3 Minima derived from the IRCs

Starting from the 10 final structures of the IRC calculations we proceeded to launch the corresponding optimizations in search for minima and products. The convergence criteria was set to *tight* and the default maximum number of optimization cycles increased up to 100. Under these conditions the median computational time of the optimizations was 4.1 hours. Some of the minima were very hard to optimize and required further human intervention, with execution times of roughly 100 hours (particularly in the case of the minima corresponding to TS5). All the optimized structures have been characterized as minima (all harmonic frequencies are positive), i.e. there are no asymptotic products in the obtained reaction network at the established level of theory. The results of this stage are presented in Figure 4.5.

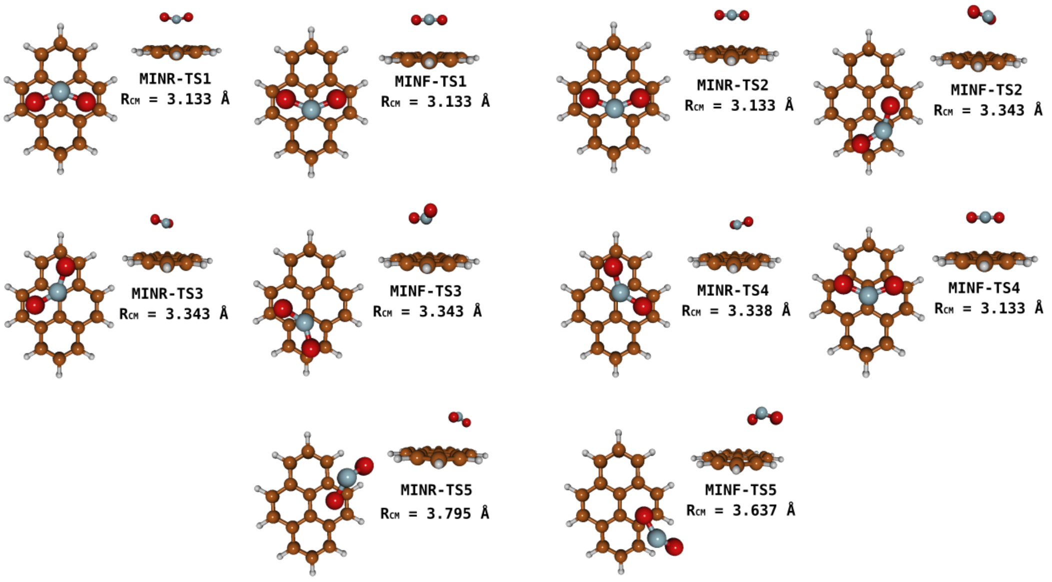

FIGURE 4.5: Frontal and transverse projections of the 10 minima of the Pyr-NO_2 system obtained after optimization of the ending structures of the 5 corresponding IRC pathways at the UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory. The distance between the centers of mass of both molecules R_{CM} is reported. The minima are labeled accordingly to their corresponding transition states and the letters **R** and **F** make reference to the reverse and forward directions respectively.

The geometries of the minima are relatively close to those of their corresponding transition states, which is in agreement with the shallow topography of the PES. The NO_2 molecule covers again a wide range of orientations, and its average distance from the Pyrene plane (distance R_{CM} in the figures) is 3.333 Å. It should be mentioned that the (vdW-)TSSCDS method does not provide all geometries generated by symmetry operations on the NO_2 molecule with respect to the Pyrene plane, and consequently further analysis must be performed in order to generate the possibly missing structures.

4.1.6 BSSE estimation

As it is customary (and necessary) in this type of studies, we studied the influence of the Basis Set Superposition Error (see Section 2.2.2). Previous results seem to indicate that for van der Waals complexes the Counterpoise correction (CP)[99] is not useful[256]. Nevertheless, we carried out standard CP calculations for all the stationary points using the Gaussian software package. Two fragments were defined, i.e. the NO_2 and the Pyrene molecules, and the same level of theory of the previous HL calculations was kept. The BSSE corrected energies are compared with the uncorrected energies in Table 4.1.

Structure	$\Delta E(cm^{-1})$	$BSSE(cm^{-1})$
MINR-TS1	4.3894e-04	-1.9603e-03
MINR-TS2	4.3894e-04	-1.9120e-01
MINR-TS3	6.4751e+01	4.8384e+01
MINR-TS4	3.8800e+01	1.7709e+01
MINR-TS5	7.6236e+01	2.1028e+02
TS1	7.4973e+01	2.6365e+02
TS2	7.4372e+01	3.8658e+01
TS3	1.3858e+02	5.3892e+01
TS4	5.1983e+01	1.1312e+01
TS5	2.8805e+02	3.3307e+02
MINF-TS1	4.3894e-04	-9.0171e-02
MINF-TS2	6.4752e+01	4.8395e+01
MINF-TS3	6.4753e+01	4.8308e+01
MINF-TS4	0.0000	0.0000
MINF-TS5	9.7619e+01	7.3458e+01

TABLE 4.1: Energies in cm^{-1} relatives to the global minimum of all stationary points of the Pyr-NO_2 PES computed with and without BSSE correction at the UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory.

Some general remarks can be made with the gathered information. Important topographical differences (different relative energies) can be observed when comparing the BSSE corrected PES with the uncorrected one. Indeed, the global minimum changes from MINF-TS4 in the uncorrected to MINR-TS1 in the BSSE PES and the RMSE of both sets of energies is 67.09 cm^{-1}. Provided the height of the barriers in the uncorrected PES, these deviations represent a substantial change in the structure of the RXN. To further check the validity of th CP results a more detailed investigation should be carried out. The above will be part of a dedicated study.

4.1.7 Reaction network

To conclude the first part of this chapter, the RXN formed by the stationary points of the Pyr-NO_2 system will be presented. We will start our analysis with the IRC corresponding to the TS1. As it can be seen in Figure 4.6, the two minima (MINF-TS1 and MINR-TS1, with F,R referring to the forward and reverse directions respectively) and the aforementioned transition state have symmetry C_s, i.e. there is only one element of symmetry in addition to the identity (E) and is the mirror plane containing the N and the four central C atoms. The importance of this piece of information will become apparent when analyzing the RXN, for which these are the central structures.

The proposed reaction network of the system is represented in Figure 4.7. As suggested before, the IRC$_{TS1}$ lies at the center of the RXN. The evidence collected from the analysis

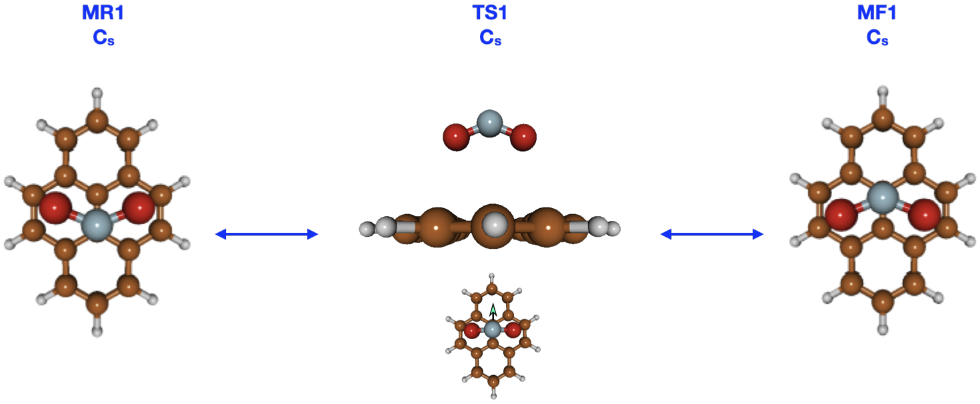

FIGURE 4.6: Stationary points of the IRC_{TS1}. The minima in the reverse and forward directions have been labeled MR1 and MF1 respectively for simplicity in the notation. The three stationary points present belong to the C_s symmetry group.

of the structures of the individual minima suggested that some of them were equivalent. In particular, MINF-TS1 is the same as MINR-TS2, which results in the connection of the central part with the upper-right part via the TS2. The primed labels in the different branches of the RXN make reference to structures generated by symmetry operations on the NO_2 molecule with respect to the Pyrene plane, which are not directly identifiable by (vdW-)TSSCDS, and must be generated in an additional step. In Figure 4.8 we have represented this situation for TS3 and TS5. Note that the corresponding TS3' and TS5' have also two minima associated with them, which are in turn equivalent to the unprimed minima.

Back to the RXN discussion, in Figure 4.7 we can appreciate that the *central* IRC_{TS1} is also connected via TS4 with the MINR-TS4. The above is a consequence of the equality between the MINF-TS1 and MINF-TS4. At the same time, MINR-TS4 is equivalent with both minima from TS3 (see Figure 4.8), which adds the extra connection in the bottom branch of the RXN. The remaining transition state (TS5) and its corresponding minima were more difficult to connect with the rest of the RXN. After a careful analysis it was concluded that an additional transition state was missing (labeled TS_{new} in Figure 4.7). In order to identify this possible stationary point, we performed QST2 [257] calculations using the Gaussian software package. This method allows for finding a transition state starting from two pivotal minima. In our case, we explored all possible combinations between the minima connected through TS1, and those connected through TS5.

The calculations were challenging to converge, and prove of it is the obtained IRC profile for the new transition state, which is represented in the left panel of Figure 4.9. As we can see, the curve presents what looks like a cusp in the vicinity of the TS, which is an undesirable topographical discontinuity. To inquiry in the fact that the LL stage

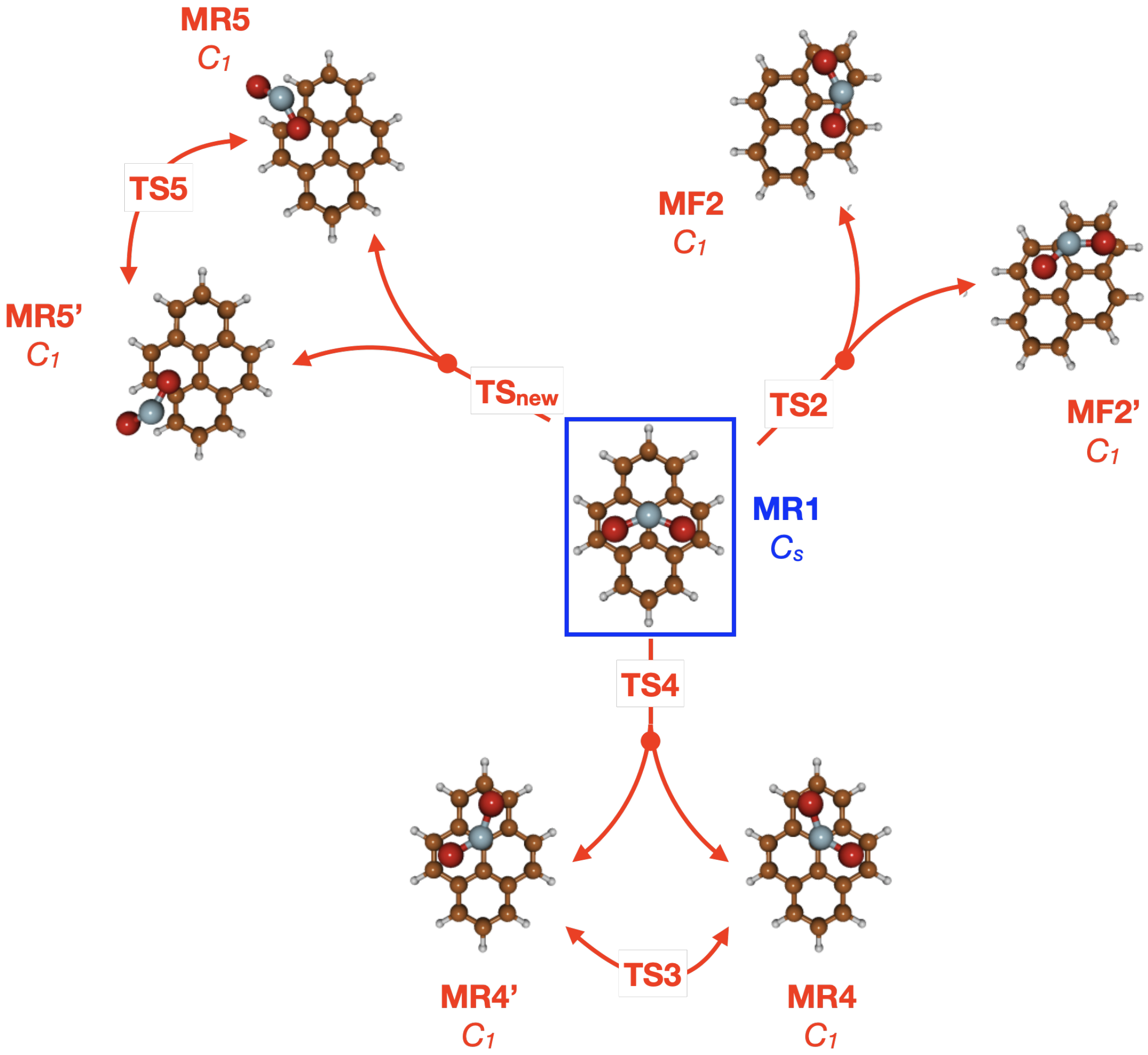

FIGURE 4.7: Reaction network of the Pyr-NO_2 system. The notation of the minima corresponding to the transition state τ has been simplified to MRτ and MFτ. Primed labels represent structures generated by symmetry operations on the NO_2 molecule with respect to the Pyrene plane. The central region (IRC$_{TS1}$) which has symmetry C_s is connected with three other regions of lower symmetry (C_1).

of (vdW-)TSSCDS did not detect the suitable guess for TS$_{new}$, we recomputed (single-point calculations) the HL IRC points at the LL at the PM7 level of theory. The results of this previous step are represented in the right panel of Figure 4.9, in which we have also included the HL curve. It can be seen that the semiempirical energy spikes in the vicinity of the TS and then forms a plateau. The above information confirms that the missing TS can be attributed to a limitation of the PM7 method and not to a flaw in the (vdW-)TSSCDS algorithm.

