# Machine Learning Coupled Potential Energy Surfaces

Tzu Yu Wang

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Science, Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Michael S. Schuurman

# Abstract

Recent advancements in machine learning (ML) for potential energy surfaces (PESs) have yielded promising results, with much of the focus on isolated ground-state surfaces. Extending machine learning to coupled excited state surfaces introduces significant challenges, however, particularly in regions of conical intersections. At these points, the adiabatic potentials are non-differentiable, complicating the application of standard ML techniques. In this work, we build on a previously proposed approach that overcomes this issue by learning the nuclear coordinate dependent characteristic polynomial coefficients of the potential matrix instead, which enables the construction of smooth, accurate machine learning models even at seams of conical intersections.

The proposed approach is first validated at the *ab initio* multi-reference configuration interaction (MRCI) level of theory for various molecules. We examine the ability of the proposed model to accurately reproduce energies near a minimum energy conical intersection (MECI) and analyze its performance in capturing seams of conical intersection against analytical MRCI solutions. The results demonstrate that, through this approach, quantitatively accurate machine learning models of seams of conical intersection may be constructed.

We further demonstrate the application of this framework to the combined density functional theory and multi-reference configuration interaction (DFT/MRCI) method. The selected configuration interaction nature of the method enables the calculation of accurate excitation energies at a low computational cost. However, the potential energy surfaces produced resemble smooth underlying surfaces contaminated with noise, rendering them locally non smooth. To mitigate this, we treat the local discontinuities in the potential surfaces as noise by explicitly optimizing a whitenoise kernel within a Gaussian process regression framework. We apply this method to optimize DFT/MRCI minimum energy conical intersection geometries and compare the results to *ab initio* MRCI solutions. While treating the locally discontinuous surface as noise limits the ability to achieve arbitrarily small energy gap at nominal intersections, the structures and branching spaces obtained are in strong agreement with ab initio data. This approach thus proves to be a viable method for generating smooth, accurate representations of DFT/MRCI(2) PESs.

# Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to Prof. Michael Schuurman for his unwavering patience and guidance over the past few years. I have learned so much under his mentorship. To Dr. Simon Neville, whose support has helped me grow in countless ways and who has become an irreplaceable mentor, I offer my deepest thanks. I am also grateful to all my lab members for their patience with my many questions and for helping me become a better scientist. Finally, I could not have embarked on this journey without the love and support of my family and my partner, who mean everything to me.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

**BS** branching space 15, 17, 19, 21, 22, 68, 70, 71, 81, 82

**CI** configuration interaction 25–28, 30

**ConIns** conical intersections 2, 3, 10, 11, 13–22, 49, 50, 64, 66–68, 70, 75, 77, 79–82, 86, 89, 94, 107, 108

**CP** characteristic polynomial 51–53, 58–61, 63, 75, 107

**CSF** configuration state function 26, 29–31

**DFT** density function theory 27, 77, 83

**DFT/MRCI** the combined density functional theory and multireference configuration interaction 3, 28–30, 77, 83, 108

**DFT/MRCI(2)** the combined density functional theory and multireference configuration interaction(2) 31, 39, 77–79, 81, 83, 84, 86, 87, 90, 91, 95–97, 99, 101, 102, 104, 105, 107, 108

**GPR** Gaussian process regression 39–41, 44, 46, 54, 56, 68, 74, 75, 77, 78, 84, 86, 98–100, 102–105, 107

**HF** Hartree-Fock 24, 25, 27–30

**KS-DFT** Kohn-Sham density function theory 27–30

**LHS** Latin hypercube sampling 54, 64, 84, 98, 99

**MECI** minimum energy conical intersection 15, 16, 39, 50, 54, 55, 64, 68, 75, 77–79, 81–87, 89, 91, 92, 95, 97, 98, 101, 102, 104, 105, 107

**ML** machine learning 2, 34, 46, 106

**MOs** molecular orbitals 23, 24, 27, 28

**MRCI** multireference configuration interaction 3, 26, 54, 70, 81, 83, 84, 86, 87, 89–91, 94, 95, 104, 105, 107

**PES** potential energy surface 2, 3, 7, 10, 79

**PESs** potential energy surfaces 1–3, 7, 10, 11, 33–35, 39, 49, 50, 58, 64, 68, 71, 74, 75, 84, 86, 101, 102, 107, 108

**SD** Slater determinant 23–27

**SOAP** smooth overlap of atomic positions 34–38, 41, 55, 85, 107

**TISE** time-independent Schrödinger equation 5, 6, 8, 10, 23

# 1

# Motivation

Understanding chemical processes at the molecular level relies heavily on the concept of potential energy surfaces (PESs), which describes the energy landscape of a molecular system as a function of atomic positions. These surfaces provide critical insight into reaction dynamics, transition states, and molecular interactions, making them central to fields such as reaction mechanism prediction, material design, and photochemistry. The generation of accurate PESs is therefore a crucial step in computational chemistry. However, producing

these surfaces requires high-level electronic structure methods, which can be computationally expensive and formally complex, especially for large systems or when multiple excited states are involved.

Recent advances in machine learning (ML) offer promising solutions to this computational bottleneck by providing efficient methods to approximate PESs, with the potential to revolutionize the way these surfaces are represented. However, when it comes to excited states and the complex phenomena associated with them, such as conical intersections (ConIns), standard ML techniques face significant challenges. ConIns are points in the PESs where electronic states become degenerate, leading to cusps in the potential surface. This non-differentiability at the intersection complicates the application of traditional ML methods, which typically rely on smooth and differentiable functions to train models.

This thesis aims to address this challenge by leveraging a framework that has been previously proposed to overcome the difficulties posed at ConIns. Rather than directly learning the adiabatic PESs, our approach involves learning the nuclear coordinate dependent characteristic polynomial coefficients of the potential matrix, which are smooth functions of nuclear coordinates and enables smooth interpolation of the potential energy surface (PES) even at ConIns. This strategy provides a computationally efficient way to construct coupled PESs without sacrificing accuracy or computational feasibility.

This thesis is structured into five main chapters: Chapter 2 and 3 introduces the background necessary to keep all the analysis done in this thesis self contained, including the theory of coupled potential energy surfaces, electronic structure theory and machine learn-

ing methods. Chapter 4 of this work applies the proposed framework to study ConIns in multireference configuration interaction (MRCI) calculations for various molecules. By benchmarking this approach against analytical MRCI solutions, we demonstrate the accuracy and robustness of the proposed method. Chapter 5 of this thesis extends the framework to the combined density functional theory and multireference configuration interaction (DFT/MRCI) method, where no analytical solutions exist. This application further highlights the versatility and potential of the framework for handling more complex, non-analytical systems, as well as gain insight into the PESs generated at the DFT/MRCI level of theory.

In summary, this thesis presents progresses in the intersection of machine learning and computational chemistry, offering a novel and efficient way to model excited-state PES, particularly in the presence of conical intersections. Through careful validation and application to a range of systems, we demonstrate the potential of this approach to transform the way PES are generated, opening new avenues for the study of machine learning chemical processes in the excited-state regime.

# 2

# Excited State Potential Energy Surfaces

## 2.1 Born-Oppenheimer Approximation

The quantum mechanical description of atoms and molecules involved in any chemical process can be extracted by solving the time-independent Schrödinger equation (TISE)

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\Psi$ is the full molecular wavefunction and $\hat{H}$ here is the field-free, non-relativistic molecular Hamiltonian in the position basis given by:

$$\begin{aligned}\hat{H} &= \hat{T}_N + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN} \\ &= -\sum_i^N \frac{1}{2M_i}\nabla^2_{\boldsymbol{R}_i} - \sum_i^n \frac{1}{2}\nabla^2_{\boldsymbol{r}_i} + \sum_i^n \sum_{j>i}^n \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \\ &\quad - \sum_i^n \sum_{j>i}^N \frac{Z_j}{|\boldsymbol{r}_i - \boldsymbol{R}_j|} + \sum_i^N \sum_{j>i}^N \frac{Z_i Z_j}{|\boldsymbol{R}_i - \boldsymbol{R}_j|}\end{aligned} \tag{2.2}$$

where the Hamiltonian is shown in terms of atomic units. Herein, all the operators and wavefunctions are expressed in the position basis.

In the first line of Eq.2.2, $\hat{T}_N$ and $\hat{T}_e$ are the nuclear and electronic kinetic energy operator, respectively, and $\hat{V}_{ee}$, $\hat{V}_{eN}$ and $\hat{V}_{NN}$ are the pair-wise Coulomb potential operators between the electronic (with subscript $e$) and the nuclei (with subscript $N$) degrees of freedom. In the second line, the sum for the nuclear components runs over the $N$ number of nuclei atoms, and similarly the electronic component runs over the $n$ number of electrons.

The gradients $\nabla_{\boldsymbol{r}_i}$ and $\nabla_{\boldsymbol{R}_i}$ are taken with respective to the $i^{th}$ electronic and nuclear position coordinate, $\boldsymbol{r}_i$ and $\boldsymbol{R}_i$, respectively, and the same notation is used throughout this thesis. Solving the TISE with the full molecular Hamiltonian (Eq. 2.2) for real systems is an extremely challenging task, and approximations are needed to effectively solve the systems beyond the Hyrdogen Atom [1].

The central approximation that underlie most of modern quantum chemistry is the Born-Oppenheimer approximation [1]. The first step in the approximation is the clamped-nuclei approximation, which, through a mass analysis of the nucleus and electrons, shows that the nuclei and electron motions evolve on drastically different timescales. As a result, the electrons are thought to adjust "instantaneously" to the nuclei positions. In terms of the molecular Hamiltonian, this means that the nuclear kinetic energy operator $\hat{T}_N$ is removed from the full molecular Hamiltonian, and the nuclear-nuclear repulsion term $V_{NN}$ becomes a constant and thus can be removed without changing the equations of motion. The resulting Hamiltonian is often referred to as the electronic Hamiltonian, which is given by:

$$\begin{aligned} H_e &= \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} \\ &= -\sum_i^n \frac{1}{2}\nabla_i^2 + \sum_i^n \sum_{j>i}^n \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_i^n \sum_{j>i}^N \frac{Z_j}{|\boldsymbol{r}_i - \boldsymbol{R}_j|} \end{aligned} \tag{2.3}$$

The clamped nuclei approximation enables the separation of the full Hamiltonian into an electronic Hamiltonian [1], defined in Eq.2.3, which can now be solved independently of the nuclear degrees of freedom. Even with this simplification, solving the TISE with

the electronic Hamiltonian is a non-trivial task, and is the subject of electronic structure theory, which will be discussed in Section 2.3. Once the electronic Hamiltonian is solved, we get the set of electronic states and electronic energies

$$H_e\psi_m(\boldsymbol{r};\boldsymbol{R}) = E_m^e(R)\psi_m(\boldsymbol{r};\boldsymbol{R}), \tag{2.4}$$

where as a consequence of the clamped-nuclei approximation, the electronic wavefunctions and energies are only parameterized by the nuclear degrees of freedom. The eigenvalue to the above equation, $E_m^e(\boldsymbol{R})$, is the adiabatic potential energy surface (PES) with $m = 0$ corresponding to the electronic ground state energy and $m = 1$ the first excited state energy etc.