Finally, we would like to discuss a very interesting topographical feature that has been found in the obtained reaction network. It was mentioned before that the stationary

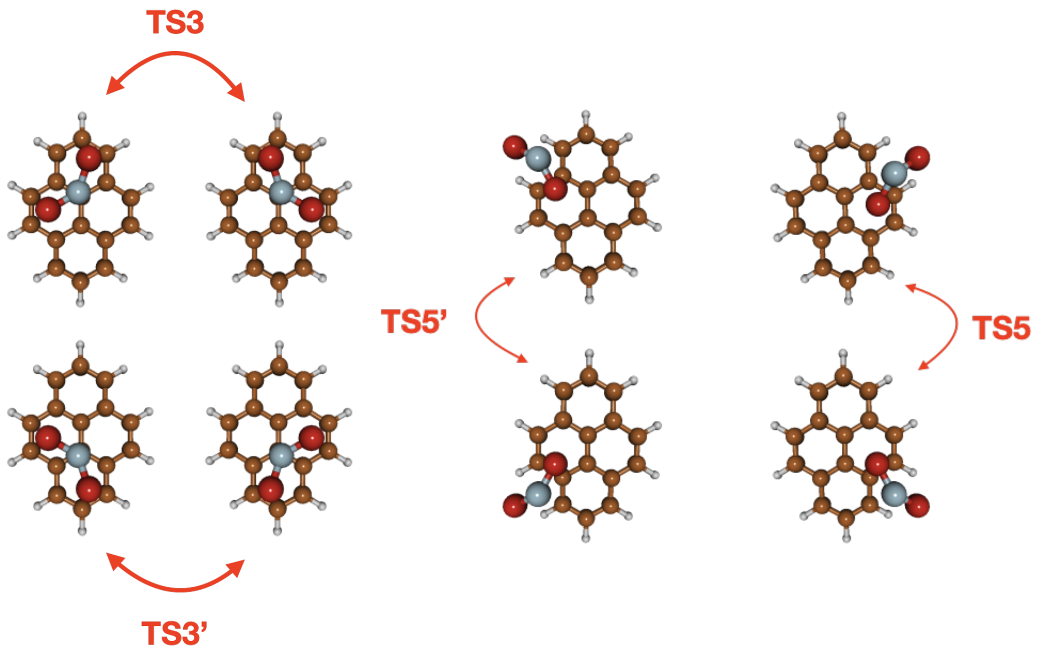

FIGURE 4.8: Equivalent structures generated by symmetry operations on the NO_2 with respect to the Pyrene plane for the TS3 and TS5. In the case of the TS3, the connection with TS1 occurs with both the vdW-TSSCDS detected structure and the equivalent TS3'.

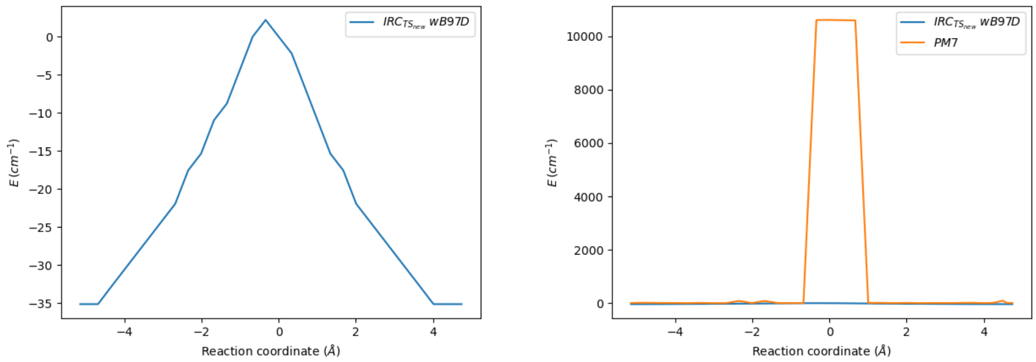

FIGURE 4.9: Intrinsic Reaction Coordinate profile for the TS_{new} structure at the (left) UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory and (right) comparison with the corresponding profile at PM7 level of theory.

points connected through the IRC_{TS1} belong to the C_s symmetry group. By simple inspection of the remaining structures shows that they belong to the C_1 symmetry group, i.e. the only symmetry operation is the identity E. A point of PES that leads from reactants to products in a way in which is energetically favorable to break the symmetry is called branching point [258]. A succinct graph-theory based representation of the RXN is presented in Figure 4.10, where we have represented the main topographical regions $\mathcal{M}_i$ with $i \in \{1, 2, 4, 5\}$ of the PES. The regions in blue and red have C_s and C_1

symmetry respectively.

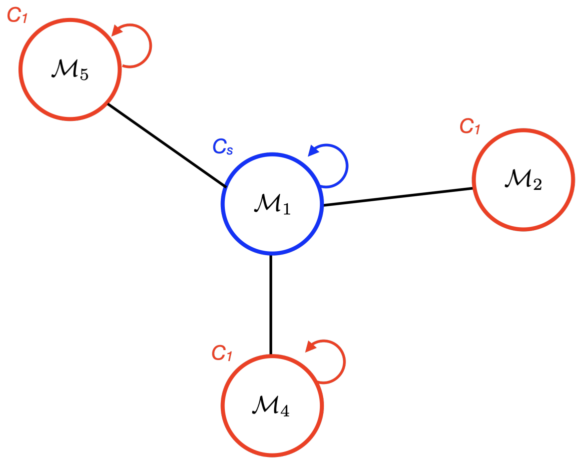

FIGURE 4.10: Graph representation of the reaction network of the Pyr-NO$_2$ system. The topographical regions $\mathcal{M}_i$ with $i \in \{1, 2, 4, 5\}$ of the PES with C_s symmetry are represented in blue, whereas the ones with C_1 symmetry are represented in red. A curved arrow indicates that the two equivalent minima are connected through a TS.

4.1.8 Partial conclusions

The present work constitutes, to the best of our knowledge, the first full dimensional (81D) theoretical study of the interaction between the Pyrene and the NO$_2$ molecules. It is also one of the first production applications of the vdW-TSSCDS method [4], a powerful tool to efficiently and *quasi* automatically sample the configuration space of a system and describe its corresponding topography. The method allowed us for finding all the stationary points of the system (transition states and minima) at the level of theory UωB97XD/cc-pVDZ Integral(SuperFineGrid).

A total of 23 transition states were found at the PM7 low level of theory after the screening of equivalent structures, in a long exploration of the configuration space with 41629 random trajectories. The structures were reoptimized at the high level of theory of choice, with the highest available size of the DFT grid. After the screening of this results, 5 transition states were found. The corresponding IRC calculations were carried out at the same level of theory. Due to the shallow topography of the PES, these calculations were hard to converge, and a benchmark of all available algorithms was made. The ending structures of the IRC calculations were then optimized in search for minima and products, resulting this into 10 different minima. The interconnections between the minima and transition states constituting the reaction network was consequently studied. Some issues regarding the LL of theory (PM7) were detected in this last process.

All the collected information suggest that the interactions in the Pyr-NO_2 system are those of a typical van der Waals complex, and no signs of reactivity between the species was found at the selected level of theory. Additional studies should address the influence of light in these physisorbed species, and explore the possible pathways of formation of the HONO molecule from them. As a final step, some BSSE calculations were carried out at the same HL of theory using the Counterpoise correction. There is some evidence that for typical van der Waals complexes this sort of correction is of no use [256]. We perform nevertheless some basic calculations on our system, and concluded that the topography of the BSSE surface differs from that of the uncorrected one, but owing to the tiny differences observed we cannot really conclude anything and further investigations are encouraged in this regard.

4.2 Reduced dimension PES (6D)

In the previous section we presented the results of the topographical studies of the Pyr-NO_2 system, and the obtained reaction network was described in detail. The next step of the study consisted in the development of the global PES surface of the system. Due to the size of the system at hand (81D), the above task was performed in a semirigid approach. The above assumption is sustained by the fact that the Pyrene geometry is almost not perturbed by the physisorbed NO_2 molecule. In this conditions, the dimensionality of the system get effectively reduced to 9D. Moreover, for the majority of the stationary points, the geometry of the NO_2 is also not considerably perturbed with respect to the isolated molecule (*vide infra*), thus allowing us to further reduce the dimensionality to 6D.

4.2.1 Coordinate transformation

As suggested before, the development of the PES of the system has been carried out in reduced dimensionality. In order to transform the Cartesian coordinates (resulting from the HL optimizations) of the stationary points obtained with the Gaussian software package to a more convenient system of internal coordinates, a suitable code has been developed (see Appendix A).

The transformation performed to pass from 81D to 9D has the form:

$$\{x_i, y_i, z_i\}_{i\in[0..N]} \longleftrightarrow \{x_i, y_i, z_i\}_{i\in[0..N-3]} \cup \{X_{cm}, Y_{cm}, Z_{cm}, \alpha, \beta, \gamma, r_1, r_2, \theta\} \qquad (4.1)$$

where N is the number of atoms, x_i, y_i, z_i are the Cartesian coordinates of the system, X_{cm}, Y_{cm}, Z_{cm} are the coordinates of the center of mass of the NO_2 molecule, α, β, γ are the corresponding Euler angles (reflecting the relative orientation of NO_2 on Pyrene), and r_1, r_2, θ its internal coordinates (distances ON and angle ONO). The unchanged $3N-9$ coordinates correspond to the frozen Pyrene molecule. The implementation of the above transformation can be a tedious task, and an adequate convention must be established in order to obtain reversible results (see Appendix A for a technical discussion). A graphical representation of the defined internal coordinates can be found in Figure 4.11.

The obtained internal coordinates for each stationary point are represented in Table 4.2. In all cases the Pyrene molecule has been placed at the origin of coordinates with its principal axes of rotation matching the Cartesian canonical axes. As it can be seen, the transformation yields geometries with are located both above and bellow the Pyrene molecular plane (YZ). The Euler angles reflect the diversity of orientations of the NO_2 already described in Section 4.1.5.3.

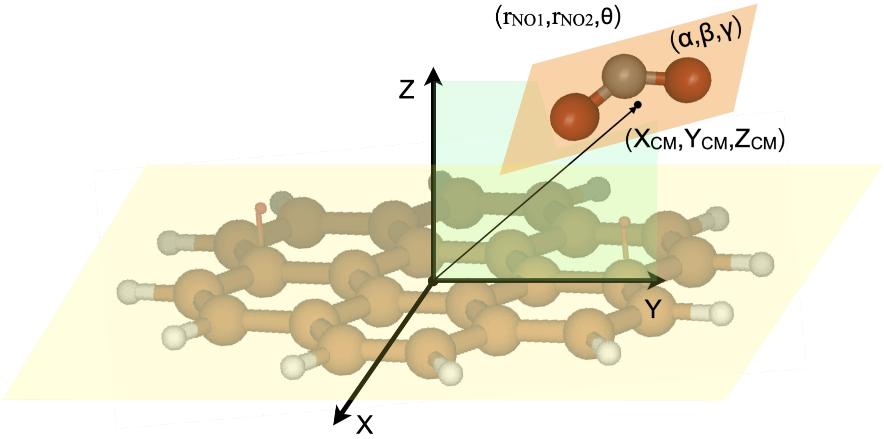

FIGURE 4.11: Internal coordinates for the system PAH-NO_2 in the reduced dimensionality frame 9D. The X_{cm}, Y_{cm}, Z_{cm} are the coordinates of the center of mass of the NO_2 molecule, α, β, γ are the corresponding Euler angles, and r_1, r_2, θ its internal coordinates.

A very important aspect here is the comparison of the internal coordinates of the NO_2 molecule with those obtained for the isolated molecule. The calculations at the level of theory UωB97XD/cc-pVDZ Integral(SuperFineGrid) showed that for latter case, the values are $r_1 = r_2 = 1.1905$ Å and $\theta = 2.3422$. For the majority of the stationary points, the aforementioned deviation occurs at the fourth decimal. In to quantify the effect of the dimensionality reduction (and its associated change in geometries), additional calculations were carried out by fixing the internal coordinates of NO_2 to their equilibrium value on the vdW-TSSCDS computed stationary states. The RMSE error between all of the original energies and the ones obtained for the modified geometries was 7.44 cm^{-1} and the MAE was 7.08 cm^{-1}.

Structure	X_{cm} (Å)	Y_{cm} (Å)	Z_{cm} (Å)	α (rad)	β (rad)	γ (rad)	r_1 (Å)	r_2 (Å)	θ (rad)
MINR-TS1	3.1334	0.0005	1.115e-2	1.5708	0.0	0.2270	1.1906	1.1906	2.3403
MINR-TS2	3.1336	0.0	-1.127e-2	1.5708	0.0	2.9137	1.1906	1.1906	2.3403
MINR-TS3	3.1550	0.3844	-1.0382	-0.6597	-1.940e-2	0.2626	1.1907	1.1908	2.3400
MINR-TS4	-3.1662	0.4331	-0.9642	-0.6272	2.371e-2	-2.7235	1.1907	1.1907	2.3396
MINR-TS5	3.0425	-1.8919	1.2520	0.7618	-7.317e-2	-0.8893	1.1887	1.1919	2.3420
TS1	-3.1145	0.0	0.0	-1.5708	0.0	1.5703	1.1903	1.1903	2.3440
TS2	-3.1830	0.2767	-0.7883	-0.9783	4.091e-2	-0.3625	1.1907	1.1907	2.3393
TS3	-3.2572	-0.4143	0.0	-3.1415	0.0	-0.7970	1.1904	1.1904	2.3404
TS4	-3.1933	0.2430	-0.5648	-0.9934	5.201e-2	-2.6571	1.1906	1.1906	2.3398
TS5	3.1340	-2.31886	0.1582	1.5061	5.080e-3	-2.2959	1.1889	1.1918	2.3423
MINF-TS1	3.1335	0.0	-1.123e-2	1.5709	0.0	2.9141	1.1906	1.1906	2.3403
MINF-TS2	3.1550	-0.3846	-1.0385	0.6614	-1.943e-2	2.8790	1.1907	1.1908	2.3400
MINF-TS3	-3.1550	-0.3838	-1.0367	-2.4848	-1.946e-2	-0.2634	1.1908	1.1907	2.3400
MINF-TS4	-3.1334	0.0	1.182e-2	-1.5708	0.0	-2.9143	1.1906	1.1906	2.3403
MINF-TS5	3.1405	-1.6451	-0.8116	2.5751	7.640e-3	3.1016	1.1893	1.1915	2.3396

TABLE 4.2: Internal coordinates of the Pyr-NO$_2$ system at the reduced dimensionality 9D. First are represented the coordinates of the relative center of mass (X_{cm}, Y_{cm}, Zcm), then the Euler angles α, β, γ, and finally the internal coordinates of the NO$_2$ molecule.

The above findings validate the approximation of further reducing the dimensionality to 6D while still getting physically meaningful results for the system of interest. In the following section we will discuss the development of the global intermolecular 6D PES.

4.2.2 SRP-MGPF fit. Guess SOP-FBR parameter set

Once a suitable coordinate transformation has been established, we can proceed to generate the reference geometries needed for the optimization process. For this purpose, we used our many body-like expansion scheme (see Chapter 3) that allows for selecting the minimal number of geometries that ensures a converged optimization. The process (described in detail in our SRP-MGPF article [7]) generates a total number of points N_{ref} which is computed as:

$$N_{ref} = N_{RXN} \cdot \Big[\sum_{i \in 1D}^{f} N_i^{(1D)} + \sum_{i \in 2D}^{f} N_i^{(2D)} + \ldots \Big] + rnd(fD) + \sum_{i}^{n_{TS}} N_i^{IRC} + \sum_{i}^{n_{asymp}} N^{(asymp)} + \ldots \tag{4.2}$$

where f is the number of DOF of the system, N is the number of generated reference geometries of a given type, for example, $N^{(nD)}$ are grid points from a n-dimensional (D) grid obtained by varying D degrees of freedom while keeping $f - D$ fixed in their equilibrium value, N_i^{IRC} are the IRC points stemming from TS_i, $rnd(fD)$ are random structures sampled in the full-D configuration space, n is the number of stationary points of a given type. The above scheme starts from the RXN elaborated by vdW-TSSCDS. In addition to the points generated by this scheme, a product grid reference tensor was also computed to be used by SOP-FBR. Here we should mention that the latter method is by no means restricted to this form of the reference geometries, but for the actual Python implementation it is very convenient (see Appendix C).

The bounds of the one-dimensional grids were set in correspondence with the internal coordinates of the stationary points (see Table 4.2). Using the coordinate transformation software we generated a reference tensor of shape $\prod_{i=1}^{6} 5$, i.e. 15625 geometries computed at the level of theory UωB97XD/cc-pVDZ Integral(SuperFineGrid). The execution time of these calculation was roughly one week, under the conditions of a shared cluster. As discussed in Section IV of Reference [11], the first step of a SOP-FBR optimization is the obtention of a adequate guess parameter set. In our case, we will use the parameter set arising from a HOOI decomposition of a SRP PES. As discussed in Chapter 3, the latter can be performed with SRP-MGPF method. In fact, the usage of this method has double purpose here: to develop a PES that is ready to be used in nuclear quantum dynamics calculations, and to offer the aforementioned guess.

With this in mind, we proceeded to carry out the SRP optimization. The reference geometries in this case were taken as a mixture of the ones yielded by the process described in Equation 4.2, and a sampling of the lower energies of the energy tensor. The semiempirical Hamiltonian reparametrized in this case was the PM7 one. Since the Pyrene molecule is kept frozen, only the parameters corresponding to the hydrogen, nitrogen, and oxygen atoms were optimized. Nevertheless, all of the two body parameters (including the ones from carbon) were optimized.

As a result of this, a set of 1071 points with energies ranging from 0 to 3346.92 cm^{-1} was optimized in a 16-core node. The CPU time of the optimization was 3.57 hours. The obtained RMSE was 292.31 cm^{-1} and a validation set of 2451 points yielded a RMSE of 449.25 cm^{-1}. In Figure 4.12 we represent the correlation plot E_{ref} vs E_{srp} and the corresponding cumulative RMSE plot for this calculation. The above are preliminary results that reflect the good properties of the SRP fit, like the obtention of RMSE bellow the Chemical Accuracy threshold.

As can be appreciated in the correlation plot, the points are still dispersed with respect to the ideal straight line. The cumulative RMSE shows a typical profile for similar optimizations (see Chapter 3). The quality of the harmonic frequencies of the stationary points was also evaluated, showing that further improvements are needed since some of the geometries optimized to higher order saddle points. This issue can be circumvented by including the harmonic frequencies in the optimization objective function, as shown in Chapter 3. On the other hand, the obtained SRP PES can be used as a suitable guess for the SOP-FBR method. In addition, the reparametrized semiempirical Hamiltonian will be used to improve the transition state search with vdW-TSSCDS, since it is a more accurate representation of the topography of the PES [259].

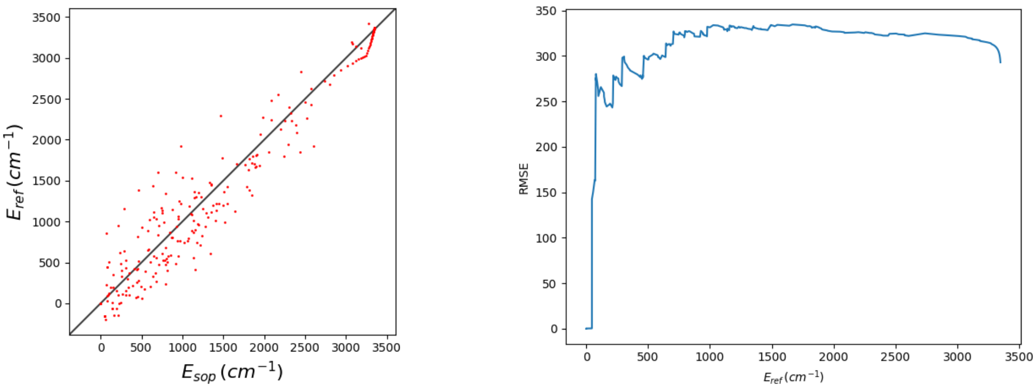

FIGURE 4.12: SRPTucker optimization results for the Pyr-NO$_2$ system. (left) Correlation plot between the reference energies (E_{ref}) and the corresponding semiempirical one E_{srp} and (right) cumulative RMSE curve.

4.3 Towards quantum dynamical studies on the Pyr-NO$_2$ system

The ground state distribution of the Pyr-NO$_2$ can be obtained by relaxation of a wave packet with the MCTDH method (see Section 2.3.2.1). In the previous section we discussed how the global PES is obtained with the SRP-MGPF method. This surface will be also used to generate a suitable guess for the SOP-FBR method, which yields a directly PES in the adequate mathematical form needed by MCTDH. To complete the Hamiltonian operator we only need the KEO, but as it was presented in Section 2.3.2.1, this problem is relatively straightforward in this case. Since in our case the surface (Pyrene molecule) is frozen, the model of a triatomic molecule in a properly defined body-fixed frame (BF), and with center of mass G that is the origin of a Space-Fixed frame (SF) [172], suffices to describe our problem.

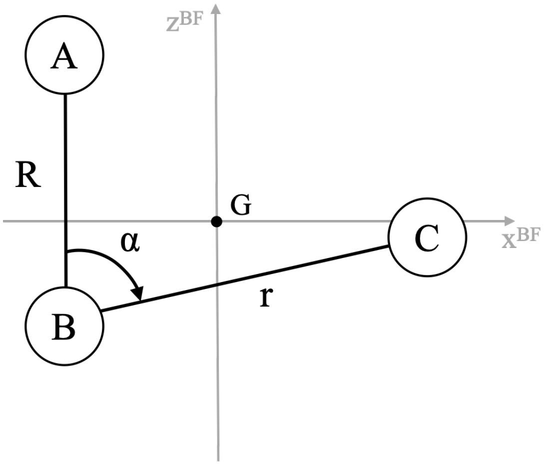

FIGURE 4.13: Body-fixed frame (BF) associated to the ABC molecule. The position vectors **R** and **r** are oriented in a way that z^{BF} is parallel to **R**, and the xy plane of the BF is parallel to the plane generated by **R** and **r**.

If we represent the triatomic molecule ABC with two relative valence position vectors **R** and **r**, and three valence coordinates R, r and α; and let the BF to be oriented in a manner that z^{BF} is parallel to **R**, and the xy plane of the BF is parallel to the plane generated by **R** and **r** as represented in Figure 4.13; then it can shown that the quantum mechanical KEO of the system ($\hat{T}$) has the form [260]:

$$
\begin{aligned}
2\hat{T} = \frac{\hat{p}_R^\dagger \hat{p}_R}{M} + \frac{\hat{p}_r^\dagger \hat{p}_r}{m} + \frac{2\hat{p}_R\hat{p}_r}{\mu}\cos\alpha - \frac{2}{\mu}\left(\frac{\hat{p}_R}{r} + \frac{\hat{p}_r}{R}\right)\sin\alpha\hat{\ell}_y^{BF} \\
+ \left(\frac{1}{MR^2} + \frac{1}{mr^2} - \frac{2\cos\alpha}{Rr\mu}\right)[\hat{\ell}^2]_{E2} + \frac{2\hat{p}_r}{\mu}R\sin\alpha\hat{\mathbf{J}}_y^{BF} \\
+ \frac{2\sin\alpha}{Rr\mu}[(\hat{\ell}_x^{BF} - \hat{\mathbf{J}}_x^{BF})\hat{\mathbf{J}}_z^{BF} + i\hbar\hat{\ell}_y^{BF}] - 2\left(\frac{1}{MR^2} - \frac{\cos\alpha}{Rr\mu}\right)[\hat{\mathbf{J}}\cdot\hat{\ell}]_{BF} + \frac{[\hat{\mathbf{J}}^2]_{BF}}{MR^2} \qquad (4.3)
\end{aligned}
$$

In the above equation all quantities with the hat $(\hat{})$ symbol are operators whose Hermitian adjoint is labeled with a dagger (†) symbol. The ordering of this operators must be respected when computing the corresponding matrix elements. The $\hat{p}_R$ and $\hat{p}_r$ are the momentum operators, $\hat{\mathbf{L}}$ and $\hat{\ell}$ are the angular momentum operators associated with the position vectors **R** and **r** respectively, **J** is the total angular momentum of the system. The subscript E_2 refers to the frame obtained after the first two Euler rotations. The masses of the individual atoms m_A, m_B and m_C are used to generate the reduced masses:

$$
\begin{aligned}
M &= \frac{m_A m_B}{m_A + m_B} \\
m &= \frac{m_B m_C}{m_B + m_C}
\end{aligned}
\qquad (4.4)
$$

The value of $\mu = m_B$ for the valence vectors in Figure 4.13. Analytical matrix elements have been developed for the operator in Equation 4.3. For more details about the subject, the reader is invited to consult the excellent material by Gatti *et al.* [172, 260]. The above KEO is fully supported by the Heidelberg implementation of the MCTDH algorithm [261].

To summarize, the next steps in oder to fulfill the objectives established in Section 1.3 are the development of the aforementioned SOP-FBR PES, and the obtention of the GS distribution of the system using MCTDH. As we have seen in this chapter, the conditions are already created to accomplish this tasks, whose results were not included in the present manuscript due to time reasons. With the GS distribution in hand and the corresponding ZPE, the absorption processes of NO_2 over Pyrene will be completely characterized. The electronic absorption spectrum will be computed as the GS-weighted sum of the individual electronic excitations of the stationary points. The vertical excitations will be obtained with the DMRG method [119], in collaboration with Morgane Vacher from the University of Nantes.