PESs are the cornerstone of modern chemistry and have been used to rationalize a plethora of chemical processes, including the dynamical and spectroscopic properties of molecules. The choice of electronic structure method is therefore crucial in any chemical studies, as the generated PESs, and subsequently the theoretical understanding, can differ drastically at different levels of theory.

The wavefunction that solves the full molecular Hamiltonian (Eq.2.2) can now be expanded in a complete basis of the adiabatic states $\{\psi_m\}$:

$$
\begin{aligned}
|\Psi\rangle &= \sum_i |\psi_i\rangle \langle\psi_i|\Psi\rangle \\
&= \sum_i |\psi_i\rangle |\chi_i\rangle \\
\Psi(\boldsymbol{r},\boldsymbol{R}) &= \langle \boldsymbol{r},\boldsymbol{R}|\Psi\rangle = \sum_i \psi_i(\boldsymbol{r};\boldsymbol{R})\chi_i(\boldsymbol{R})
\end{aligned}
\tag{2.5}
$$

The expansion in Eq.2.5 is known as the Born-Huang expansion [2], and is an exact solution given the electronic states form a complete basis. Substituting the Born-Huang expansion into the full molecular TISE (Eq.2.1) and taking the expectation value with respect to $\psi_j$ (left multiply by $\psi_j^\dagger$ and integrating out the electronic degrees of freedom), we arrive at:

$$
\left[ T_N + E_j^e(\boldsymbol{R}) - \sum_i \Lambda_{ji} \right] \chi_i(\boldsymbol{R}) = E\chi_i(\boldsymbol{R}) \tag{2.6}
$$

The $\Lambda$ term is the non-adiabatic coupling [2], given by $\Lambda_{ji} = \delta_{ji}T_N - \langle\psi_j|T_N|\psi_i\rangle$. It couples the electronic and nuclear motions through the nuclear kinetic energy operator acting on the nuclear coordinate parameterized electronic adiabatic states. Intuitively, if the electronic states do not change with respect to $\boldsymbol{R}$, then the nonadiabatic coupling is small or becomes zero in the limit that the electronic states are independent of $\boldsymbol{R}$. Setting $T_N = \frac{-1}{2\mu}\nabla_R^2$, where $\mu$ is the reduced nuclear mass of the system, we can rewrite the non-adiabatic coupling $\Lambda$ in Eq.2.6 after some algebra as [2]:

$$\begin{aligned}\Lambda_{ji} &= \frac{1}{2\mu}\left[2\boldsymbol{F}_{ji}\cdot\nabla + G_{ji}\right] \\ \boldsymbol{F}_{ji}(\boldsymbol{R}) &= \langle\psi_j|\nabla_R\psi_i\rangle = \frac{\langle\psi_j|\nabla H_e|\psi_i\rangle}{E_i(\boldsymbol{R}) - E_j(\boldsymbol{R})} \\ G_{ji} &= \langle\psi_j|\nabla_R^2\psi_i\rangle\end{aligned} \tag{2.7}$$

The second part of the Born-Oppenheimer approximation then assumes that the full molecular wavefunction can be described solely as a single product of the nuclear and electronic wavefunction:

$$\Psi = \psi_m(\boldsymbol{r};\boldsymbol{R})\chi_m(\boldsymbol{R}), \tag{2.8}$$

and after going through the same algebra as the Born-Huang expansion leads to a similar equation

$$\left[T_N - \frac{G}{2M} + E_e(\boldsymbol{R})\right]\chi_m(\boldsymbol{R}) = E\chi_m \tag{2.9}$$

Unlike the full non-adiabatic coupling term in Eq.2.6, however, the $\frac{G}{2M}$ above is just a scalar that modifies the energy, as shown in Eq 2.7, and is often called the diagonal Born-Oppenheimer correction. The equation can be further approximated by neglecting the $\frac{G}{2M}$ term, resulting in the final Born-Oppenheimer adiabatic approximation:

$$\left[T_N + E_e(\boldsymbol{R})\right]\chi_m = E\chi_m \tag{2.10}$$

It is evident from Eq 2.10 that the nuclear coordinate-dependent electronic energies are referred to as potential energy surfaces because they directly serve as the potential in the nuclear TISE. This picture is central to understanding many thermochemical processes, where the nuclear configuration evolves along a single PES, typically the ground state. However, this single product approximation breaks down in excited state photochemistry, where multiple electronic states can couple due to significant non-adiabatic effects, as shown in Eq 2.6. In such cases, it becomes necessary to consider multiple PESs simultaneously. This can lead to regions of degeneracy between PESs, known as conical intersections (ConIns), which play a critical role in photochemical processes. Conical intersections will be discussed in more detail in Section 2.2.

### 2.1.1 Diabatic States

It is clear from examining Eq.2.7 that in regions where the adjacent adiabatic energies are close-lying or become degenerate, such as at a conical intersection, the coupling term $\boldsymbol{F}_{ji}$ diverges due to the inversely proportional energy gap, which is a major problem of studying coupled PESs in the adiabatic representation. To overcome this problem, a different electronic representation is introduced by a change of basis $\boldsymbol{U}(\boldsymbol{R})$:

$$\boldsymbol{\psi^d}(\boldsymbol{r};\boldsymbol{R}) = \boldsymbol{U}(\boldsymbol{R})\boldsymbol{\psi}(\boldsymbol{r};\boldsymbol{R}) \tag{2.11}$$

The states $\boldsymbol{\psi^d}$ above are called the diabatic states, and are defined such that the derivative coupling is removed. That is, $\boldsymbol{F}^d_{ji} = 0$, where the superscript $\boldsymbol{d}$ denotes the

coupling in a diabatic basis. When the coupling term can be rigorously set to zero, the states are referred to as strict diabatic states.

A solution to $\boldsymbol{F}^d_{ji} = 0$, however, do not in general exist for polyatomic molecules when the electronic states are truncated [3]. In practice, a transformation that minimizes $\boldsymbol{F}^d_{ji}$ is used, resulting in quasi-diabatic states [2]. Since strict diabatic states do not in general exist, diabatization approaches seek to remove or minimize the singular parts of the $\boldsymbol{F}^d_{ji}$, which can be accomplished in many different ways. The construction of quasi-diabatic states is thus not uniquely defined, and multiple approaches of obtaining them exist [2].

The diagonal potential matrix originally in the adiabatic state basis, now also transforms into a diabatic potential matrix which contains non-zero off-diagonal terms.

$$\boldsymbol{V^d} = \boldsymbol{U^{-1}VU} \tag{2.12}$$

An important property of the diabatic representation is, in addition to removing the singularity of $\boldsymbol{F}_{ji}$ at ConIns, the diabatic surfaces (elements of diabatic potential matrix) do not exhibit cusps at conical intersections that is otherwise present in the adiabatic basis, which will be detailed in Section 2.2, and are thus the preferred basis when fitting or learning coupled PESs. This is an important property that is used in the next Chapter.

While diabatic states are the preferred basis for quantum dynamics, the global rotation of the surfaces is in general not feasible for on the fly dynamics. This can be overcome if the global surfaces are known a priori, but for high dimensional systems, they are generally not. In many instances, adiabatic surfaces are adequate and is a more intuitive basis to

work in despite their issues, which may include propagating a trajectory or optimizing a structural minimum.

## 2.2 Conical Intersections

Conical intersections (ConIns) are defined as regions in the nuclear coordinate space where the adiabatic PESs intersect and the degeneracy is lifted to first order [2]. They are ubiquitous in excited state manifolds and are crucial to understanding many excited state photo-dynamical phenomena [2, 4]. ConIns also have important theoretical consequence as it give rise to geometric phases in the electronic wavefunction, and can affect the quantum dynamics. There can in principle be an arbitrary number $n$ of intersecting states forming an $n$-state ConIns, but the most commonly encountered and studied are two-state and three-state ConIns, which are also the models studied in this thesis and outlined below.

### 2.2.1 Two-State Conical Intersections

A two-state ConIns has two degenerate (intersecting) adiabatic energy surfaces, and is the most studied model due to its simple analytical solution but significance to the modeling of many phenomena. To start, we take a general $2 \times 2$ Hamiltonian matrix:

$$\boldsymbol{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix} \tag{2.13}$$

where $H_{ij} = \langle \phi_i | H | \phi_j \rangle$ and $\phi$ is the arbitrary basis used to expand the electronic states. The two adiabatic energies, $E_1(\boldsymbol{R})$ and $E_2(\boldsymbol{R})$ ($E_1 \leq E_2$), then take on a simple form:

$$E_{1,2}(\boldsymbol{R}) = \omega(\boldsymbol{R}) \pm \sqrt{\Delta H(\boldsymbol{R}) + H_{12}^2(\boldsymbol{R})} \tag{2.14}$$

$$\begin{aligned} \omega(\boldsymbol{R}) &\equiv \frac{H_{11}(\boldsymbol{R}) + H_{22}(\boldsymbol{R})}{2} \\ \Delta H(\boldsymbol{R}) &\equiv \frac{H_{11}(\boldsymbol{R}) - H_{22}(\boldsymbol{R})}{2} \end{aligned} \tag{2.15}$$

The two energy surfaces are degenerate ($E_1 = E_2$) at a ConIns, $\boldsymbol{R}_{CI}$, which can only be achieved if the matrix elements $\Delta H$ and $H_{12}$ are both zero. The conditions needed for $\Delta H = H_{12} = 0$ and insight into the surfaces in the vicinity of the ConIns is obtained by Taylor expanding the matrix elements about a ConIns to first order [5]:

$$\begin{aligned} \Delta H(\boldsymbol{R}) &\approx \Delta H(\boldsymbol{R}_{CI}) + \nabla_{\boldsymbol{R}}(\Delta H)|_{\boldsymbol{R}_{CI}} \cdot \boldsymbol{R} \\ &= \nabla_{\boldsymbol{R}}(\Delta H) \cdot \boldsymbol{R} \\ H_{12}(\boldsymbol{R}) &\approx H_{12}(\boldsymbol{R}_{CI}) + \nabla_{\boldsymbol{R}}(H_{12})|_{\boldsymbol{R}_{CI}} \cdot \boldsymbol{R} \\ &= \nabla_{\boldsymbol{R}}(H_{12}) \cdot \boldsymbol{R} \end{aligned} \tag{2.16}$$

For the degeneracy condition to hold, the second line for both quantities in Eq.2.16 need to be zero. That is,

$$\begin{aligned} \nabla_{\boldsymbol{R}}(\Delta H) \cdot \boldsymbol{R} &= 0 \\ \nabla_{\boldsymbol{R}}(H_{12}) \cdot \boldsymbol{R} &= 0 \end{aligned} \tag{2.17}$$

which implies that for the degeneracy to be preserved by a (linear) displacement, it needs

to be orthogonal to the two vectors $\nabla(\Delta H)$ and $\nabla(H_{12})$.