Chapter 5

Conclusions

> *Mais -que tous les enfants le sachent bien- avec de l'ordre, du zèle, du courage, il n'est pas de situations, si périlleuses soient-elles, dont on ne puisse se tirer.*
>
> Jules Verne, *Deux ans de vacances*

In the present dissertation we addressed the full quantum description of interaction between the NO_2 molecule and soot particles. The surface of the soot particles has been modeled as a large PAH molecule, Pyrene in our case. Although drastic in appearance, the previous approximation is sustained by some of the physicochemical properties of soot particles (see Section 1.1). The environmental and human health impact of the system of choice is elevated. In particular, it has been recently shown that in the presence of solar light the target system can lead to formation of HONO, which further decomposes yielding the radical OH, which is one of the strongest oxidants of the atmosphere. After an appropriate bibliographic research to establish the *state-of-the-art* of the subject, a number of (perhaps ambitious) objectives were set in order to organize the research. This objectives are answered with the following conclusions:

- The determination of the adsorption sites of the system and the transition states connecting them.

  The above task was performed with the powerful vdW-TSSCDS, the generalization of the TSSCDS method recently developed in our research group. The vdW-TSSCDS method allows for an automated an efficient exploration of the full configuration space starting from a small set of judiciously chosen initial structures. To the best of our knowledge, our study constitutes the first full dimensional (81D) theoretical description of the interaction between the Pyrene and the NO_2 molecules, as well as the first production application of vdW-TSSCDS. The PES was characterized at the UωB97XD/cc-pVDZ Integral(SuperFineGrid) level of theory, which has been successfully applied to the description of similar systems. The reaction network of the system was elucidated and some interesting topographical features were identified, for example, the presence of branching point in the PES. In addition, some limitations on the low level Electronic Structure calculations were identified. The former can be overcome with the usage of a properly reparametrized semiempirical PES, which better captures the Physics of the problem (*Precook* vdW-TSSCDS).

- The development of a global representation of the interaction potential (PES) in the mathematical form necessary to study the nuclear quantum dynamics of the system.

  This part constituted the bulk of this thesis, involving a large software development activity. Two new methodologies and the associated software packages were developed in order to fulfill this objective, both of them start from the topographical information gathered from (vdW-)TSSCDS. The first one of them is the Specific Reaction Parameter Multigrid POTFIT (SRP-MGPF) method, which is implemented in the SRPTucker package [7]. The former allows for automatically obtaining a PES using a reparametrized semiempirical method. These PES are then tensor decomposed by the MGPF algorithm [8] which permit to perform nuclear quantum dynamics calculations with the Heidelberg implementation of the MCTDH algorithm [9, 10]. The methodology was benchmarked with a number of reference problems, like the *cis-trans* isomerization of the HONO molecule. The second method is the Sum-Of-Products Finite-Basis-Representation (SOP-FBR) [11], which allows for obtaining the PES of the system directly in the mathematical form required by MCTDH. From the basis of a low-rank Sum of Products expansion on the grid, the method replaces the grid-based basis (discrete Schmidt basis) functions by an expansion in terms of orthogonal polynomials. The proposed approach can be seen as an alternative to other NN-based optimization methods. The MCTDH package was modified by us in order to include support for the expansion of the Schmidt basis of choice (series of Chebyshev polynomials in this case). Due to its adequate mathematical form, the integration of SOP-FBR with MCTDH was direct. The method was also benchmarked with quantum dynamical calculations on several reference problems [11].

- The determination of the probability distribution function (ground state) of the system.

  This stage of the project is still ongoing research. At the present moment, an initial SRP-MGPF PES has been developed and is susceptible to improvements. The latter will be used to generate a high quality guess for the corresponding SOP-FBR PES. The reference points for the aforementioned calculation have been already computed. The KEO of the system has been also determined and can be readily implemented in the Heidelberg version of MCTDH. Together with the PES, the GS distribution of the system will be obtained by wave packet relaxation with MCTDH.

- The determination of the adsorption energy of the system and the electronic absorption spectrum.

  The first of these quantities will be directly obtained from the results of the previous stage. Indeed, with the GS distribution in hand and the corresponding ZPE, the absorption processes of NO_2 over Pyrene will be completely characterized. As a first approximation, the electronic absorption spectrum will be computed as the GS-weighted sum of the individual vertical excitations at the minima. These calculations are currently being computed at the DMRG level of theory [119], in collaboration with Morgane Vacher from the University of Nantes.

Appendix A

Coordinate transformation

In this appendix we present the generalities of a code developed to transform Cartesian coordinates to internal coordinates of the Pyr-NO_2 system, although applicable to any triatomic molecule over a surface. The reference Pyrene geometry has been optimized at the ωB97x/cc-pvDZ level of theory (see Chapter 4) and is centered on the canonical Cartesian frame. The Pyrene molecule is considered frozen in its equilibrium geometry and lying on the YZ plane, with its principal axis of rotation aligned with the canonical ones. The code has been written in Python 3.8.5 and uses the SciPy library [262] to perform the conversion of the rotation matrices into Euler angles. For a general discussion about the definition of the latter see Reference [263].

Let us first recall the transformation introduced in the expression 4.1 of Section 4.2.1. The formulated bijection has been implemented in both directions, but for the sake of brevity only the forward (CART2INT) will be described here. The program workflow is completely procedural and is represented in Algorithm 9. The code receives as input an array G of shape $(N, 4)$, i.e. the three Cartesian coordinates and corresponding masses. The previous array is then divided by the `split` procedure into four arrays containing the masses (M_n, M_p) and the Cartesian coordinates (G_n, G_p) of the two fragments. Here the subscripts n, p make reference to the NO_2 and Pyrene molecules respectively. The next step is the determination of the NO_2 internal coordinates with the `no2int` procedure. The above consist in finding the norm and the angle between two Euclidean vectors formed with origin in the nitrogen atom and end in the oxygen atoms.

The centers of mass of the individual fragments are then computed with the procedure `cmass`, and the resulting vectors C_n, C_p are used to translate the molecules to the origin (by subtracting from the initial geometry). The difference of the individual centers of mass is then unpacked into the values of the center of mass of the NO_2[1]. The next

---

[1] To be more precise, this is the center of mass of the NO_2 once the Pyrene is centered on the origin.

```
Algorithm 9: CART2INT
Result: $X_{cm}, Y_{cm}, Z_{cm}, \alpha, \beta, \gamma, r_1, r_2, \theta$
Input: $G$ Cartesian coordinates of the system;
Function paxes (M,G):
    $I_{ij} \leftarrow \sum_{k=1}^{N} M[k](\|G[k]\|^2\delta_{ij} - G[k,i] \cdot G[k,j])$;
    $\Lambda, A \leftarrow$ eigendec($I$);
return $A$
$M_n, G_n, M_p, G_p \leftarrow$ split($G$);
$r_1, r_2, \theta \leftarrow$ no2int($G_n$);
$C_n \leftarrow$ cmass($M_n, G_n$);
$C_p \leftarrow$ cmass($M_p, G_p$);
$G_n \leftarrow G_n - C_n$;
$G_p \leftarrow G_p - C_p$;
$X_{cm}, Y_{cm}, Z_{cm} \leftarrow$ unpack($C_n - C_p$);
$A_n \leftarrow$ paxes($M_n, G_n$);
$A_p \leftarrow$ paxes($M_p, G_p$);
$R \leftarrow A_n \times A_p^T$;
$\alpha, \beta, \gamma \leftarrow$ aseuler($R$);
```

step is the determination of the rotation matrix (R) between the coordinate frames of both molecules. Our approach here consisted in finding the principal axis of rotation (contained in the array A) of the molecules by diagonalization of the inertia tensor I. This operations are implemented in the `paxes` function. Finally, the rotation matrix is transformed into Euler angles using the `aseuler` procedure (as implemented in SciPy).

Although simple in concept, the development of the above algorithm can be a tedious process. Indeed, care must be take in defining (and respecting) suitable conventions for the sign of the rotation matrices and for the ordering of the frames in the Euler angles. In practice, many of the well tested and powerful *eigensolvers* available yield eigenvectors which are defined up to a multiplicative constant, which implies that after normalization a spurious negative sign might appear. This is in frank contradiction whit the first of the aforementioned conditions, and an extra verification of the axis orientations obtained with `paxes` must be introduced.

Appendix B

SRPTucker software

In the present appendix the SRPTucker software package will be described in detail. We will assume that the reader if familiar with the SRP-MGPF method [7], that was introduced in Section 3.2. SRPTucker have been developed in Python 3.8.5, and has been tested for back compatibility until Python 3.6.0. The actual implementation of the code is written as combination of the procedural and object oriented (OOP) paradigms. Parallelization has been ensured, as the bottleneck of the process is an *embarrassingly parallel* problem (*vide infra*). Note that due to the specificities of Python (in particular the global interpreter lock or GIL[1]), the parallelization is achieved by using subprocesses and not threads, i.e. in distributed memory.

The workflow of SRPTucker is represented in Algorithm 10. We will use indistinctly the term meta-algorithm to refer to SRPTucker, as it uses the results of individual algorithms. The process starts with the definition of a set of guess parameters x_{guess}, included in a single array. This parameters are bounded to the underlaying semiempirical level of theory of choice (see Section 2.2.3 for more details). Additionally, the number of reference geometries N_g and a threshold ϵ for the computed RMSE must be defined. The geometries are generated by the `geogen` procedure, which employs the systematic many body like expansion scheme described in [7]. It is important to notice that this step relays on the stationary points determined by (vdW-)TSSCDS, and in the selection of an appropriated *ab initio* level of theory for the calculations, which must be performed in an independent Electronic Structure package. In our particular case, we used the Gaussian [248] and MOLPRO [264] software packages. Finally, the procedure yields the reference geometries G_{ab} and corresponding energies $E_{ab}(G_{ab})$ contained in arrays of suitable shape.

---

[1] The memory management in the CPython implementation is not thread-safe. The function of the GIL is then to prevent that multiple threads execute Python bytecode at the same time.

A fundamental step in any optimization process is the definition of the objective (or target) function. In the case of SRPTucker, this function is not analytical and consequently, analytical derivatives are not available for it. Indeed, the values of the energies are obtained with the `mopac` procedure, which is a python interface to the MOPAC software package [247]. As it was discussed in Sections 2.2.1.1 and 2.2.3, the semiempirical energies are calculated by solving a set of coupled integro-differential equations with a SCF procedure. The computations of the individual single point energies is parallelized by the `multiprocessing` procedure[2]. In this way, our objective function `target`, which is nonlinear and *black box*, returns the possibly weighted L_2 norm (ρ) of the difference between the *ab initio* and semiempirical energies.

Algorithm 10: SRPTucker

```
Result: $x_{opt}$
Input: $x_{guess}$ guess semiempirical parameters, $N_g$ number
  of geometries, $\epsilon$ threshold, $N$ maximum number of
  iterations ;
$k \leftarrow 0$;
$x_0 \leftarrow x_{guess}$;
$G_{ab}, E_{ab} \leftarrow$ geogen($N_g$);
Function target ($x_{val}$):
    $E_{srp} \leftarrow$ multiprocessing(mopac)($G_{ab}, x_{val}$);
    $\rho \leftarrow \|E_{ab} - E_{srp}\|_{L_2}$;
return $\rho$
repeat
    $\rho, x_{k+1} \leftarrow$ MLSL(BOBYQA, target($x_k$));
    $k \leftarrow k + 1$;
until $\rho < \epsilon \vee k < N$;
$x_{opt} \leftarrow x_k$
```

The next step consists on a series of global optimizations using the MLSL algorithm (see Section 2.6), which in turn requires the BOBYQA [265] local optimizer. The meta-algorithm is not restricted to the previous combination, and many other algorithms can be explored as suggested by the *No Free Lunch* Theorems for Optimization [220]. All the optimizers employed in this work are included in the Python implementation of the NLopt library [266]. Suitable termination conditions must be set for both the main loop and the optimization with MLSL. In our case, we controlled the values of ρ with the threshold ϵ, and the total number of iterations N. The optimal parameter set x_{opt} defines the SRP PES of the system, which is afterwards tensor-decomposed using the MGPF algorithm [8] using suitable FORTRAN external process calls.

---

[2] This procedure uses the homonymous Python library, but in the pseudocode makes reference to a larger set of logical steps.

Appendix C

SOP-FBR software

In this appendix the SOP-FBR software package will be discussed in more detail. It will be assumed that the reader is familiar with corresponding method [11] that was introduced in Section 3.2. SOP-FBR have been implemented in Python in Python 3.8.5, and was tested for back compatibility until Python 3.6.0. The code is written mainly in an object oriented (OOP) fashion. In this case, the meta-algorithm makes heavy use of the SciPy [262] and TensorLy [267] libraries for the optimization and tensor decomposition algorithms. Parallelization is provided by the powerful BLAS interfaces of SciPy, which automatically optimize the code to run as fast as possible in the provided architecture. This feature is particularly convenient for matrix and tensor operations.

The parameters in the SOP-FBR *ansatz* can be optimized in several ways. The simplest strategy is the direct usage of a global optimization algorithm (like the Basinhopping procedure described in Section 2.6). However, for systems with dimensionality bigger than 3D the parameter space might be too large, and in addition, the condition number of the Chebyshev series coefficients is large (see [11] for a detailed discussion). Without an appropriated guess, the optimization might also become a lengthy process. To address this problem, we have employed a more elaborated strategy, namely a decoupled[1] optimization of the core tensor and the SPP matrices. The proposed workflow is represented in Algorithm 11.

The process starts by defining the guess parameters, which are included in a flattened array $x_{guess} = \mathrm{vec}(B) || \mathrm{vec}(C)$ where B, C represent a tensor containing the Chebyshev series coefficients for all factor matrices (see Section 3.1.1.2) and the core tensor respectively. As outlined above, the selection of an initial parameter set reasonably close to the optimal value can significantly speed up the process. The previous task is as

---

[1] Not to be confused with uncoupled optimization, in which the two target functions are optimized using independent underlying physical principles

hard as the optimization itself, but fortunately, we have already developed a method can offer suitable guesses: SRPTucker (see Chapter 3 and Appendix B). Additionally, the dimensionality of the problem D, an array M of length D containing the number of basis functions for each DOF, and an array T of length D containing the degree of the Chebyshev series for each DOF[2] must be inputed.

Algorithm 11: SOP-FBR

```
Result: $x_{opt}$
Input: $x_{guess}$ guess parameters, $D$ dimensionality, $M$
  number of basis functions, $T$ degree of Chebyshev series,
  $N_g$ number of geometries, $\epsilon$ threshold, N maximum
  number of iterations ;
$k \leftarrow 0$;
$x_0 \leftarrow x_{guess}$;
$G_{ab}, E_{ab} \leftarrow$ geogen($N_g$);
Function sopfbr ($B, C$):
    $l \leftarrow 0$;
    for $k \leftarrow 0$ to $D$ do
        for $j \leftarrow 0$ to $M[k]$ do
            for $i \leftarrow 0$ to $G_{ab}[:, k]$ do
                $U_{ij}^{(k)} \leftarrow$ chebyshev($G_{ab}[i, k], B(l : l + T[k])$);
            end
            $l \leftarrow l + T[k]$;
        end
    end
    $E_{sop} \leftarrow C \times_1 U^{(1)} \cdots \times_D U^{(D)}$;
return $E_{sop}$
Function target ($B, C$):
    $E_{sop} \leftarrow$ sopfbr($B, C$);
    $\rho \leftarrow \|E_{ab} - E_{sop}\|_{L_2}$;
return $\rho$
repeat
    $B, C \leftarrow$ split($x_k, T \times M$);
    $B \leftarrow$ BFGS(target($B, \bar{C}$));
    $\rho, C \leftarrow$ Powell(target($\bar{B}, C$));
    $x_{k+1} \leftarrow$ concatenate($B, C$);
    $k \leftarrow k + 1$;
until $\rho < \epsilon \vee k < N$;
$x_{opt} \leftarrow x_k$
```

The `geogen` procedure used by SOP-FBR does not differ substantially from the one of SRPTucker: it yields N_g reference geometries G_{ab} generated in a systematic manner. The main discrepancy here is that the former might need the reference energies E_{ab} to a

---

[2] We assume that the degree of the Chebyshev series is the same for each basis functions and for each DOF. This is not a limitation of the method, but rather a simplification in the discussion.

tensor of suitable shape. We have opted to impose this condition because the implementation of the function `sopfbr` that evaluates the analytical SOP-FBR *ansatz* in the set G_{ab} only can use the full parallelization power of SciPy if tensors are used (in contrast to flattened arrays). To overcome this drawback of the Python application, extension routines are being developed in FORTRAN. Regarding this previous function, it works by first computing the factor matrices $U^{(k)}$ evaluating the corresponding Chebyshev series (implemented in the procedure `chebyshev`), and then performing a series of tensor n-mode products. The obtained energy E_{sop} is then used by the objective function `target`, which returns the L_2 norm (ρ) of the difference between the former and the reference energy.

The main loop of the meta-algorithm starts by splitting and reshaping the flattened the parameters of the k-th iteration (x_k) into the tensors B and C. This step is performed with the `split` procedure. Then the decoupled scheme begins alternating the optimization of the B tensor while keeping C fixed[3], and the optimization of C while keeping B fixed. Since B is usually smaller than C, we have used the BFGS algorithm for the former and the Powell algorithm (which is computationally less demanding) for the latter. As termination conditions for the loop, we used the threshold value ϵ for RMSE error ρ, and the total number of iterations N. The resulting optimal parameter set x_{opt} is obtained by flattening and concatenating B and C, which is achieved with the `concatenate` procedure. The optimized analytical *ansatz* is already in the Tucker form needed by MCTDH, and consequently, a straightforward interfacing with the latter has been accomplished [11]. In addition, the Heidelberg implementation of the MCTDH software package has been modified by us in order to include the support for the Schmidt basis (Chebyshev polynomials in the above example) employed by SOP-FBR.

---

[3] Here we have used the notation $\bar{C}$ (bar) to indicate that the quantity is kept constant in the optimization process.

Bibliography

[1] Chun Guan, Xinling Li, Wugao Zhang, and Zhen Huang. Identification of nitration products during heterogeneous reaction of NO_2 on soot in the dark and under simulated sunlight. *The Journal of Physical Chemistry A*, 121(2):482–492, 2017.

[2] Maria Eugenia Monge, Barbara D'Anna, Linda Mazri, Anne Giroir-Fendler, Markus Ammann, DJ Donaldson, and Christian George. Light changes the atmospheric reactivity of soot. *Proceedings of the National Academy of Sciences*, 107 (15):6605–6609, 2010.

[3] Ann M Holloway and Richard P Wayne. *Atmospheric chemistry*. Royal Society of Chemistry, 2015.

[4] Sabine Kopec, Emilio Martínez-Núñez, Juan Soto, and Daniel Peláez. vdW-TSSCDS an automated and global procedure for the computation of stationary points on intermolecular potential energy surfaces. *International Journal of Quantum Chemistry*, 119(21):e26008, 2019.

[5] Emilio Martínez-Núñez. An automated method to find transition states using chemical dynamics simulations. *Journal of computational chemistry*, 36(4):222–234, 2015.

[6] Emilio Martínez-Núñez. An automated transition state search using classical trajectories initialized at multiple minima. *Physical Chemistry Chemical Physics*, 17 (22):14912–14921, 2015.

[7] Ramón Lorenzo Panadés-Barrueta, Emilio Martínez-Núñez, and Daniel Peláez. Specific reaction parameter multigrid potfit (SRP-MGPF): Automatic generation of sum-of-products form potential energy surfaces for quantum dynamical calculations. *Frontiers in chemistry*, 7:576, 2019.

[8] D. Peláez and H.-D. Meyer. The multigrid POTFIT (MGPF) method: Grid representations of potentials for quantum dynamics of large systems. 138:014108, 2013.

[9] H.-D. Meyer, U. Manthe, and L. S. Cederbaum. The multi-configurational time-dependent Hartree approach. 165:73–78, 1990.

[10] G. A. Worth, M. H. Beck, A. Jäckle, and H.-D. Meyer. *The MCTDH Package, Version 8.1.* University of Heidelberg, Heidelberg, 2000.

[11] Ramón Lorenzo Panadés-Barrueta and Daniel Peláez. Low-rank sum-of-products finite-basis-representation (SOP-FBR) of potential energy surfaces. *The Journal of Chemical Physics*, (in review), 2020.

[12] Michel Hénon and Carl Heiles. The applicability of the third integral of motion: some numerical experiments. *The Astronomical Journal*, 69:73, 1964.

[13] Oleg L. Polyansky, Per Jensen, and Jonathan Tennyson. The potential energy surface of $H_2\ ^{16}O$. 105:6490–6497, 1996. PES, water.

[14] F. Richter, M. Hochlaf, P. Rosmus, F. Gatti, and H.-D. Meyer. A study of mode–selective trans–cis isomerisation in HONO using ab initio methodology. 120:1306–1317, 2004.

[15] Frank Arens, Lukas Gutzwiller, Urs Baltensperger, Heinz W Gäggeler, and Markus Ammann. Heterogeneous reaction of no2 on diesel soot particles. *Environmental science & technology*, 35(11):2191–2199, 2001.

[16] Harald Muckenhuber and Hinrich Grothe. The heterogeneous reaction between soot and NO_2 at elevated temperature. *Carbon*, 44(3):546–559, 2006.

[17] Tahrizi Andana, Marco Piumetti, Samir Bensaid, Nunzio Russo, and Debora Fino. Heterogeneous mechanism of nox-assisted soot oxidation in the passive regeneration of a bench-scale diesel particulate filter catalyzed with nanostructured equimolar ceria-praseodymia. *Applied Catalysis A: General*, 583:117136, 2019.

[18] C. A. Pope and D. W. Dockery. Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air and Waste Management Association*, 56(6): 709–742, 2006. ISSN 21622906. doi: 10.1080/10473289.2006.10464485.

[19] B. A. Maher, I. A.M. Ahmed, V. Karloukovski, D. A. MacLaren, P. G. Foulds, D. Allsop, D. M.A. Mann, R. Torres-Jardón, and L. Calderon-Garciduenas. Magnetite pollution nanoparticles in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 113(39):10797–10801, 2016. ISSN 10916490. doi: 10.1073/pnas.1605941113.

[20] H. Bové, E. Bongaerts, E. Slenders, E. M. Bijnens, N. D. Saenen, W. Gyselaers, P. V. Eyken, M. Plusquin, M. B. J. Roeffaers, M. Ameloot, and T. S. Nawrot. Ambient black carbon particles reach the fetal side of human placenta. *Nature Communications*, 10:1–7, 2019. doi: 10.1038/s41467-019-11654-3.

[21] Henning Rodhe, Christer Persson, and Ove Åkesson. An investigation into regional transport of soot and sulfate aerosols. *Atmospheric Environment (1967)*, 6(9):675–693, 1972.

[22] Veerabhadran Ramanathan and Gregory Carmichael. Global and regional climate changes due to black carbon. *Nature geoscience*, 1(4):221–227, 2008.

[23] Øivind Hodnebrog, Gunnar Myhre, and Bjørn H Samset. How shorter black carbon lifetime alters its climate effect. *Nature communications*, 5(1):1–7, 2014.

[24] Jordi Sunyer, Mikel Esnaola, Mar Alvarez-Pedrerol, Joan Forns, Ioar Rivas, Mònica López-Vicente, Elisabet Suades-González, Maria Foraster, Raquel Garcia-Esteban, Xavier Basagaña, et al. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. *PLoS Med*, 12(3):e1001792, 2015.

[25] G. Grimmer. *Environmental carcinogens: Polycyclic aromatic hydrocarbons: Chemistry, occurrence, biochemistry, carcinogenicity.* CRC Press, 2018. ISBN 9781351080200. doi: 10.1201/9781351071758.

[26] Meyer M Melicow. Percivall Pott (1713–1788) 200th anniversary of first report of occupation-induced cancer of scrotum in chimney sweepers (1775). *Urology*, 6(6):745–749, 1975.

[27] Keith Denkler. Sir Percivall Pott, sir james paget, and soot cancer of the hand. *The Lancet*, 364(9434):582, 2004.

[28] Kristin A Miller, David S Siscovick, Lianne Sheppard, Kristen Shepherd, Jeffrey H Sullivan, Garnet L Anderson, and Joel D Kaufman. Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine*, 356(5):447–458, 2007.

[29] PR Buseck, K Adachi, A Gelencsér, É Tompa, and M Pósfai. Are black carbon and soot the same? *Atmospheric Chemistry & Physics Discussions*, 12(9), 2012.

[30] Song Han, Xiaoxia Li, Fengguang Nie, Mo Zheng, Xiaolong Liu, and Li Guo. Revealing the initial chemistry of soot nanoparticle formation by ReaxFF molecular dynamics simulations. *Energy & Fuels*, 31(8):8434–8444, 2017.

[31] Árpád B Palotás, Lenore C Rainey, Christian J Feldermann, Adel F Sarofim, and John B Vander Sande. Soot morphology: An application of image analysis in high-resolution transmission electron microscopy. *Microscopy research and technique*, 33(3):266–278, 1996.

[32] Viktória Kovács Kis, Mihály Pósfai, and János L Lábár. Nanostructure of atmospheric soot particles. *Atmospheric Environment*, 40(29):5533–5542, 2006.

[33] Michael Frenklach and Hai Wang. Detailed modeling of soot particle nucleation and growth. In *Symposium (International) on Combustion*, volume 23, pages 1559–1566. Citeseer, 1991.

[34] Ivar SA Isaksen, Claire Granier, G Myhre, TK Berntsen, Stig B Dalsøren, Michael Gauss, Zbigniew Klimont, Rasmus Benestad, Philippe Bousquet, W Collins, et al. Atmospheric composition change: climate–chemistry interactions. *Atmospheric Environment*, 43(33):5138–5192, 2009.

[35] Robert C Chapleski, Yafen Zhang, Diego Troya, and John R Morris. Heterogeneous chemistry and reaction dynamics of the atmospheric oxidants, O_3, NO_3, and OH, on organic surfaces. *Chemical Society Reviews*, 45(13):3731–3746, 2016.