The two vectors, $\nabla(\Delta H)$ and $\nabla(H_{12})$, are commonly referred to as the energy difference gradient $\boldsymbol{g}$ and derivative coupling vector $\boldsymbol{h}$, respectively [2, 5, 6]:

$$\begin{aligned} \boldsymbol{g} &\equiv \nabla_{\boldsymbol{R}}(\Delta H) \\ \boldsymbol{h} &\equiv \nabla_{\boldsymbol{R}}(H_{12}) \end{aligned} \tag{2.18}$$

The $\boldsymbol{g}$ and $\boldsymbol{h}$ vector form the two-dimensional subspace that lifts the degeneracy to first order, and is termed the branching space (BS), while the $N-2$ orthogonal complement which preserves the degeneracy is termed the seam space. ConIns are therefore not isolated points on the energy surfaces, but rather form a seam in the $N-2$ subspace. The nuclear configuration on the seam with the lowest energy is called the minimum energy conical intersection (MECI) structure, and is an important stationary point used to compare the ConIns of different methods and molecules.

The nascent BS vectors in Eq.2.18 are in general not orthogonal to each other, but can be orthogonalized via a rotation:

$$\begin{aligned} \tilde{\boldsymbol{g}} &= -\cos(\beta)\boldsymbol{g} + \sin(\beta)\boldsymbol{h} \\ \tilde{\boldsymbol{h}} &= \sin(\beta)\boldsymbol{g} + \cos(\beta)\boldsymbol{h} \\ \tan(2\beta) &= \frac{2\langle \boldsymbol{g}, \boldsymbol{h} \rangle}{\langle \boldsymbol{h}, \boldsymbol{h} \rangle - \langle \boldsymbol{g}, \boldsymbol{g} \rangle} \end{aligned} \tag{2.19}$$

where $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{h}}$ are the orthgonalized BS vectors, and $<,>$ refers to the inner (dot) product.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the unit branching vectors,

$$\begin{aligned} \boldsymbol{x} &= \tilde{\boldsymbol{g}}/\tilde{g} \\ \boldsymbol{y} &= \tilde{\boldsymbol{h}}/\tilde{h} \end{aligned} \tag{2.20}$$

where $\tilde{g} \equiv |\tilde{g}|$ and $\tilde{h} \equiv |\tilde{h}|$ are the $l^2$-norms, then the tilt parameters $s_x$ and $s_y$ are defined as the projection of the average energy gradient onto the unit branching vectors:

$$\begin{aligned} s_x &= \nabla\omega(\boldsymbol{R}) \cdot \boldsymbol{x} \\ s_y &= \nabla\omega(\boldsymbol{R}) \cdot \boldsymbol{y} \end{aligned} \tag{2.21}$$

Using the parameters $s_x$, $s_y$ and the branching vectors, the $2\times2$ Hamiltonian of Eq.2.13 can be written in the intersection adapted coordinate $\boldsymbol{x}$ and $\boldsymbol{y}$ to first order as [2, 5]:

$$\boldsymbol{W}^{(1)}(x,y) = (s_x x + s_y y)\,\mathbf{1}_2 + \begin{bmatrix} \tilde{g}x & \tilde{h}y \\ \tilde{h}y & -\tilde{g}x \end{bmatrix} \tag{2.22}$$

Eq.2.22 is the main working equation for a 2-state ConIns used to characterize the intersection topography, such as at a MECI. The set of parameters $\{s_x, s_y, \tilde{g}, \tilde{h}\}$ can fully characterize the first-order intersection topography of a two-state ConIns. The energy expressed in the intersection adapted coordinates then read as:

$$E_{1,2}(x,y) = s_x x + s_y y \pm \sqrt{(\tilde{g}x)^2 + (\tilde{h}y)^2}, \tag{2.23}$$

which shows the cusp at a ConIns in the BS (x-y plane) that renders it non-smooth, as shown in Fig.2.1

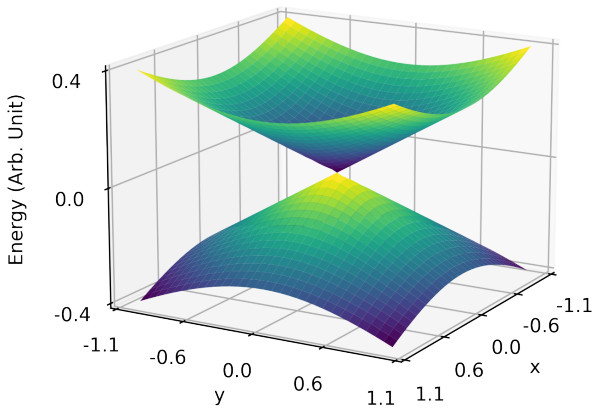


Figure 2.1: Example of a typical ConIns double cone in the $x-y$ branching plane plotted using energies given by Eq.2.23, with $s_x = s_y = 0$ and $g = h = 0.3$. The energy is given in arbitrary units, and the origin exhibits a cusp that is characteristic of ConIns

A two-state ConIns can be fully characterized in a first-order model by the set of parameters $\{s_x, s_y, \tilde{g}, \tilde{h}\}$. Specifically, the norms of the BS vectors, $\tilde{g}$ and $\tilde{h}$, characterizes the asymmetry of the cone. When $\tilde{g} = \tilde{h}$, the typical double cone structure, as shown in Fig.2.1, projected onto the $x-y$ plane will be circular (symmetrical), whereas if $\tilde{g}$ and $\tilde{h}$ are different, the resulting $x-y$ plane projection will be elliptical. The $s_x$ and $s_y$ parameters describe the tilt of the double cone relative to the principle axis, defined as the normal vector of the $x-y$ plane. When both $s_x$ and $s_y$ are zero or close to zero, the ConIns exhibits a vertical double cone. If one or both of the tilt parameter is non-zero, the intersection will

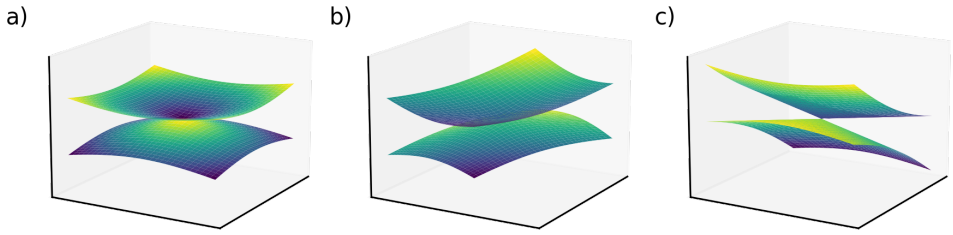


Figure 2.2: Comparison of different ConIns intersection topography given by Eq 2.23 for: a) same parameters as Fig 2.1, namely $s_x = s_y = 0$ and $g = h = 0.3$, b) $s_x = s_y = 0.5$ and $g = h = 0.3$, c)$s_x = 0.5$, $s_y = 0$ and $g = h = 0.3$.

be sloped. The intersection topographies are important for a simple characterization of the dynamic as it passes through the ConIns. For example, it can be seen from using simple trajectory arguments that vertical cones facilitate surface transitions more efficiently than a tilted cone. [5] To illustrate some examples of different intersection topography, Fig.2.2 shows a first-order ConIns with varying asymmetry and tilt parameters.

### 2.2.2 Three-State Conical Intersections

A brief introduction of a three-state ConIns is given, which follows the same analysis as the two-state ConIns. The $3 \times 3$ Hamiltonian is given as:

$$\boldsymbol{H} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{12} & H_{22} & H_{23} \\ H_{13} & H_{23} & H_{33} \end{bmatrix} = H_{22}\mathbf{1}_3 + \begin{bmatrix} \Delta H_{13} & H_{12} & H_{13} \\ H_{12} & \Delta H_{23} & H_{23} \\ H_{13} & H_{23} & 0 \end{bmatrix}, \tag{2.24}$$

where the last diagonal term is subtracted out, and $\Delta H_{ij} \equiv H_{ii} - H_{jj}$.

Using the same arguments as the two-state case, at a ConIns geometry $\boldsymbol{R}_{CI}$, the conditions $H_{12} = H_{13} = H_{23} = 0$ and $H_{11} = H_{22} = H_{33}$ are required for the eigenvalues to be degenerate. Taylor expanding the matrix elements to first order, we get a set of condition at a ConIns [7]:

$$\begin{aligned} \Delta H_{ij}(\boldsymbol{R}) &\approx \Delta H_{ij}(\boldsymbol{R}_{CI}) + \nabla_{\boldsymbol{R}}(\Delta H_{ij})|_{\boldsymbol{R}_{CI}} \cdot \boldsymbol{R} \\ &= \nabla_{\boldsymbol{R}}(\Delta H_{ij}) \cdot \boldsymbol{R} \\ H_{ij}(\boldsymbol{R}) &\approx H_{ij}(\boldsymbol{R}_{CI}) + \nabla_{\boldsymbol{R}}(H_{ij})|_{\boldsymbol{R}_{CI}} \cdot \boldsymbol{R} \\ &= \nabla_{\boldsymbol{R}}(H_{ij}) \cdot \boldsymbol{R} \end{aligned} \tag{2.25}$$

for which the necessary conditions for the degeneracy results in a set of five equations

$$\begin{aligned} \nabla_{\boldsymbol{R}}(\Delta H_{ij}) \cdot \boldsymbol{R} = 0 \\ \nabla_{\boldsymbol{R}}(H_{kl}) \cdot \boldsymbol{R} = 0, \end{aligned} \tag{2.26}$$

with $\{ij\} = \{13, 23\}$ and $\{kl\} = \{12, 13, 23\}$.

The BS for a three-state ConIns is therefore a 5 dimensional subspace, with a $N - 5$ dimensional seam space.

### 2.2.3 Types of Conical Intersections

There are in general three types of ConIns: the (1) symmetry allowed, (2) symmetry required and (3) accidental ConIns. In this thesis, we focus and examine the symmetry required (2) and accidental ConIns (3).

The symmetry allowed (1) ConIns refer to crossings when the two states have different symmetries. Since the Hamiltonian is totally symmetric and the two states have different symmetry, the condition $H_{12} = 0$ is automatically satisfied because only totally symmetric integrals are non-zero.

The symmetry required (2) ConIns, also commonly known as the Jahn-Teller effect, are the earliest studied type of ConIns. For molecules exhibiting degenerate irreducible representations (irreps), which are defined as those having dimensions greater than 1, the electronic states that transform as the degenerate irrep will have degenerate energies. The Jahn-Teller theorem tells us that at these symmetry-induced degeneracy, the BS vectors [1] (vibrational modes) are those whose irreps are contained in the symmetric product of the electronic state irreps. That is, $\Gamma^{vib} \subset [\Gamma^e \times \Gamma^e]_s$, where $\Gamma^{vib}$ and $\Gamma^e$ are the irreps of the vibrational mode and electronic state, respectively, and $[\Gamma^{elec} \times \Gamma^{elec}]_s$ the symmetrized product. An implication of the symmetry required ConIns is the double cone will be symmetrical ($\tilde{g} = \tilde{h}$).

Finally, the accidental (3) ConIns refers to degeneracies that do not arise due to any

[1] In the context of Jahn-Teller systems, the degeneracy lifting BS vectors are often referred to as the Jahn-Teller active modes

symmetry argument, and thus cannot be located *a priori* by symmetry. For many years, accidental ConIns were considered rare, as it was believed that the degeneracy conditions (Eq 2.17) would generally be difficult to satisfy without symmetry. However, advances in computations and simulations over the past two decades have revealed that accidental ConIns are ubiquitous in practical molecular systems, which in general lack high symmetry.

### 2.2.4 Methods of Branching Space Determination

While the BS vectors have simple analytical forms given in Eq.2.18, they may not be readily available from general quantum chemistry programs, and their implementations are often non-trivial. In lieu of analytical solutions, the BS vectors can also be obtained by considering the energy difference squared surface $\Delta E^2(\boldsymbol{R})$ at a ConIns [8, 9]. Since the BS and seam space lifts degeneracy at first and second order, respectively, $\Delta E^2(\boldsymbol{R}_{CI})$ will be lifted to second-order in the BS, and to fourth-order in the seam space.

The $\Delta E^2(\boldsymbol{R})$ surface at a ConIns will therefore only have non-zero curvatures in the BS, while the seam space will have zero curvature. The Hessian of the energy difference squared surface at a ConIns

$$\mathcal{H}_{ab} = \left.\frac{\partial^2 \Delta E^2}{\partial R_a \partial R_b}\right|_{\boldsymbol{R}_{CI}}, \tag{2.27}$$

thus has a column space corresponding to the BS and a null space corresponding to the seam space.

In addition, the intersection topography $g$ and $h$ can also be obtained from the non-zero eigenvalues of the $\Delta E^2(\boldsymbol{R})$ Hessian. Note that from the first-order intersection adapted energies (Eq.2.23), the energy difference squared surface is given as $\Delta E^2(\boldsymbol{R}) = 4((\tilde{g}x)^2 + (\tilde{h}y)^2)$. Taking the Hessian and solving for the eigenvalues yield $\lambda_1 = 8\tilde{g}^2$ and $\lambda_2 = 8\tilde{h}^2$, where $\lambda$ are the eigenvalues of the $\Delta E^2$ Hessian [9]. The norms of the orthogonalized BS vectors are therefore

$$\begin{aligned} \tilde{g} &= \sqrt{\lambda_1/8} \\ \tilde{h} &= \sqrt{\lambda_2/8} \end{aligned} \tag{2.28}$$

For a general $n$-state ConIns with multiple intersecting surfaces, each pairs of surfaces needs to be considered as degeneracy can be lifted within any pairs of surface. The equivalent quantity would therefore generalize to:

$$\begin{aligned} \mathcal{H}_{ab} &= \left.\frac{\partial^2 F}{\partial R_a \partial R_b}\right|_{\boldsymbol{R}_{CI}} \\ F &= \sum_{i<j}^{n} \Delta E^2 \end{aligned} \tag{2.29}$$

The orthogonalized branching (seam) space vectors may be computed as the eigenvectors of the $\Delta E^2(\boldsymbol{R}_{CI})$ Hessian with nonzero (zero) eigenvalues.

## 2.3 Electronic Structure Theory

Electronic structure theory refers to the study of electrons in atoms and molecules, which amounts to solving the electronic TISE (Eq.2.4) encountered earlier. Even without the nuclear degrees of freedom, it is an extremely difficult problem to solve and approximations are needed. We give here only briefly some of the underlying theory.

### 2.3.1 Hartree-Fock Theory

To solve the electronic eigenvalue problem, the wavefunction is expanded in a single particle bases, commonly known as molecular orbitals (MOs), which are themselves expanded in a basis set. Taking the anti-symmetrized product of $n$ MOs so that the electronic wavefunction satisfies the correct fermionic (electrons) symmetry, we arrive at the Slater determinant (SD) [1]:

$$\psi^{SD}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\boldsymbol{x_1}) & \ldots & \phi_n(\boldsymbol{x_1}) \\ \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{x_n}) & \ldots & \phi_n(\boldsymbol{x_n}) \end{vmatrix}, \tag{2.30}$$

where $\phi_i$ are the single particle MOs, and $\boldsymbol{x_i} = \{\boldsymbol{r_i}, \omega\}$ is a composite coordinate that describes an electron's Cartesian $\boldsymbol{r_i} \in \mathbb{R}^3$ and spin $\omega \in \{\uparrow, \downarrow\}$ coordinates.

Given this trial wavefunction, the energy can be obtained by invoking the variational principle, which provides a lower bound on the approximated ground state energy. Solving

the stationary point for the energy of the trial wavefunction with respect to linear orbital variations subsequently returns a variationally optimized wavefunction. The assumption that the ground state wavefunction can be expressed by a single SD is the core of Hartree-Fock (HF) theory, and is the foundation for more complex wavefunction type methods.

The variational optimization of the orbitals in the SD results in a pseudo-eigenvalue problem that needs to be solved self-consistently:

$$f(\boldsymbol{x_1})\phi_i(\boldsymbol{x_1}) = \epsilon_i\phi_i(\boldsymbol{x_1}) \tag{2.31}$$

The $f(\boldsymbol{x_1})$ in Eq.2.31 is the Fock operator, which is a mean-field approximation of the electronic Hamiltonian that arises as a consequence of the SD ansatz:

$$f(\boldsymbol{x_1}) = h(\boldsymbol{x_1}) + \sum_{j\neq i} J_j(\boldsymbol{x_1}) - K_j(\boldsymbol{x_1}) \tag{2.32}$$

$$\begin{aligned} J_j(\boldsymbol{x_1}) &= \int d\boldsymbol{x_2}\phi_j^*(\boldsymbol{x_2})\frac{1}{r_{12}}\phi_j(\boldsymbol{x_2}) \\ K_j(\boldsymbol{x_1})\phi_i(\boldsymbol{x_1}) &= \left[\int d\boldsymbol{x_2}\phi_j^*(\boldsymbol{x_2})\frac{1}{r_{12}}\phi_i(\boldsymbol{x_2})\right]\phi_j(\boldsymbol{x_1}) \end{aligned} \tag{2.33}$$

The HF ground state wavefunction then corresponds to a SD formed by the set of converged and variationally optimized MOs $\{\phi\}$. The mean-field approximation results in a solvable problem, but necessarily introduces errors. The energy that is missing or otherwise not captured by the HF mean-field approximation is known as the correlation energy, defined by

$$E_{corr} \equiv E_{exact} - E_{HF}, \tag{2.34}$$

where $E_{corr}$ is the correlation energy, $E_{exact}$ is the unknown exact electronic energy, and $E_{HF}$ corresponds to the mean-field, HF energy.

SDs are particularly important because they form a complete basis for the anti-symmetric subspace of the N-particle Hilbert space. That is, any anti-symmetric N-particle state can be expressed as a linear combination of N-particle SDs [1]. This is an important result that is the foundation for more advanced, post-HF methods, which attempts to retrieve the correlation energy that is unaccounted for in HF by expanding the wavefunction in a basis of SDs.

### 2.3.2 Multireference Configuration Interaction

The idea behind configuration interaction (CI) is to use SDs as the many-body Hilbert space bases to build more complicated wavefunctions that cannot be adequately described by a single SD alone. In particular, CI wavefunctions are built by taking (fermionic) excitations from occupied to virtual orbitals out of a single SD (usually the HF solution), given by the general form [1]:

$$|\Psi\rangle = c_0 |\Psi_{HF}\rangle + \sum_{ra} c_a^r |\Psi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs} |\Psi_{ab}^{rs}\rangle + \dots \tag{2.35}$$

where $|\Psi_a^r\rangle$ refers to a different SD formed by exciting an electron from the occupied

orbital $\phi_a$ to an unoccupied orbital $\phi_r$, and the linear expansion coefficients $\{c\}$ now being the variational parameters. As a technical side-note, a single SD is in general not an eigenfunction of the total spin operator, unless it has equal spin densities. In practice, spin-adapted configurations called configuration state function (CSF) is used, which is obtained by taking linear combinations of SDs, but this does not change the arguments made here.

When the zeroth order wavefunction in a CI expansion is a single SD, as implied in Eq. 2.35, it is referred to as a single-reference CI, or simply CI. In contrast, when the zeroth order wavefunction is a linear combination of SDs, it is referred to as a multireference CI (MRCI). The idea behind MRCI is intuitive: The zeroth order description of some electronic states are not well described by a single SD, especially with electronic states that involve near-degenerate configurations [10]. The correlation arising from such near-degenerate electronic configurations is referred to as static correlation [1, 10], and is ubiquitous when describing non-adiabatically coupled excited states. The remaining electronic correlation, aptly named dynamical correlation, refers to the instantaneous electron-electron repulsion that is absent in a mean-field approximation, and is typically recovered by perturbation theory or a CI expansion [10, 11]. Finally, if the expansion in Eq.2.35 is truncated to include only single excitations, it is referred to as (MR-)CI singles (MR-CIS). Similarly, if both single and double excitations are included, it is referred to as (MR-)CI singles doubles (MR-CISD). A typical method to generate the multiple reference CSFs needed in MRCI is to use the restricted active space- (RASSCF) and complete active space- (CASSCF) self consistent method. The main idea behind both RASSCF and CASSCF is

to do a full configuration interaction (considering all possible excitations) within a small chosen subset of orbitals [12], where both the CI and MO expansion coefficients are simultaneously optimized. RASSCF is a more flexible method than CASSCF, where aside from a full CI within an orbital subspace, an additional subspace is defined with truncated excitations carried out within it [12].

### 2.3.3 Density Functional Theory

A very brief summary of density function theory (DFT) is presented in order to keep the introduction of the method in Section 2.3.4 self contained.

The idea behind DFT is to avoid the direct construction of high dimensional electronic wavefunctions by describing the system solely with the real space electron density, which has three degrees of freedom. The Hohenberg-Kohn theorems proved that the ground state properties of a many-particle system can be uniquely described by an electron density, and that the total energy is a functional of the electron density [13]. Kohn and Sham further introduced a non-interacting kinetic energy functional and constrained the density to be formed from a set of single particle MOs [13], i.e. SD for electrons.

While Kohn-Sham density function theory (KS-DFT) has a different derivation and origin, the resulting working equations are equivalent to that of HF. That is, a pseudo-eigenvalue problem (Eq.2.31) that needs to solved self-consistently. Similarly, it results in a set of optimized MOs that form the density, where in place of the Fock operator (Eq.2.32), it is replaced with a KS effective potential $v_{ks}$ [13]:

$$
\begin{aligned}
v_{ks}(\boldsymbol{r}) &= v(\boldsymbol{r}) + \int d\boldsymbol{r'} \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r'}|} + \frac{\delta E_{xc}[\rho]}{\delta \rho}, \\
\rho(\boldsymbol{r}) &= \sum_{i}^{n} |\phi_i(\boldsymbol{r})|^2,
\end{aligned}
\tag{2.36}
$$

where $\rho(\boldsymbol{r})$ is the non-interacting $\nu$-representable electron density and $E_{xc}[\rho]$ is the exchange-correlation functional [13], and is the main quantity of interest in KS-DFT. The main strength of KS-DFT lies in its ability to efficiently capture dynamic correlation compared to standard post-HF methods. The recovery of dynamic correlation in post-HF methods relies on a CI expansion, as shown in Eq.2.35, which converges slowly and grows factorially in dimension. In comparison, a well-chosen exchange-correlation functional in KS-DFT can capture a high degree of dynamic correlation more efficiently [14]. The accuracy of the method is thus dependent on the choice and design of the exchange-correlation function, and is an active area of research [14].