[36] Manuel T Lerdau, J William Munger, and Daniel J Jacob. The NO_2 flux conundrum. *Science*, 289(5488):2291–2293, 2000.

[37] JL Devalia, C Rusznak, MJ Herdman, CJ Trigg, RJ Davies, and H Tarraf. Effect of nitrogen dioxide and sulphur dioxide on airway response of mild asthmatic patients to allergen inhalation. *The Lancet*, 344(8938):1668–1671, 1994.

[38] M Ammann, M Kalberer, DT Jost, L Tobler, E Rössler, D Piguet, HW Gäggeler, and U Baltensperger. Heterogeneous production of nitrous acid on soot in polluted air masses. *Nature*, 395(6698):157–160, 1998.

[39] Narumol Jariyasopit, Melissa McIntosh, Kathryn Zimmermann, Janet Arey, Roger Atkinson, Paul Ha-Yeon Cheong, Rich G Carter, Tian-Wei Yu, Roderick H Dashwood, and Staci L Massey Simonich. Novel nitro-PAH formation from heterogeneous reactions of PAHs with NO_2, NO_3/N_2O_5, and OH radicals: Prediction, laboratory studies, and mutagenicity. *Environmental science & technology*, 48(1):412–419, 2014.

[40] Pascale Desgroux, Xavier Mercier, and Kevin A Thomson. Study of the formation of soot and its precursors in flames using optical diagnostics. *Proceedings of the Combustion Institute*, 34(1):1713–1738, 2013.

[41] Junteng Wu, Alessandro Faccinetto, Symphorien Grimonprez, Sébastien Batut, Jérôme Yon, Pascale Desgroux, and Denis Petitprez. Influence of the dry aerosol particle size distribution and morphology on the cloud condensation nuclei activation. An experimental and theoretical investigation. *Atmospheric Chemistry & Physics*, 20(7), 2020.

[42] Dumitru Duca, Cornélia Irimiea, Alessandro Faccinetto, JA Noble, Marin Vojkovic, Yvain Carpentier, Ismael K Ortega, Claire Pirim, and Cristian Focsa. On the benefits of using multivariate analysis in mass spectrometric studies of combustion-generated aerosols. *Faraday discussions*, 218:115–137, 2019.

[43] Shing-Kuo Shih, Sigrid D Peyerimhoff, and Robert J Buenker. Theoretical study of the vertical electronic spectrum of NO_2. *Chemical Physics Letters*, 46(2):201–207, 1977.

[44] S Mahapatra, H Köppel, LS Cederbaum, P Stampfuss, and W Wenzel. Nonadiabatic wave packet dynamics on the coupled X2A1/A2B2 electronic states of NO_2 based on new ab initio potential energy surfaces. *Chemical Physics*, 259(2): 211–226, 2000.

[45] Saikat Mukherjee, Bijit Mukherjee, Subhankar Sardar, and Satrajit Adhikari. Ab initio constructed diabatic surfaces of NO_2 and the photodetachment spectra of its anion. *The Journal of chemical physics*, 143(24):244307, 2015.

[46] E Haller, H Köppel, and LS Cederbaum. On the statistical behaviour of molecular vibronic energy levels. *Journal of Molecular Spectroscopy*, 111(2):377–397, 1985.

[47] Wolfgang Schneider, Geert K Moortgat, Geoffrey S Tyndall, and John P Burrows. Absorption cross-sections of NO_2 in the uv and visible region (200–700 nm) at 298 k. *Journal of photochemistry and photobiology A: Chemistry*, 40(2-3):195–217, 1987.

[48] Daniel Peláez Ruiz. *Mecanismos de Reacción Fotoquímicos de Nitro-y Nitrosocompuestos desde las Aproximaciones CASSCF y CASPT2: Relevancia de los Procesos No Adiabáticos.* PhD thesis, Universidad de Málaga, 2007.

[49] S Yu Grebenshchikov, C Beck, H Flöthman, R Schinke, and S Kato. Unimolecular dissociation of NO_2. I. classical trajectory and statistical calculations on a global potential energy surface. *The Journal of chemical physics*, 111(2):619–632, 1999.

[50] Reinhard Schinke, S Yu Grebenshchikov, and Hui Zhu. The photodissociation of NO_2 in the second absorption band: Ab initio and quantum dynamics calculations. *Chemical Physics*, 346(1-3):99–114, 2008.

[51] Fabrizio Santoro and Carlo Petrongolo. Nonadiabatic wave packet dynamics of NO_2 on the $\tilde{X}$2A'/$\tilde{A}$2A' conical intersection. *The Journal of chemical physics*, 110 (9):4419–4427, 1999.

[52] Uwe Manthe, H-D Meyer, and LS Cederbaum. Multiconfigurational time-dependent hartree study of complex dynamics: Photodissociation of NO_2. *The Journal of chemical physics*, 97(12):9062–9071, 1992.

[53] Chong Han, Yongchun Liu, and Hong He. Heterogeneous photochemical aging of soot by NO_2 under simulated sunlight. *Atmospheric environment*, 64:270–276, 2013.

[54] D Perner and U Platt. Detection of nitrous acid in the atmosphere by differential optical absorption. *Geophysical Research Letters*, 6(12):917–920, 1979.

[55] Trevor C VandenBoer, Cora J Young, Ranajit K Talukdar, Milos Z Markovic, Steven S Brown, James M Roberts, and Jennifer G Murphy. Nocturnal loss and daytime source of nitrous acid through reactive uptake and displacement. *Nature Geoscience*, 8(1):55–60, 2015.

[56] Jinzhu Ma, Yongchun Liu, Chong Han, Qingxin Ma, Chang Liu, and Hong He. Review of heterogeneous photochemical reactions of NOy on aerosol. A possible daytime source of nitrous acid (HONO) in the atmosphere. *Journal of Environmental Sciences*, 25(2):326–334, 2013.

[57] Jonathan Raff. Air quality: Night-time sinks, daytime sources. *Nature Geoscience*, 8(1):5–7, 2015.

[58] Jörg Kleffmann. Daytime sources of nitrous acid (HONO) in the atmospheric boundary layer. *ChemPhysChem*, 8(8):1137–1144, 2007.

[59] Barbara J Finlayson-Pitts. Introductory lecture: atmospheric chemistry in the anthropocene. *Faraday discussions*, 200:11–58, 2017.

[60] PC Chan. NTP technical report on toxicity studies of 1-nitropyrene (Cas No. 5522-43-0) administered by inhalation to F344/N rats. toxicity report series. Technical report, National Toxicology Program, Research Triangle Park, NC (United States), 1996.

[61] Chad Orzel. *Breakfast with Einstein: The exotic physics of everyday objects*. BenBella Books, 2018.

[62] C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Mécanique Quantique - Tome 1: Nouvelle édition*. EDP Sciences, 2018. ISBN 9782759822881. URL https://books.google.de/books?id=3jRyDwAAQBAJ.

[63] Erwin Schrödinger. Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinem. *Annalen der Physik*, 384(8):734–756, 1926.

[64] Amritanshu Prasad. An easy proof of the Stone-von Neumann-Mackey theorem. *Expositiones Mathematicae*, 29(1):110–118, 2011.

[65] HS Perlman. Equivalence of the Schrödinger and Heisenberg pictures. *Nature*, 204 (4960):771–772, 1964.

[66] Richard Phillips Feynman. Space-time approach to non-relativistic quantum mechanics. In *Feynman's Thesis—A New Approach To Quantum Theory*, pages 71–109. World Scientific, 2005.

[67] Paul Adrien Maurice Dirac. A new notation for quantum mechanics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 35, pages 416–418. Cambridge University Press, 1939.

[68] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory.* Courier Corporation, 2012.

[69] Frank Jensen. *Introduction to computational chemistry.* John wiley & sons, 2017.

[70] Fabien Gatti, Benjamin Lasorne, Hans-Dieter Meyer, and André Nauts. *Applications of quantum dynamics in chemistry*, volume 98. Springer, 2017.

[71] Henry Pierce Stapp. The Copenhagen interpretation. *American Journal of Physics*, 40(8):1098–1116, 1972.

[72] Erwin Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23(49):823–828, 1935.

[73] Graham A Worth and Lorenz S Cederbaum. Beyond Born-Oppenheimer: molecular dynamics through a conical intersection. *Annu. Rev. Phys. Chem.*, 55:127–158, 2004.

[74] Max Born and Kun Huang. *Dynamical theory of crystal lattices.* Clarendon press, 1954.

[75] Michael Baer. *Beyond Born-Oppenheimer: electronic nonadiabatic coupling terms and conical intersections.* John Wiley & Sons, 2006.

[76] Wolfgang Domcke, David R Yarkony, and Horst Köppel. *Conical intersections: theory, computation and experiment*, volume 17. World Scientific, 2011.

[77] Axel D Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of chemical physics*, 140(18):18A301, 2014.

[78] Douglas Rayner Hartree and W Hartree. Self-consistent field, with exchange, for beryllium. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 150(869):9–33, 1935.

[79] Pablo Echenique and José Luis Alonso. A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry. *Molecular Physics*, 105(23-24):3057–3098, 2007.

[80] Walther Gerlach and Otto Stern. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik*, 9(1):349–352, 1922.

[81] Robert S Mulliken. Electronic structures of polyatomic molecules and valence. ii. general considerations. *Physical Review*, 41(1):49, 1932.

[82] Tjalling Koopmans. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms. *Physica*, 1(1-6):104–113, 1934.

[83] Wolfgang Pauli. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik*, 31(1):765–783, 1925.

[84] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Quantum mechanics: non-relativistic theory*, volume 3. 2013.

[85] Jochen Schirmer. *Many-Body Methods for Atoms, Molecules and Clusters*. Springer, 2018.

[86] Mariano Giaquinta and Stefan Hildebrandt. *Calculus of variations II*, volume 311. Springer Science & Business Media, 2013.

[87] James Daniel Whitfield, Peter John Love, and Alán Aspuru-Guzik. Computational complexity in electronic structure. *Physical Chemistry Chemical Physics*, 15(2):397–411, 2013.

[88] Clemens Carel Johannes Roothaan. New developments in molecular orbital theory. *Reviews of modern physics*, 23(2):69, 1951.

[89] GG Hall. The molecular orbital theory of chemical valency VIII. a method of calculating ionization potentials. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 205(1083):541–552, 1951.

[90] Per-Olov Löwdin. On the nonorthogonality problem. In *Advances in quantum chemistry*, volume 5, pages 185–199. Elsevier, 1970.

[91] Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2):151–177, 1957.

[92] Axel D Becke. Density functionals for static, dynamical, and strong correlation. *The Journal of Chemical Physics*, 138(7):074109, 2013.

[93] Alice Cuzzocrea, Anthony Scemama, Wim J Briels, Saverio Moroni, and Claudia Filippi. Variational principles in quantum Monte Carlo: the troubled story of variance minimization. *Journal of Chemical Theory and Computation*, 2020.

[94] Seymour Michael Blinder and James E House. *Mathematical Physics in Theoretical Chemistry*. Elsevier, 2018.

[95] John C Slater. Atomic shielding constants. *Physical Review*, 36(1):57, 1930.

[96] Frank Jensen. Polarization consistent basis sets: Principles. *The Journal of Chemical Physics*, 115(20):9113–9125, 2001.

[97] Karl N Kirschner, Dirk Reith, and Wolfgang Heiden. The performance of Dunning, Jensen, and Karlsruhe basis sets on computing relative energies and geometries. *Soft Materials*, pages 1–15, 2020.

[98] Ryan M Richard, Brandon W Bakr, and C David Sherrill. Understanding the many-body basis set superposition error: Beyond Boys and Bernardi. *Journal of chemical theory and computation*, 14(5):2386–2400, 2018.

[99] SF Boys and FJMP Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970.

[100] Hans Martin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

[101] RB Gerber, D Shemesh, ME Varner, J Kalinowski, and B Hirshberg. Ab initio and semi-empirical molecular dynamics simulations of chemical reactions in isolated molecules and in clusters. *Physical Chemistry Chemical Physics*, 16(21):9760–9775, 2014.

[102] Michael JS Dewar and Walter Thiel. A semiempirical model for the two-center repulsion integrals in the nddo approximation. *Theoretica chimica acta*, 46(2): 89–104, 1977.

[103] Tamara Husch, Alain C Vaucher, and Markus Reiher. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *International journal of quantum chemistry*, 118(24):e25799, 2018.

[104] John A Pople, D Po Santry, and Gerald A Segal. Approximate self-consistent molecular orbital theory. I. invariant procedures. *The Journal of Chemical Physics*, 43(10):S129–S135, 1965.

[105] James JP Stewart. Optimization of parameters for semiempirical methods II. applications. *Journal of computational chemistry*, 10(2):221–264, 1989.

[106] James JP Stewart. Optimization of parameters for semiempirical methods VI: more modifications to the nddo approximations and re-optimization of parameters. *Journal of molecular modeling*, 19(1):1–32, 2013.

[107] Ramón Lorenzo Panadés-Barrueta, Oriol Vendrell, and Daniel Peláez-Ruiz. Non-adiabatical quantum dynamical study of the lowest-lying electronic excitations of NO_2 using explicit electron dynamics. Deutsche Akademische Austauschdienst Short Reseach Grant, .