### 2.3.4 The combined density function theory and multireference configuration interaction

As the name of the method implies, the combined density function theory and multireference configuration interaction (DFT/MRCI) method seeks to combine the advantages of KS-DFT and a CI expansion. In essence, DFT/MRCI uses optimized KS-DFT MOs (contrary to HF MOs in a usual CI expansion) as the basis for a CI type expansion. With this ansatz, the dynamic correlation can be more efficiently captured by the KS-DFT MOs,

while the static correlations can be accounted for with a MR-CISD expansion [15, 16]. Before showing the DFT/MRCI Hamiltonian, we first note that the *ab initio* diagonal Hamiltonian elements can be written in second quantization as [16]

$$\begin{aligned}\langle \boldsymbol{w}\omega|\hat{H} - E_{SCF}|\boldsymbol{w}\omega\rangle =& \sum_i F_{ii}\Delta w_i + \frac{1}{2}\sum_{ij}\left(V_{iijj} - \frac{1}{2}V_{ijji}\right)\Delta w_i \Delta w_j \\ &+ \frac{1}{2}\sum_{i,j\in S_w} V_{ijji}\left[\langle \boldsymbol{w}\omega|\hat{E}_i^j\hat{E}_j^i|\boldsymbol{w}\omega\rangle\,(1-\delta_{ij}) - \frac{1}{2}\right]\end{aligned} \tag{2.37}$$

Here, each CSF $|\boldsymbol{w}\omega\rangle$ is represented as a spatial occupation $\boldsymbol{w}$ and spin coupling pattern $\omega$ in Fock space [15, 16]. $\Delta w_i \equiv w_i - \overline{w}_i$ is the occupation number difference of the $i^{th}$ spatial orbital to a reference configuration $\overline{\boldsymbol{w}}$, typically the HF occupation. $V_{ijkl}$ denotes the two electron integral in chemist notation, $F_{ii}$ denotes the Fock matrix elements, and $\hat{E}_k^l$ is a singlet excitation operator. The set $S_w$ summed over in the third term is the set of indices that has single orbital occupation in the configuration $\boldsymbol{w}$.

The DFT/MRCI ansatz using the KS-DFT orbitals then results in diagonal matrix elements

$$\begin{aligned}\langle \boldsymbol{w}\omega|\hat{H}^{DFT} - E_{DFT}|\boldsymbol{w}\omega\rangle =& (E_{SCF} - E_{DFT}) + \sum_i \epsilon_{ii}^{KS}\Delta w_i + \frac{1}{2}\sum_{ij}\left(\tilde{V}_{iijj} - \tilde{V}_{ijji}\right)\Delta w_i \Delta w_j \\ &+ \frac{1}{2}\sum_{i,j\in S_w} V_{ijji}\left[\langle \boldsymbol{w}\omega|\hat{E}_i^j\hat{E}_j^i|\boldsymbol{w}\omega\rangle\,(1-\delta_{ij}) - \frac{1}{2}\right]\end{aligned} \tag{2.38}$$

where $\tilde{V}_{iijj}$ and $\tilde{V}_{ijji}$ are the DFT/MRCI down-scaled Coulomb and exchange integrals given by

$$\begin{aligned} \tilde{V}_{iijj} &= \left(1 - \frac{\beta_c}{n_{exc}}\right) V_{iijj} \\ \tilde{V}_{ijji} &= \left(\frac{1}{2} - \frac{\beta_x(n_{open})}{n_{exc}}\right) V_{ijji} \end{aligned} \tag{2.39}$$

By using the KS-DFT instead of HF orbital energy differences, a larger portion of dynamic correlation can be incorporated but necessitates the down-scaling of the Coulomb and exchange terms. However, since static and dynamic correlation cannot be clearly partitioned, damping terms in the off-diagonal Hamiltonian matrix elements are used to mitigate the double counting of dynamic correlation that arise from a CI expansion [15, 16]. The damping term is at the heart of the method, as it allows for the decoupling of a large portion of the base and excited CSFs, thereby discarding a large portion of the excited CSFs and often leads to many order of magnitude reduction on the basis size. In practice, an *a priori* orbital energy criterion is used to decide whether a CSF is included. The off diagonal DFT/MRCI matrix elements are simply given by

$$\langle \boldsymbol{w}\omega|\hat{H}^{DFT} - E_{DFT}|\boldsymbol{w}'\omega'\rangle = \langle \boldsymbol{w}\omega|\hat{H} - E_{SCF}|\boldsymbol{w}'\omega'\rangle \, D(\Delta E_{\boldsymbol{w}\omega,\boldsymbol{w}'\omega'}) \tag{2.40}$$

where $D(\Delta E_{\boldsymbol{w}\omega,\boldsymbol{w}'\omega'})$ is the damping function typically chosen to decay rapidly with increasing $\Delta E$.

Additionally, a perturbative variant of the parent DFT/MRCI method, simply termed

the combined density functional theory and multireference configuration interaction(2) (DFT/MRCI(2)) has been developed in our group, which further reduces the computational cost without sacrificing the accuracy [16–18]. In essence, DFT/MRCI(2) partitions the Hamiltonian based on the reference CSFs and CSFs generated via single and double excitations out of the reference space, from which an effective Hamiltonian is constructed using quasi-degenerate perturbation theory [16]. The energy can then be obtained as the eigenvalues of the effective Hamiltonian, and the wavefunction with perturbative corrections. In practice, this reduces the diagonalization of the original $N_{CSF} \times N_{CSF}$ Hamiltonian to one that is $N_s \times N_s$, where $N_{CSF}$ is the number of CSFs generated and $N_s$ is the number of states requested, and typically $N_s << N_{CSF}$ [16].

# 3

# Machine Learning of Potential Energy Surfaces

## 3.1 Smooth Overlap of Atomic Positions

To learn or model any structure-property relationship of a molecule, there naturally needs to be a way to represent the nuclear structure. While there is a plethora of methods to represent a molecule, the most straightforward being the Cartesian coordinates of the $3N$ atoms, the prevalent way of representing molecules in the context of machine learning is with molecular descriptors.

There exists many molecular descriptor variants, with some descriptors tailored for certain systems. For the learning of interatomic potential and PESs, however, most descriptors share a common set of important qualities [19]:

1. Invariant to spatial translations
2. Invariant to rotation of the coordinate system
3. Invariant with permutations of equivalent atoms
4. Complete and unique: the structure and descriptor representation has a bijective mapping and each property corresponds to a unique descriptor
5. Continuous and differentiable: 'small' changes in the structure result in a 'small' change in the descriptor.
6. Non-redundant and computationally cheap to construct

The descriptor should be both translational and rotationally invariant, as the field-field electronic Hamiltonian (Eq. 2.3) and the resulting PESs are invariant to these transformations. Similarly, permutation of atomic indices do not affect the electronic Hamiltonian. The descriptor's completeness and uniqueness ensure that any molecular structure can be accurately represented, with distinct descriptors corresponding to different molecular structures up to a symmetry. Its continuous and differentiable nature guarantees that PESs can be differentiated to yield quantities like force. Moreover, the continuous nature ensures that small perturbations to atomic positions lead to similar descriptors without exhibiting rapid changes. This is particularly important for kernel methods such as Gaussian process regression, which will be discussed in the next section. These methods rely on kernels to assess the similarity between atomic positions, which are in this work represented by a descriptor. Finally, the descriptor's non-redundant and computationally efficient design ensures that it represents molecular structures without unnecessary components, and the construction of descriptors do not become a bottleneck in the ML procedure.

In this work, we employ the smooth overlap of atomic positions (SOAP) descriptor. The SOAP descriptor is a three dimensional descriptor that satisfies the general qualities of a descriptor listed above [19, 20]. Other commonly used three-dimensional descriptors include variants of the Coulomb matrix [21], many-body tensor representation [22], and atom-centered symmetry functions [23]. We have chosen to initially use the SOAP descriptor, as it has proven to be an efficient representation for predicting the energies of organic molecules with chemical accuracy using kernel methods [24–26]. Additionally, it has demonstrated consistently higher accuracy in predicting the formation energies of inor-

ganic crystals compared to the other descriptors mentioned above [25]. As will be shown in Chapters 4 and 5, we found that the SOAP descriptor can be used to adequately learn the quantities of interest in this thesis, and thus more complex descriptors were not explored. Additionally, as will be outlined below, the SOAP descriptor do not directly depend on the number of atoms, but rather on the number of atomic species. For the learning of PESs of organic molecules, the number of species is often small and constrained, thus making the scaling of SOAP descriptors favorable. In brief, the SOAP descriptor encodes local atomic environments through a sum of Gaussian-type densities placed at a number of local centers, and the densities are subsequently expanded in terms of Gaussian radial and spherical harmonic angular basis functions. These local environments are expressed as a power spectrum and can be placed arbitrarily in real space, but they are often, and in this work, placed on each atomic centre in the molecule. These power spectra satisfy the above requirements of a descriptor, the proof and derivations can be found in Reference [19]. A partial power spectrum placed on a single centre is given by [19, 24]

$$p_{nn'l}^{Z_1 Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m (c_{nlm}^{Z_1})^* c_{n'lm}^{Z_2}, \tag{3.1}$$

where $n$ is the radial basis indices, and $l$ and $m$ index the angular spherical harmonic basis functions. The coefficient $c_{nlm}^Z$ is defined through the following inner product:

$$c_{nlm}^Z = \iiint_V dV g_n(r) Y_{lm}(\theta, \phi) \rho^Z(r), \tag{3.2}$$

$$\rho^Z(r) = \sum_i^{|Z_i|} e^{-\frac{1}{2\sigma^2|r-R_i|^2}}, \tag{3.3}$$

$$g_{nl}(r) = \sum_n^{n_{max}} \beta_{nn'l} r^l e^{-\alpha_{nl} r^2}, \tag{3.4}$$

where $\rho^Z$ in Equation 3.3 is a species-dependent pseudo-atomic density built from a Gaussian controlled by the width parameter $\sigma$, and the $Y_{lm}$ are spherical harmonics. The decay parameter $\alpha_n$ in the Gaussian radial function $g_n$ is chosen such that each term in $g_n$ decays to a threshold value of $10^{-3}$ at a cutoff radius $r_{cut}$. The coefficient $c_{nlm}$ is invariant to translation as well as permutation of atomic indices, as permutation of indices would simply amount to changes in the summation order for Eq.3.3, which does not affect the results. Regarding rotational invariance, it can be shown that an arbitrary rotation operator acting on the coefficient $c_{nlm}$ (specifically the spherical harmonics $Y_{lm}$, as the radial component is not affected by rotation) transforms it via the Wigner matrix, which form the irrep of the three dimensional rotation group SO(3). The Wigner matrix is unitary, and therefore the power spectrum of Eq. 3.1 is invariant under rotation. More detail on the derivation for this result can be found in Ref. [19]. Since the coefficients $c_{nlm}$ are continuous and differentiable with respect to the nuclear coordinates, the SOAP descriptor will be smooth. It has been determined that the SOAP power spectrum is overcomplete, i.e. redundant components in the power spectrum [19, 27], thus reducing the efficiency of the method. However, this does not significantly impact our studies, and therefore, we do not consider it a concern.

A full power spectrum that represents a local environment is taken as the concatenation of the partial power spectrum $p_{nn'l}^{Z_1 Z_2}$ for all unique pairs of atomic species. A global SOAP descriptor $\boldsymbol{P}$ with elements

$$P_I = P_{nn'l}^{Z_1 Z_2} \sim \sum_m \left( \frac{1}{n} \sum_i c_{nlm}^{i,Z_1} \right) \left( \frac{1}{n} \sum_i c_{nlm}^{i,Z_2} \right) \tag{3.5}$$

is then built from a sum over the local power spectra centered at each site. The size of the SOAP descriptor $\boldsymbol{P}$ therefore does not depend directly on the size of the molecule, but rather the number of angular and radial basis $\{Y_{lm}, g_{nl}\}$, and the number of atomic species included in Equation 3.1. The length of the resulting power spectrum can be calculated as [28]

$$L = \frac{1}{2} n_{max} S_n (n_{max} S_n + 1)(l_{max} + 1), \tag{3.6}$$

where $S_n$ is the number of atomic species included. For a fixed number of radial and spherical expansion terms, $n_{max}$ and $l_{max}$, the length of the SOAP descriptor only scales with the number of species. For studying organic molecules, the species are restricted to the first two rows of the periodic table, with the molecules studied in this thesis only containing C and H.

The SOAP descriptor is specified by four parameters: $r_{cut}$, $n_{max}$, $l_{max}$, $\sigma$. The parameters $n_{max}$ and $l_{max}$ specifies the angular expansion of the spherical harmonic in Equation 3.2, $\sigma$ controls the width of the Gaussian-type pseudo-density in Equation 3.3, and $r_{cut}$ implic-

itly changes the decay parameter in Equation 3.4, which controls the maximum reach of the radial basis. Computations of the SOAP descriptor were carried out using the DScribe Python package [25, 29].

## 3.2 Gaussian Process Regression

### 3.2.1 Introduction

In this thesis, we focus on regression problems, specifically the construction of surrogate models for mapping nuclear coordinates to the PESs. Various machine learning models, particularly neural networks in their many forms, are commonly employed to predict excited-state adiabatic energies[30]. The ultimate objective of constructing these surrogate PESs is to enable simulations of molecular dynamics, including both *ab initio* and quantum dynamics. In this context, we identify three key reasons for choosing Gaussian process regression (GPR) for this task.

First, GPR performs well with small amounts of training data and is less data intensive compared to neural network architectures [31], thus making it suitable for on-the-fly generation of PESs. That is, relevant local surfaces will be built and updated throughout the dynamic as opposed to building a global surface. Second, the differentiability class of the predictor can be guaranteed by the choice of kernel, and different types of noise can be accounted for surprisingly well with a stochastic noise model in a straightforward manner. Finally, uncertainty quantification is easily achieved via the predicted variance, which also aids in the construction of on-the-fly surfaces, as it can be used to determine areas where more training data is needed. The uncertainty quantification via predicted variance also enable us to quantify the nominal DFT/MRCI(2) MECIs, which will be discussed in detail in Chapter 5.

### 3.2.2 Derivation of Working Equations

GPR is a classical machine learning technique that falls under a broader category called kernel methods, which includes support vector machines and kernel ridge regression. While the derivation and interpretation of each kernel method may differ, they all employ the kernel trick, which may loosely be thought of as casting a non-linear regression problem into a linear one by mapping the original inputs into some higher (possibly infinite) dimensional feature space [32]. GPR in particular is different from the others in the kernel method category due to its Bayesian interpretation and derivation of the working equations.

Following Rasumussen and William, there are two equivalent but alternative derivations to arrive at the working equations of GPR, namely the weight- or function- space view. A brief summary of the function-space view is given here, and interested readers can find details in Ref. [32]. The following derivation closely follows Rasumussen and William [32].

A Gaussian process is a collection of (possibly infinite) random variables, any finite number of which have a joint Gaussian distribution. The notation for a Gaussian Process is written as:

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x'})) \tag{3.7}$$

$$\begin{aligned} m(\boldsymbol{x}) &= \mathbb{E}\left[f(\boldsymbol{x})\right] \\ k(\boldsymbol{x}, \boldsymbol{x'}) &= \mathbb{E}\left[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x'}) - m(\boldsymbol{x'}))\right] \end{aligned} \tag{3.8}$$

where the function, $f(\boldsymbol{x})$, is a Gaussian process that is fully specified by a mean function $m(\boldsymbol{x})$ and the covariance function $k(\boldsymbol{x}, \boldsymbol{x'})$. Gaussian processes could also be thought of as a distribution over functions, and Gaussian process regression (GPR) starts by assuming the function $f(\boldsymbol{x})$ that is being learned is modeled by a Gaussian process. In this work, the $f(\boldsymbol{x})$ are either the adiabatic energies or the $\omega$-CP surfaces, which will be introduced in the next chapter, and $\boldsymbol{x}$ is the nuclear coordinate represented via the SOAP descriptor. In Cartesian representation, $\boldsymbol{x} \in \mathbb{R}^{3N}$, where $N$ is the number of atoms in the molecule. In SOAP representation, $\boldsymbol{x} \in \mathbb{R}^{L}$, where $L$ is the length of the SOAP power spectrum given by Eq. 3.6.

The covariance function, also commonly referred to as the kernel, specifies the covariance between pairs of random variables, in this case the image (outputs) of the function $f(\boldsymbol{x})$ being learned [1]. The covariance between the output is written as a function of the inputs:

$$cov(f(x), f(x')) = k(x, x') \tag{3.9}$$

By the consistency (also known as the marginalization) property of Gaussian processes, the function evaluated at a set of finite input points, $\{\boldsymbol{X}_i, i = 1, \ldots, N\}$, is simply a multivariate Gaussian distribution specified by the mean and covariance function. i.e. some finite dimensional vector

[1] We consider only functions that map to a single output. That is, $f : X \to Y, Y \in \mathbb{R}$. This can be generalized for a vector-valued output by considering multi-output GP, which is beyond the scope of this thesis

$$\boldsymbol{f} \equiv \begin{bmatrix} f(\boldsymbol{x_1}) \\ f(\boldsymbol{x_2}) \\ \vdots \\ f(\boldsymbol{x_N}) \end{bmatrix} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}) \tag{3.10}$$

where for simplicity, the mean is taken to be a zero vector, which can always be achieved by subtracting the mean from the data set, and $\boldsymbol{K}$ corresponds to an $N \times N$ covariance matrix with elements $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. This property allows us to consider only the subset of points of interest, as supposed to requiring the consideration of the full infinite dimensional distribution.

Sampling from the unconditioned prior distribution (Eq.3.10) of $f$ amounts to drawing random functions. To turn it into a useful regression method requires conditioning of the prior using Bayes inference. Let $\mathcal{D} = \{(\boldsymbol{x}_i, f_i), i = 1, \ldots, N\}$ be the training set and $f* \equiv f(x_*)$ be the function evaluated at a test point $x_*$. The joint (Gaussian) distribution of the training and test outputs is given as [32]

$$\begin{bmatrix} f \\ f* \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{K}_* \\ \boldsymbol{K}_* & \boldsymbol{K}_{**} \end{bmatrix}\right) \tag{3.11}$$

where for notational convenience, $\boldsymbol{K}$ is the $N \times N$ covariance matrix for the training data, $\boldsymbol{K}_*$ is a vector with elements $k_i(\boldsymbol{x}_i, \boldsymbol{x}_*)$ for $\boldsymbol{x}_i \in \mathcal{D}$, and $\boldsymbol{K}_{**} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$.

The Gaussian property ensures that the posterior will also be Gaussian, and more

importantly, the inference can be done analytically, which after some algebra shows that

$$\begin{aligned} f_*|\boldsymbol{x}_*, \mathcal{D} &\sim \mathcal{N}(\mu_{post}, \Sigma_{post}) \\ \mu_{post} &= \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{f} \\ \boldsymbol{\Sigma}_{post} &= \boldsymbol{K}_{**} - \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{K}_* \end{aligned} \tag{3.12}$$

where the predicted value is obtained by conditioning the distribution of $f_*$ given the training data and the prediction point $x_*$, and $\boldsymbol{f}$ refers to a vector of the training values.

The above corresponds to the learning of a noiseless model. For more general cases, the function is assumed to take on the form $y = f(x) + \epsilon$, where $\epsilon$ is assumed to be an additive independent identically distributed (i.i.d.) Gaussian noise [2]. That is, $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$, and the covariance function is given as

$$cov(y(\boldsymbol{x}_i), y(\boldsymbol{x}_j)) = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sigma_n^2 \delta_{ij} \tag{3.13}$$

The joint distribution is now in terms of the 'noisy' data $\boldsymbol{y}$, but as everything is Gaussian, the posterior is still simply a Gaussian with modified posterior mean and variance:

$$\begin{aligned} f_*|\boldsymbol{x}_*, \mathcal{D} &\sim \mathcal{N}(\mu_{post}, \Sigma_{post}) \\ \mu_{post} &= \boldsymbol{K}_* (\boldsymbol{K} + \sigma_n^2 \mathbf{1})^{-1} \boldsymbol{y} \\ \boldsymbol{\Sigma}_{post} &= \boldsymbol{K}_{**} - \boldsymbol{K}_* (\boldsymbol{K} + \sigma_n^2 \mathbf{1})^{-1} \boldsymbol{K}_* \end{aligned} \tag{3.14}$$

[2] In practice, even when learning a noiseless function, a small additive noise is included for numerical stability

It can be seen from the posterior mean that GPR is determined solely by the choice of kernel. An advantage of GPR that makes it stand out from other kernel methods is the access to the variance at a predicted point. Note that the variances corresponds to the diagonal of the covariance matrix. This information allows for the uncertainty quantification of the model at a prediction point, which is a quality that will be utilized in the analysis in Section 5.4.

### 3.2.3 Choice of Kernel

The posterior mean in Eq.3.12 and Eq.3.14 is the main working equations of GPR, which can be rewritten in a more compact form that clearly shows the role of the kernel:

$$\tilde{f}(\boldsymbol{x}_*) = \sum_i^N \alpha_i k(\boldsymbol{x}_i, \boldsymbol{x}_*) \tag{3.15}$$

where $\boldsymbol{\alpha} = (K + \sigma_n^2 \mathbf{1})^{-1} \boldsymbol{y}$ is an $N$-dimensional vector with elements $\alpha_i$.

Specifically, it shows that the non-linearity in GPR arises from the linear combination of kernels evaluated between the training and prediction points. An important property of the kernel is it determines the differentiability class of the model, which can be clearly seen by taking the derivative of Eq.3.15, and noting that $\alpha_i$ is a constant with respect to the inputs $\boldsymbol{x}_*$. The kernels should thus be chosen such that it has the same differentiability class as the function being learned, if that information is *a priori* known.

A kernel takes two inputs as arguments and assigns a "similarity" between them. It

is specified by a set of hyperparameters that can be optimized to yield a better fit, which will be discussed in the following subsection.

Below we show a list of the kernels that was used in this thesis,

- **Squared Exponential Kernel:**

  The squared exponential (SE) (also commonly known as radial basis function) kernel is given by

$$k_{SE}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2 e^{-\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^2}{2l^2}}, \tag{3.16}$$

  where the hyperparameters are the variance ($\sigma^2$) and length-scale($l$), which can be loosely thought of as the amplitude and frequency of the fitted function. A SE kernel is infinitely differentiable, thus functions fitted using this kernel are very smooth: they belong to the $C^\infty$ class.