[108] Werner Kutzelnigg and Pasquale von Herigonte. Electron correlation at the dawn of the 21st century. In *Advances in quantum chemistry*, volume 36, pages 185–229. Elsevier, 2000.

[109] Egil A Hylleraas. Über den Grundterm der Zweielektronenprobleme von H^-, He, Li^+, Be^{++} usw. *Zeitschrift für Physik*, 65(3-4):209–225, 1930.

[110] Bruno Klahn and Werner A Bingel. The convergence of the Rayleigh-Ritz Method in quantum chemistry. *Theoretica chimica acta*, 44(1):9–26, 1977.

[111] Elda Rossi, Gian Luigi Bendazzoli, Stefano Evangelisti, and Daniel Maynau. A full-configuration benchmark for the N_2 molecule. *Chemical physics letters*, 310 (5-6):530–536, 1999.

[112] Wibe A De Jong, Eric Bylaska, Niranjan Govind, Curtis L Janssen, Karol Kowalski, Thomas Müller, Ida MB Nielsen, Hubertus JJ van Dam, Valera Veryazov, and Roland Lindh. Utilizing high performance computing for chemistry: parallel computational chemistry. *Physical Chemistry Chemical Physics*, 12(26):6896–6920, 2010.

[113] C David Sherrill and Henry F Schaefer III. The configuration interaction method: Advances in highly correlated approaches. In *Advances in quantum chemistry*, volume 34, pages 143–269. Elsevier, 1999.

[114] Péter R Surján. *Second quantized approach to quantum chemistry: an elementary introduction*. Springer Science & Business Media, 2012.

[115] Lucjan Piela. *Ideas of quantum chemistry*. Elsevier, 2013.

[116] Per-Åke Malmqvist and Björn O Roos. The CASSCF state interaction method. *Chemical physics letters*, 155(2):189–194, 1989.

[117] Per Åke Malmqvist, Alistair Rendell, and Bjoern O Roos. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *Journal of Physical Chemistry*, 94(14):5477–5482, 1990.

[118] Konstantinos D Vogiatzis, Dongxia Ma, Jeppe Olsen, Laura Gagliardi, and Wibe A De Jong. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *The Journal of chemical physics*, 147(18):184111, 2017.

[119] Sebastian Wouters and Dimitri Van Neck. The density matrix renormalization group for ab initio quantum chemistry. *The European Physical Journal D*, 68(9): 272, 2014.

[120] Hans-Joachim Werner and Peter J Knowles. An efficient internally contracted multiconfiguration–reference configuration interaction method. *The Journal of chemical physics*, 89(9):5803–5814, 1988.

[121] Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Physical review*, 46(7):618, 1934.

[122] Dieter Cremer. Møller–plesset perturbation theory: from small molecule methods to methods for thousands of atoms. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(4):509–530, 2011.

[123] Reinhold F Fink. Why does MP_2 work? *The Journal of chemical physics*, 145(18): 184101, 2016.

[124] Kerstin Andersson, Per Aake Malmqvist, Björn O Roos, Andrzej J Sadlej, and Krzysztof Wolinski. Second-order perturbation theory with a CASSCF reference function. *Journal of Physical Chemistry*, 94(14):5483–5488, 1990.

[125] James Finley, Per-Åke Malmqvist, Björn O Roos, and Luis Serrano-Andrés. The multi-state CASPT2 method. *Chemical physics letters*, 288(2-4):299–306, 1998.

[126] Rodney J Bartlett. Coupled-cluster approach to molecular structure and spectra: a step toward predictive quantum chemistry. *The Journal of Physical Chemistry*, 93(5):1697–1708, 1989.

[127] Rodney J Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291, 2007.

[128] Julien Toulouse. Introduction to perturbation theory and coupled-cluster theory for electron correlation, 2019. URL: `https://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction_pt_cc.pdf`.

[129] Troy Van Voorhis and Martin Head-Gordon. A nonorthogonal approach to perfect pairing. *The Journal of Chemical Physics*, 112(13):5633–5638, 2000.

[130] James B Robinson and Peter J Knowles. Approximate variational coupled cluster theory. *The Journal of chemical physics*, 135(4):044113, 2011.

[131] Krishnan Raghavachari, Gary W Trucks, John A Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989.

[132] Monika Kodrycka and Konrad Patkowski. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *The Journal of Chemical Physics*, 151(7):070901, 2019.

[133] Gerald Knizia, Thomas B Adler, and Hans-Joachim Werner. Simplified CCSD(T)-F12 methods: Theory and benchmarks. *The Journal of chemical physics*, 130(5): 054104, 2009.

[134] Rick A Kendall and Herbert A Früchtl. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theoretical Chemistry Accounts*, 97(1-4):158–163, 1997.

[135] Xinguo Ren, Patrick Rinke, Volker Blum, Jürgen Wieferink, Alexandre Tkatchenko, Andrea Sanfilippo, Karsten Reuter, and Matthias Scheffler. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New Journal of Physics*, 14(5):053020, 2012.

[136] Claudia Filippi and Cyrus J Umrigar. Multiconfiguration wave functions for quantum monte carlo calculations of first-row diatomic molecules. *The Journal of Chemical Physics*, 105(1):213–226, 1996.

[137] Michel Caffarel and Roland Assaraf. A pedagogical introduction to quantum monte-carlo. In *Mathematical models and methods for ab initio quantum chemistry*, pages 45–73. Springer, 2000.

[138] Xinguo Ren, Patrick Rinke, Christian Joas, and Matthias Scheffler. Random-phase approximation and its applications in computational chemistry and materials science. *Journal of Materials Science*, 47(21):7447–7471, 2012.

[139] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.

[140] Robert O Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of modern physics*, 87(3):897, 2015.

[141] Roberto Car. Introduction to density-functional theory and ab-initio molecular dynamics. *Quantitative Structure-Activity Relationships*, 21(2):97–104, 2002.

[142] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[143] LJ Sham and Walter Kohn. One-particle properties of an inhomogeneous interacting electron gas. *Physical Review*, 145(2):561, 1966.

[144] LE Elsholz. Differential equations and variational calculus, 1969.

[145] Mel Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proceedings of the National Academy of Sciences*, 76(12):6062–6065, 1979.

[146] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008.

[147] Erich Runge and Eberhard KU Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52(12):997, 1984.

[148] J Schirmer and A Dreuw. Critique of the foundations of time-dependent density-functional theory. *Physical Review A*, 75(2):022513, 2007.

[149] Thomas A Niehaus and Norman H March. Brief review related to the foundations of time-dependent density functional theory. *Theoretical Chemistry Accounts*, 125 (3-6):427–432, 2010.

[150] J Schirmer. Reexamination of the runge-gross action-integral functional. *Physical Review A*, 86(1):012514, 2012.

[151] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: basic theory and advanced methods.* Cambridge University Press, 2009.

[152] Fabien Gatti. *Molecular quantum dynamics: from theory to applications.* Springer, 2014.

[153] Michael P Allen et al. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23(1):1–28, 2004.

[154] Herbert Goldstein, Charles Poole, and John Safko. Classical mechanics, 2002.

[155] James M Haile. *Molecular dynamics simulation: elementary methods.* John Wiley & Sons, Inc., 1992.

[156] Eric Paquet and Herna L Viktor. Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *BioMed research international*, 2015, 2015.

[157] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.

[158] Naomichi Hatano and Masuo Suzuki. *Finding Exponential Product Formulas of Higher Orders*, pages 37–68. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-31515-5. doi: 10.1007/11526216_2. URL https://doi.org/10.1007/11526216_2.

[159] Martin Karplus. Development of multiscale models for complex chemical systems: from H + H_2 to biomolecules (Nobel lecture). *Angewandte Chemie International Edition*, 53(38):9992–10005, 2014.

[160] J. Frenkel. *Wave Mechanics*. Clarendon Press, Oxford, 1934.

[161] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51:455–500, 2009.

[162] J. C. Light, I. P. Hamilton, and J. V. Lill. Generalized discrete variable approximation in quantum mechanics. 82:1400, 1985.

[163] M. D. Feit, J. A. Fleck Jr., and A. Steiger. Solution of the Schrödinger equation by a spectral method. 47:412–433, 1982.

[164] T. J. Park and J. C. Light. Unitary quantum time evolution by iterative Lanczos reduction. 85:5870, 1986.

[165] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Q. Appl. Math.*, 9:17, 1951.

[166] R. H. Bisseling, R. Kosloff, R. B. Gerber, M. A. Ratner, L. Gibson, and C. Cerjan. Exact time-dependent quantum mechanical dissociation dynamics of I_2He: Comparison of exact time-dependent quantum calculation with the quantum time-dependent self-consistent field (TDSCF) approximation. 87:2760, 1987.

[167] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer. The multi-configuration time-dependent Hartree (MCTDH) method: A highly efficient algorithm for propagating wave packets. *Phys. Rep*, 324:1–105, 2000.

[168] Z. Kotler, A. Nitzan, and R. Kosloff. Multiconfiguration time-dependent self-consistent field approximation for curve crossing in presence of a bath. A Fast Fourier Transform study. 153:483, 1988.

[169] J. Zanghellini, M. Kitzler, C. Fabian, T. Brabec, and A. Scrinzi. An MCTDHF approach to multi–electron dynamics in laser fields. 13:1064, 2003.

[170] O. E. Alon, A. I. Streltsov, and L. S. Cederbaum. Fragmentation of Bose-Einstein condensates in multi-well three-dimensional traps. 347:88, 2005.

[171] A. Raab, G. Worth, H.-D. Meyer, and L. S. Cederbaum. Molecular dynamics of pyrazine after excitation to the S_2 electronic state using a realistic 24-mode model Hamiltonian. 110:936–946, 1999.

[172] F. Gatti and C. Iung. Exact and constrained kinetic energy operators for polyatomic molecules: The polyspherical approach. 484:1–69, 2009.

[173] Haobin Wang and Michael Thoss. Numerically exact quantum dynamics for indistinguishable particles: The multilayer multiconfiguration time-dependent Hartree theory in second quantization representation. 131(2):024114, 2009.

[174] O. Vendrell and H.-D. Meyer. Multilayer multiconfiguration time-dependent Hartree method: Implementation and applications to a Henon-Heiles Hamiltonian and to pyrazine. 134:044135, 2011.

[175] Uwe Manthe and Thomas Weike. On the multi-layer multi-configurational time-dependent hartree approach for bosons and fermions. *The Journal of Chemical Physics*, 146(6):064117, 2017.

[176] H. Wang and M. Thoss. Multilayer formulation of the multiconfiguration time-dependent Hartree theory. 119:1289–1299, 2003.

[177] E. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. 17:87, 1975.

[178] J. M. Bowman, S. Carter, and X. Huang. MULTIMODE: a code to calculate rovibrational energies of polyatomic molecules. *Int. Rev. Phys. Chem.*, 22:533–549, 2003.

[179] H. Rabitz and O. F. Alis. General foundations of high-dimensional model representations. 25:197, 1999.

[180] M. Schröder, F. Gatti, and H.-D. Meyer. Theoretical studies of the tunneling splitting of malonaldehyde using the multiconfiguration time-dependent Hartree approach. 134:234307, 2011.

[181] G.-J. Kroes and H.-D. Meyer. Using n-mode potentials for reactive scattering: Application to 6D H_2 + Pt(111). 440:334–340, 2007.

[182] A. Jäckle and H.-D. Meyer. Product representation of potential energy surfaces. 104:7974, 1996.

[183] A. Jäckle and H.-D. Meyer. Product representation of potential energy surfaces II. 109:3772, 1998.

[184] I. Burghardt, K. Giri, and G. A. Worth. Multi-mode quantum dynamics: The G-MCTDH method applied to the absorption spectrum of pyrazine. 129:174104, 2008.

[185] G. W. Richings, I. Polyak, K. E. Spinlove, G. A. Worth, I. Burghardt, and B. Lasorne. Quantum dynamics simulations using gaussian wavepackets: the vMCG method. 34:269, 2015.

[186] Iakov Polyak, Charlotte SM Allan, and Graham A Worth. A complete description of tunnelling using direct quantum dynamics simulation: Salicylaldimine proton transfer. *The Journal of chemical physics*, 143(8):084121, 2015.

[187] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4): 1253–1278, 2000.

[188] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-(R_1,R_2,. . .,R_N) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000.

[189] Venera Khoromskaia and Boris N Khoromskij. *Tensor numerical methods in quantum chemistry.* Walter de Gruyter GmbH & Co KG, 2018.

[190] Salman Ahmadi-Asl, Andrzej Cichocki, Anh Huy Phan, Ivan Oseledets, Stanislav Abukhovich, and Toshihisa Tanaka. Randomized algorithms for computation of Tucker decomposition and higher order SVD (HOSVD). *arXiv preprint arXiv:2001.07124*, 2020.

[191] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[192] D Brandoni and Valeria Simoncini. Tensor-train decomposition for image recognition. *Calcolo*, 57(1):9, 2020.

[193] H.-D. Meyer, F. Gatti, and G. A. Worth, editors. *Multidimensional Quantum Dynamics: MCTDH Theory and Applications.* Wiley-VCH, Weinheim, 2009.