- **Exponential Kernel:**

  The absolute exponential kernel is given by

$$k_{exp}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2 e^{-\frac{|\boldsymbol{x}_i - \boldsymbol{x}_j|}{l}}, \tag{3.17}$$

  where the hyperparameters play a similar role as that of the SE kernel, with the main difference that it is only continuous but not differentiable.

- **Whitenoise Kernel:**

The whitenoise kernel is given by

$$k_{white}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma_n^2 \delta_{ij} \tag{3.18}$$

It has a simple form that models the noise as independently and identically normally-distributed. The parameter $\sigma_n^2$ simply equals the variance of the noise.

### 3.2.4 Kernel Optimization

As seen in the previous section, the kernel is only specified by the hyperparameters [3]. While in principle the GPR mean (the $\mu_{post}$ in Eq 3.14) does not require an explicit optimization of the hyperparameters, doing so usually improves the prediction results in practice.

As in most regression or supervised ML methods, a cost function needs to be designed, in which the optimal (hyper)parameters are those that minimizes or maximizes the cost function. Some of the common cost functions include the root mean squared displacement or mean absolute error.

A natural cost function in the GPR framework is the marginal log-likelihood:

$$\begin{aligned} \log(\mathrm{p}(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})) &= \int \mathrm{p}(\mathbf{y}|\mathbf{f}, \mathbf{X})\mathrm{p}(\mathbf{f}|\mathbf{X})\mathrm{d}\mathbf{f} \\ &= -\frac{1}{2}\mathbf{y}^{\mathbf{T}}(\mathbf{K} + \mathbf{\alpha e_n^2 1})^{-\mathbf{1}}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \mathbf{\alpha e_n^2 1}| - \frac{n}{2}\log(2ß) \end{aligned} \tag{3.19}$$

[3] They are called hyperparameters (as supposed to parameters) because they are parameters of a non-parametric model

where $\boldsymbol{\theta}$ is the set of kernel hyperparmeters that needs to be optimized, and the marginalization is over the function values $\mathbf{f}$, which has an analytical solution due to all the distributions being Gaussian. The marginal likelihood has a simple interpretation: The first term encourages the model fit to the data, while the second term is the complexity penalty and last term is a normalization constant [32]. The optimal hyperparameter values are thus those that maximizes the marginal log-likelihood.

# 4

# Learning Coupled Potential Energy Surfaces

## 4.1 Introduction

This Chapter is based on the results and work published in Machine Learning Seams of Conical Intersection: A Characteristic Polynomial Approach. The chapter reuses a large portion of the work.



Adiabatic PESs are the natural quantities used to rationalize many chemical processes. The generation of high quality PESs, however, is a non-trivial task, and is heavily dependent on the level of electronic structure theory used, which is often computationally expensive for accurate *ab initio* methods. The advent of machine learning offers a promising way to accelerate the generation of high quality PESs with reduced quantum chemistry calculations [33–37]. While machine learning of PESs has seen rapid advances in recent years, majority of the work has focused on the learning of isolated ground state PESs. Extension of standard machine learning methods from the ground state to coupled excited state PESs poses challenges due to the ubiquitous existence of ConIns seams, whose description was outlined in section 2.2. The existence of an $n$-state ConIns seam means that the corresponding adiabatic PESs belong to the $C^0$ differentiability class over any domain containing it. This lack of differentiability, or non-smoothness, of the PESs on the locus

of seam points is problematic for many standard machine learning approaches, including kernel methods, and neural networks employing gradient-based optimization. This has traditionally been overcome by learning the diabatic surfaces instead [38–40] , however, diabatic representations are not unique and constructing a general adiabatic to diabatic transformation is a non-trivial task.

In this chapter, we show that this can be overcome instead by learning the nuclear coordinate dependent characteristic polynomial coefficients of the potential matrix, which will be outlined in detail in section 4.2. This proposal follows directly from the previous work of Opalka and Domcke who used expansions of the CP coefficients in terms of permutationally invariant polynomials to fit models of intersecting adiabatic potentials [41]. Using *ab intio* MR-CIS calculations, we learn the coupled PESs with the characteristic polynomial coefficients at the MECI of three prototypical systems: the accidental $S_0/S_1$ ConIns of twisted-pyramidalized ethylene, the symmetry required $(E \otimes e)$ two-state ConIns between the two components of the $D_1$ state of $NH_3^+$, and the symmetry-required $(T_2 \otimes (e \oplus t_2))$ three-state intersection between the components of the $D_0$ state of $CH_4^+$.

## 4.2 Characteristic Polynomial Coefficients

This section introduces the characteristic polynomial coefficient and the Frobenius companion matrix, and shows a few important properties relevant to its application to the learning of adiabatic energies.

The characteristic polynomial (CP) $p_{\boldsymbol{A}}(x)$ of an $m \times m$ invertible matrix $\boldsymbol{A}$ is defined as the m$^{\text{th}}$ order monic [1] polynomial

$$p_{\boldsymbol{A}}(x) \equiv \det(x\mathbf{1} - A) = \prod_{i=1}^{m}(x - \lambda_i) = \sum_{i=0}^{m} c_i x^i, \tag{4.1}$$

of which its roots[2] are the eigenvalues of $\boldsymbol{A}$. In Eq. 4.1, $\{\lambda_i, i = 1, \ldots, m\}$ are the eigenvalues of $\boldsymbol{A}$, and $\{c_i, i = 0, \ldots, m\}$ are the CP coefficients. The set of CP coefficients corresponds to the elementary symmetric polynomials of the eigenvalues up to a sign, where the k$^{\text{th}}$ elementary symmetric polynomial of $m$ eigenvalues is expressed as [42]

$$e_k(\lambda_1, \ldots, \lambda_m) = \sum_{1 \le i_1 \le i_2 \le \ldots \le i_k \le n} \lambda_{i_1} \ldots \lambda_{i_k}, \tag{4.2}$$

There are up to $k = 0, \ldots, m$ non-zero elementary symmetric polynomials, with the $k = 0$ elementary symmetric polynomial having a simple form $e_0(\lambda_1, \ldots, \lambda_m) = 1$. For

[1] A monic polynomial is defined such that its leading coefficient $c_m = 0$

[2] The roots of polynomial are the set of values $\{\lambda\}$ such that the polynomial maps to zero. i.e. the set of $\lambda_i$ such that $p_{\boldsymbol{A}}(\lambda_i) = 0, i = 0, \ldots, m$

example, let $\boldsymbol{A}$ be a $3 \times 3$ matrix with three eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$. Then the elementary symmetric polynomials are (excluding $e_0(\lambda_1, \ldots, \lambda_m) = 1$)

$$\begin{aligned} e_1(\lambda_1, \lambda_2, \lambda_3) &= \lambda_1 + \lambda_2 + \lambda_3 \\ e_2(\lambda_1, \lambda_2, \lambda_3) &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 \\ e_3(\lambda_1, \lambda_2, \lambda_3) &= \lambda_1\lambda_2\lambda_3 \end{aligned} \tag{4.3}$$

The CP coefficients $\{c_i, i = 0, \ldots, m\}$ are related to the elementary symmetric polynomials by the following relationship [42]:

$$e_k = (-1)^k c_{m-k} \tag{4.4}$$

To illustrate this, consider a simple $2 \times 2$ matrix $\boldsymbol{A}$ with eigenvalues $\{\lambda_1, \lambda_2\}$. The corresponding CP of A is

$$p_{\boldsymbol{A}}(x) = (x - \lambda_1)(x - \lambda_1) = x^2 - (\lambda_1 + \lambda_2)x + \lambda_1\lambda_2, \tag{4.5}$$

and the elementary symmetric polynomials of the eigenvalues are

$$\begin{aligned} e_1(\lambda_1, \lambda_2) &= \lambda_1 + \lambda_2, \\ e_2(\lambda_1, \lambda_2) &= \lambda_1\lambda_2, \end{aligned} \tag{4.6}$$

which comparing Eq.4.5 and Eq.4.6 shows that $e_0 = (-1)^0 c_2$, $e_1 = (-1)^1 c_1$ and $e_2 =$

$(-1)^2 c_0$.

An important property of CPs are their invariance under similarity transforms, meaning that the set of CP coefficients $\{c_i\}_{\boldsymbol{V}}$ for the matrix $V$ and the set of CP coefficients $\{c_i\}_{\boldsymbol{W}}$ for the matrix $\boldsymbol{W} = \boldsymbol{U^{-1}VU}$ for some invertible $U$, are equivalent:

$$\begin{aligned} \det(x\mathbf{1} - \boldsymbol{W}) &= \det(\boldsymbol{U^{-1}}x\boldsymbol{U} - \boldsymbol{U^{-1}VU}) \\ &= \det(\boldsymbol{U^{-1}}(x\mathbf{1} - \boldsymbol{V})\boldsymbol{U}) \\ &= \det(x\mathbf{1} - \boldsymbol{V})\det(\boldsymbol{U^{-1}U}) \\ &= \det(x\mathbf{1} - \boldsymbol{V}) \end{aligned} \tag{4.7}$$

Just as the CP casts an eigenvalue problem to a root finding problem, the inverse can also be inquired. That is, given a polynomial $p(x) = \sum_{i=0}^{m} c_i x^i$, what is the corresponding matrix $\boldsymbol{C}$ whose eigenvalues are the roots. One of the possible method is by constructing the so called Frobenius companion matrix [42]:

$$\boldsymbol{C} = \begin{pmatrix} 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & \dots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -c_{m-1} \end{pmatrix}, \tag{4.8}$$

the diagonalization and eigenvalues of which are the roots of $p(x)$.

## 4.3 Computational Details

Before showing the results and discussion, this section defines the computational details used in the study. Namely, the electronic structure method employed, the sampling scheme used to generate the training and test geometries, and the GPR details used to learn the relevant quantities.

### 4.3.1 Quantum Chemistry Calculations

The *ab inito* MR-CIS single point energies are computed using the COLOMBUS quantum chemistry package [43]. The MECI geometries about which the adiabatic energies were sampled were computed at the same (MRCI) level of theory. In these calculations, complete active space self-consistent field (CASSCF) calculations were used to generate the molecular orbital bases and to provide the MRCI reference spaces. CAS(2,2), CAS(5,3), and CAS(5,3) reference spaces were used for $C_2H_4$, $NH_3^+$, and $CH_4^+$, respectively. In all CASSCF calculations, state averaging was performed over the three lowest-lying states. MRCI configurations were then generated by allowing all single excitations out of the reference space. The cc-pVDZ basis was used in all calculations.

### 4.3.2 Latin Hybercube Sampling

The training data is generated with atom-centered Cartesian Latin hypercube sampling (LHS). LHS is a space-filling design which seeks to spread out the sampling whilst main-

taining some degree of random sampling [44]. That is, rather than randomly or uniformly sampling over the space, LHS seeks to find a balance. In brief, this is done by partitioning the volume evenly into cubes, and each dimension into equally sized segments, and ensuring that each segment contains exactly one point [44]. For this thesis, the space to be sampled is the $3N$-dimensional hypercube with the origin corresponding to a specific nuclear configuration, and the bounds being displacements from the origin. For the following studies, the origin simply corresponds to the MECI geometries obtained using COLUMBUS, and the bounds are displacements of each atomic Cartesian coordinate by $\pm 0.05$Å. The scikit-learn library [45] was used perform the LHS.