[194] Markus Schröder and Hans-Dieter Meyer. Transforming high-dimensional potential energy surfaces into sum-of-products form using monte carlo methods. *The Journal of Chemical Physics*, 147(6):064105, 2017.

[195] Markus Schröder. Transforming high-dimensional potential energy surfaces into a canonical polyadic decomposition using monte carlo methods. *The Journal of Chemical Physics*, 152(2):024108, 2020.

[196] Frank Otto. Multi-layer potfit: an accurate potential representation for efficient high-dimensional quantum dynamics. *The Journal of Chemical Physics*, 140(1): 014106, 2014.

[197] Frank Otto, Ying-Chih Chiang, and Daniel Peláez. Accuracy of potfit-based potential representations and its impact on the performance of (ML-)MCTDH. *Chemical Physics*, 509:116–130, 2018.

[198] D. Peláez, K. Sadri, and H.-D. Meyer. Full-dimensional MCTDH/MGPF study of the ground and lowest lying vibrational states of the bihydroxide $H_3O_2^-$ complex. 119:42–51, 2014. doi: 10.1016/j.saa.2013.05.008.

[199] Satoshi Maeda and Koichi Ohno. A new approach for finding a transition state connecting a reactant and a product without initial guess: applications of the scaled hypersphere search method to isomerization reactions of HCN,$(H_2O)_2$, and alanine dipeptide. *Chemical physics letters*, 404(1-3):95–99, 2005.

[200] Silvia Bonfanti and Walter Kob. Methods to locate saddle points in complex landscapes. *The Journal of Chemical Physics*, 147(20):204104, 2017.

[201] Andreas Heyden, Alexis T Bell, and Frerich J Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *The Journal of chemical physics*, 123(22):224101, 2005.

[202] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics*, 113(22):9901–9904, 2000.

[203] Elena F Koslover and David J Wales. Comparison of double-ended transition state search methods. *The Journal of chemical physics*, 127(13):134102, 2007.

[204] Jorge M del Campo and Andreas M Köster. A hierarchical transition state search algorithm. *The Journal of chemical physics*, 129(2):024107, 2008.

[205] Aurelio Rodríguez, Roberto Rodríguez-Fernández, Saulo A. Vázquez, George L. Barnes, James JP Stewart, and Emilio Martínez-Núñez. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *Journal of computational chemistry*, 39(23):1922–1930, 2018.

[206] Manyi Yang, Jingxiang Zou, Guoqiang Wang, and Shuhua Li. Automatic reaction pathway search via combined molecular dynamics and coordinate driving method. *The Journal of Physical Chemistry A*, 121(6):1351–1361, 2017.

[207] Leif D Jacobson, Art D Bochevarov, Mark A Watson, Thomas F Hughes, David Rinaldo, Stephan Ehrlich, Thomas B Steinbrecher, S Vaitheeswaran, Dean M Philipp, Mathew D Halls, et al. Automated transition state search and its application to diverse types of organic reactions. *Journal of chemical theory and computation*, 13 (11):5780–5797, 2017.

[208] Robert A DiStasio Jr, Vivekanand V Gobre, and Alexandre Tkatchenko. Many-body van der waals interactions in molecules and condensed matter. *Journal of Physics: Condensed Matter*, 26(21):213202, 2014.

[209] E Herbst, Q Chang, and HM Cuppen. Chemistry on interstellar grains. In *Journal of Physics Conference Series*, volume 6, pages 18–35, 2005.

[210] C Garcia-Fernandez, B Radola, L Martin-Gondre, S Picaud, MT Rayez, JC Rayez, FX Ouf, and J Rubayo-Soneira. A DFT study of the interaction between large PAHs and atomic chlorine or hydrogen chloride molecule: Toward a modelling of the influence of chlorinated species on the trapping of water by soot. *Chemical Physics*, 483:46–55, 2017.

[211] Dirk Porezag, Th Frauenheim, Th Köhler, Gotthard Seifert, and R Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Physical Review B*, 51(19):12947, 1995.

[212] Béla Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[213] Fabio Pietrucci and Wanda Andreoni. Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. *Physical review letters*, 107(8):085504, 2011.

[214] Kenichi Fukui. The path of chemical reactions-the IRC approach. *Accounts of chemical research*, 14(12):363–368, 1981.

[215] Michael P Metz, Konrad Piszczatowski, and Krzysztof Szalewicz. Automatic generation of intermolecular potential energy surfaces. *Journal of Chemical Theory and Computation*, 12(12):5895–5919, 2016.

[216] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[217] Andreas Antoniou and Wu-Sheng Lu. *Practical optimization: algorithms and engineering applications.* Springer Science & Business Media, 2007.

[218] Sergiy Butenko and Panos M Pardalos. *Numerical methods and optimization: An introduction.* CRC Press, 2014.

[219] Luis Miguel Rios and Nikolaos V Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.

[220] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[221] David J Wales and Jonathan PK Doye. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[222] Ana M Ferreiro-Ferreiro, Jose A Garcia-Rodriguez, Luis Souto, and Carlos Vázquez. Basin hopping with synched multi L-BFGS local searches. parallel implementation in multi-CPU and GPUs. *Applied Mathematics and Computation*, 356:282–298, 2019.

[223] AH Rinnooy Kan and GT Timmer. Stochastic global optimization methods part II: multi level methods. *Mathematical Programming: Series A and B*, 39(1):57–78, 1987.

[224] László Pál. Benchmarking a hybrid multi-level single-linkage algorithm on the bbob noiseless testbed. In *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pages 1145–1152, 2013.

[225] Hee Sun Hong and Fred J Hickernell. Algorithm 823: Implementing scrambled digital sequences. *ACM Transactions on Mathematical Software (TOMS)*, 29(2): 95–109, 2003.

[226] Lawrence Davis. Handbook of genetic algorithms. 1991.

[227] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[228] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.

[229] Constantino Tsallis and Daniel A Stariolo. Generalized simulated annealing. *Physica A: Statistical Mechanics and its Applications*, 233(1-2):395–406, 1996.

[230] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[231] Henry Leung and Simon Haykin. The complex backpropagation algorithm. *IEEE Transactions on signal processing*, 39(9):2101–2104, 1991.

[232] Michael JD Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2): 155–162, 1964.

[233] Martin Buhmann, Roger Fletcher, Arieh Iserles, and Philippe Toint. Michael James David Powell: 29 july 1936-19 april 2015. *Biogr. Mems Fell R. Soc.(London)*, 64, 2018.

[234] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[235] William Kahan. Pracniques: further remarks on reducing truncation errors. *Communications of the ACM*, 8(1):40, 1965.

[236] George L Barnes, David R Glowacki, Sabine Kopec, Emilio Martinez-Nunez, Daniel Pelaez-Ruiz, Aurelio Rodriguez, Roberto Rodriguez-Fernandez, Robin J Shannon, James JP Stewart, Pablo G Tahoces, et al. Auto mekin. `https://rxnkin.usc.es/index.php/AutoMeKin`.

[237] Jesús A Varela, Saulo A Vázquez, and Emilio Martínez-Núñez. An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis. *Chemical science*, 8(5):3843–3851, 2017.

[238] Helena R Lourenço, Olivier C Martin, and Thomas Stützle. Iterated local search. In *Handbook of metaheuristics*, pages 320–353. Springer, 2003.

[239] David Ferro-Costas, Emilio Martínez-Núñez, Jesús Rodríguez-Otero, Enrique Cabaleiro-Lago, Carlos M Estévez, Berta Fernández, Antonio Fernández-Ramos, and Saulo A Vázquez. Influence of multiple conformations and paths on rate constants and product branching ratios. thermal decomposition of 1-propanol radicals. *The Journal of Physical Chemistry A*, 122(21):4790–4800, 2018.

[240] Yann Fenard, Adrià Gil, Guillaume Vanhove, Hans-Heinrich Carstensen, Kevin M Van Geem, Phillip R Westmoreland, Olivier Herbinet, and Frédérique Battin-Leclerc. A model of tetrahydrofuran low-temperature oxidation based on theoretically calculated rate constants. *Combustion and Flame*, 191:252–269, 2018.

[241] Raúl Pérez-Soto, Saulo A Vázquez, and Emilio Martínez-Núñez. Photodissociation of acryloyl chloride at 193 nm: Interpretation of the product energy distributions, and new elimination pathways. *Physical Chemistry Chemical Physics*, 18(6):5019–5026, 2016.

[242] Michael J Wilhelm, Emilio Martínez-Núñez, Jesús González-Vázquez, Saulo A Vázquez, Jonathan M Smith, and Hai-Lung Dai. Is photolytic production a viable source of HCN and HNC in astrophysical environments? a laboratory-based feasibility study of methyl cyanoformate. *The Astrophysical Journal*, 849(1):15, 2017.

[243] Saulo A Vázquez and Emilio Martínez-Núñez. HCN elimination from vinyl cyanide: product energy partitioning, the role of hydrogen–deuterium exchange reactions and a new pathway. *Physical Chemistry Chemical Physics*, 17(10):6948–6955, 2015.

[244] Estefanía Rossich Molina, Jean-Yves Salpin, Riccardo Spezia, and Emilio Martinez-Nunez. On the gas phase fragmentation of protonated uracil: a statistical perspective. *Physical Chemistry Chemical Physics*, 18(22):14980–14990, 2016.

[245] H Bernhard Schlegel. Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):790–809, 2011.

[246] Satoshi Maeda, Yu Harabuchi, Yuriko Ono, Tetsuya Taketsugu, and Keiji Morokuma. Intrinsic reaction coordinate: Calculation, bifurcation, and automated search. *International Journal of Quantum Chemistry*, 115(5):258–269, 2015.

[247] James Stewart. *MOPAC2016*. URL `http://openmopac.net/`.

[248] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[249] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44):6615–6620, 2008.

[250] Andrea N. Bootsma and Steven Wheeler. Popular integration grids can result in large errors in dft-computed free energies, 2019. URL `https://chemrxiv.org/articles/Popular_Integration_Grids_Can_Result_in_Large_Errors_in_DFT-Computed_Free_Energies/8864204/5`.

[251] C Garcia-Fernandez, Sylvain Picaud, Marie-Thérèse Rayez, Jean-Claude Rayez, and J Rubayo-Soneira. First-principles study of the interaction between NO and large carbonaceous clusters modeling the soot surface. *The Journal of Physical Chemistry A*, 118(8):1443–1450, 2014.

[252] Hrant P Hratchian and H Bernhard Schlegel. Accurate reaction paths using a hessian based predictor–corrector integrator. *The Journal of chemical physics*, 120 (21):9918–9924, 2004.

[253] Michael Page and James W McIver Jr. On evaluating the reaction path hamiltonian. *The Journal of chemical physics*, 88(2):922–935, 1988.

[254] Shiro Koseki and Mark S Gordon. Intrinsic reaction coordinate calculations for very flat potential energy surfaces: application to singlet disilenylidene isomerization. *The Journal of Physical Chemistry*, 93(1):118–125, 1989.

[255] Kathryn Zimmermann, Narumol Jariyasopit, Staci L Massey Simonich, Shu Tao, Roger Atkinson, and Janet Arey. Formation of nitro-PAHs from the heterogeneous reaction of ambient particle-bound pahs with $N_2O_5/NO_3/NO_2$. *Environmental science & technology*, 47(15):8434–8442, 2013.

[256] Xiao Wei Sheng, Lukasz Mentel, Oleg V Gritsenko, and Evert Jan Baerends. Counterpoise correction is not useful for short and van der waals distances but may be useful at long range. *Journal of computational chemistry*, 32(13):2896–2901, 2011.

[257] Chunyang Peng and H Bernhard Schlegel. Combining synchronous transit and quasi-newton methods to find transition states. *Israel Journal of Chemistry*, 33(4): 449–454, 1993.

[258] J Baker and PMW Gill. An algorithm for the location of branching points on reaction paths. *Journal of computational chemistry*, 9(5):465–475, 1988.

[259] Ramón Lorenzo Panadés-Barrueta, Emilio Martínez-Núñez, and Daniel Peláez. Improving the efficiency of vdw-tsscds through the use of srp. *In preparation*, .

[260] Fabien Gatti, Christophe Iung, and Xavier Chapuisat. Vector parametrization of the three-atom problem in quantum mechanics II. valence vectors. *Journal of Molecular Structure: THEOCHEM*, 430:201–208, 1998.

[261] G. A. Worth, M. H. Beck, A. Jäckle, and H.-D. Meyer. The MCTDH Package, H.-D. Meyer, Version 8.4.12. See http://mctdh.uni-hd.de/, 2016.

[262] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.

[263] F. Gatti, C. Iung, M. Menou, Y. Justum, A. Nauts, and X. Chapuisat. Vector parametrization of the n-atom problem in quantum mechanics. I. Jacobi vectors. 108:8804, 1998.

[264] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and A. Wolf. Molpro, version 2010.1, a package of ab initio programs, 2010. see http://www.molpro.net.

[265] M. J. D. Powell. *Cambridge NA Report NA2009/06, University of Cambridge*, pages 26–46, 2009.

[266] S. G. Johnson. The NLopt nonlinear-optimization package. *http://ab-initio.mit.edu/nlopt*, 2011.

[267] Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. Tensorly: Tensor learning in Python. *CoRR*.