### 4.3.3 Choice of Kernel and Hyperparameter Optimization

The SOAP descriptor outlined in Section 3.1 is used to represent the molecular structure. The SE kernel was used, which was introduced in Section 3.2.3. The $\omega$-CP surfaces are hypothesized to be smooth and slowly varying functions of nuclear coordinates, which can also be visually appreciated in Fig. 4.1. We thus consider kernels that learn smooth function, and the "default" choice is the SE kernel due to its simplicity. While there are certainly more kernel options, starting with a SE kernel showed to give accurate results, as will be shown in subsequent sections. We consider the ability for a SE kernel to describe the $\omega$-CP surfaces despite its simplicity a strength rather than weakness of the proposed method. The scikit-learn library [45] was used to construct the GPR surrogates.

In this section, the SOAP parameters were optimized in tandem with the kernel hy-

perparameters. Specifically, the combined parameters $\{r, n_{max}, l_{max}, \sigma, l\}$ are optimized simultaneously. Since the spherical harmonic expansion terms $n_{max}, l_{max}$ only take integer values, however, standard gradient based optimization cannot be employed. To overcome this, a genetic algorithm (GA) is used. The GA search algorithm follows a generic procedure of creating a population, selecting top candidates based on chosen metric and random selection, then allowing combination and random mutation which form a new set of population. The process is repeated until an optimal solution is found. First, an initial population is generated by randomly sampling from the bounded set of parameter values:

- $r_{cut} \epsilon [1, 9] \subset \mathbb{R}$
- $n_{max} \epsilon [2, 9] \subset \mathbb{Z}^+$
- $l_{max} \epsilon [2, 9] \subset \mathbb{Z}^+$
- $\sigma \epsilon [10^{-3}, 10^{-1}] \subset \mathbb{R}$
- $l_{kernel} \epsilon [10^{-5}, 10^{-1}] \subset \mathbb{R}$

Once a population (parameter set) is generated, a metric is used to score each individual (sets) in the population, here taken as the root mean squared displacement (RMSD) of the GPR model predictions relative to a test set. The training/test sets were generated by randomly splitting the data set into 15%/85% subsets. Once scored by their RMSDs, survival of the fittest is applied, and from the remaining survivors, crossover and mutation operations are performed to generate a new population (set of parameters) of the same

length, and the whole process is repeated until convergence. Simulated Binary Crossover (SBX) [46] and Polynomial Mutation (PM) [46] algorithms were used for the crossover and mutation procedure, respectively. The GA search is done with the pymoo library [47]

## 4.4 Characterizing Branching Spaces with Characteristic Polynomial Coefficients

To connect the CP framework to the learning of adiabatic PESs, we begin with the decomposition of the diagonal adiabatic potential matrix with elements corresponding to the adiabatic PESs, into an average energy and traceless (by construction) splitting matrix [41]:

$$\boldsymbol{V}(\boldsymbol{R}) = \begin{bmatrix} E_1(\boldsymbol{R}) & 0 & \dots & 0 \\ 0 & E_2(\boldsymbol{R}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & E_n(\boldsymbol{R}) \end{bmatrix} = \omega(\boldsymbol{R})\mathbf{1}_n + \boldsymbol{Z}(\boldsymbol{R}) \tag{4.9}$$

where $\omega = \frac{1}{n}\sum_{i=1}^{n} E_i$ is the average energy of $n$ adiabatic energies, $\mathbf{1}_n$ is the $n \times n$ identity matrix, and $\boldsymbol{Z}$ is called the splitting matrix, with elements $Z_{ij} = (E_i - \omega)\delta_{ij}$. Since the average energy is a smooth function of nuclear coordinates, which we state as a fact for now and we will discuss shortly, this decomposition separates out the component in the potential matrix that give rise to the discontinuous derivatives (cusps) along the seams of CIs. To overcome the non-smoothness in the splitting term $\boldsymbol{Z}$, which prohibit standard machine learning techniques, the CP coefficients of $\boldsymbol{Z}$ is learned instead,

$$p^Z(\lambda) = \det\left[\lambda\mathbf{1}_n - \boldsymbol{Z}(\boldsymbol{R})\right] = \sum_{i=0}^{n-1} c_i^Z(\boldsymbol{R})\lambda^i, \tag{4.10}$$

with CP coefficients $c_i^Z(\boldsymbol{R})$ given by Eq.4.4, and below a few terms are explicitly listed:

$$\begin{aligned} c_0^z(\boldsymbol{R}) &= (-1)^n \prod_i^n Z_{ii}(\boldsymbol{R}) \\ &\vdots \\ c_{n-2}^z(\boldsymbol{R}) &= \sum_{i<j}^n Z_{ii}(\boldsymbol{R}) Z_{jj}(\boldsymbol{R}) \\ c_{n-1}^z(\boldsymbol{R}) &= -\sum_i^n Z_{ii}(\boldsymbol{R}) = 0 \\ c_n^z(\boldsymbol{R}) &= 1 \end{aligned} \tag{4.11}$$

Given the CP coefficients, the adiabatic energies can then be easily obtained as the eigenvalues of the following Frobenius companion matrix defined in Eq.4.8:

$$\boldsymbol{C}(\boldsymbol{R}) = \begin{bmatrix} \omega(\boldsymbol{R}) & 0 & \cdots & 0 & 0 & -c_0^z(\boldsymbol{R}) \\ 1 & \omega(\boldsymbol{R}) & \cdots & 0 & 0 & -c_1^z(\boldsymbol{R}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \omega(\boldsymbol{R}) & -c_{n-2}^z(\boldsymbol{R}) \\ 0 & 0 & \cdots & 0 & 1 & \omega(\boldsymbol{R}) \end{bmatrix} \tag{4.12}$$

where the $-c_{n-1}^z$ term is zero because $\boldsymbol{Z}$ is traceless. Since $\omega \mathbf{1}_n$ commutes with $\boldsymbol{Z}$, the eigenvalue of the potential matrix (which is the adiabatic energies) is simply $\omega + z_i$. In Eq.4.12, the $\omega$ is added into the diagonal for a compact representation, but only the eigenvalues of the companion matrix for $p_{\boldsymbol{Z}}$ is needed. For clarity, there are $n-1$ number

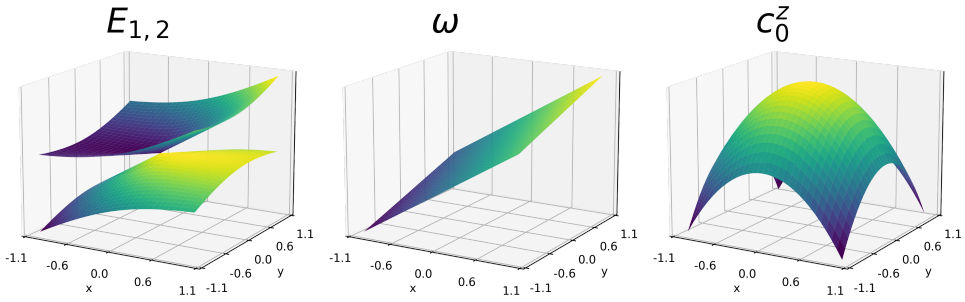


Figure 4.1: The left panel shows the adiabatic PESs at a conical intersection given by Eq.2.23. The middle and right panel shows the corresponding $\omega$ and $c_0^z$ surface. It shows the removable of the cusp and smoothness of the functions compared to the adiabatic PESs

of non-trivial CP coefficients that needs to be fitted for an $n$-state CI due to $-c_{n-1}^z = 0$ plus the $\omega$ surface. For example, a two-state CI would have only $c_0^z(\boldsymbol{R})$ and the average energy $\omega(\boldsymbol{R})$ surface, and a three-state CI would have $c_0^z(\boldsymbol{R})$, $c_1^z(\boldsymbol{R})$ and the $\omega(\boldsymbol{R})$ surface. The method of learning the CP coefficients of $\boldsymbol{Z}$ and the average energy $\omega$ is herein referred to as the $\omega$-CP model.

The first and main advantage of the $\omega$-CP approach is that the CP coefficients are smooth functions of the nuclear coordinates. As an example, the $c_0^z$ and $\omega$ surface for the two-state model (Eq.2.23) is shown in figure 4.1, which visually shows the smoothness of the $c_0^z$ and $\omega$ surfaces in the BS.

The motivation behind learning the CP coefficients is its invariant with respect to the choice of electronic basis. That is, the potential matrix in a diabatic basis $\boldsymbol{W}$ is given by

$$\boldsymbol{W} = \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S} \tag{4.13}$$

where $\boldsymbol{S}$ is the adiabatic-to-diabatic transformation introduced in Section 2.1.1. Since the CP coefficients are similarity transform invariant (by Eq.4.7), the same set of CP coefficients are generated in the adiabatic representation as for a diabatic representation. The same holds for the $\omega(\boldsymbol{R})$ term, being as it is the trace of the potential matrix, which is also invariant to the choice of basis[3]. Since the diabatic surfaces (which are in general not eigenvalues of the electronic Hamiltonian) are smooth functions of nuclear coordinates, the CP coefficients, which are sums and products of diabatic surfaces, are also expected to be smooth functions.

A second reason is that all of the BS information is contained in the single CP coefficient surface $c^z_{n-2}(\boldsymbol{R})$, irrespective of the number of intersecting states. Recall that in lieu of direct access to the BS vectors (Eq.2.18), they can also be obtained by considering the eigenpairs of the generalized energy difference squared surface (Eq.2.29). That is, the eigenvectors with non-zero (zero) eigenvalues form the branching (seam) space.

It was determined in the work of this thesis that the following relation $c^z_{n-2} = \frac{-1}{2n}\sum^n_{i<j}\Delta E^2$ holds, thus the CP coefficient $c^z_{n-2}$ captures all of the branching space information. The derivation is as follows:

Eq.2.29 is first expanded out to have the same elements as the splitting matrix:

[3]Note that the average energy, which is the trace of the potential matrix divided by the number states, also corresponds to the $c_{n-1}$ coefficient of the potential matrix

$$\sum_{i<j}^{n}(E_i - E_j)^2 = \sum_{i<j}^{n}\left[(E_i - \omega) - (E_j - \omega)\right]^2 = \sum_{i<j}^{n}(Z_{ii} - Z_{jj})^2 \tag{4.14}$$

Expanding out Equation 4.14 and multiplying out the summation we obtain

$$\sum_{i<j}^{n}(E_i - E_j)^2 = \sum_{i<j}^{n}(Z_{ii}^2 + Z_{jj}^2) - 2\sum_{i<j}^{n} Z_{ii}Z_{jj} \tag{4.15}$$

Taking the first term on the right-hand side of Equation 4.15, expanding it, and simplifying the double ordered summation to a single summation, we obtain

$$\sum_{i<j}^{n}(Z_{ii}^2 + Z_{jj}^2) = (n-1)\sum_{i}^{n} Z_{ii}^2 \tag{4.16}$$

We now make use of the trace property of a Kronecker product

$$Tr(\boldsymbol{Z} \otimes \boldsymbol{Z}) = Tr(\boldsymbol{Z})Tr(\boldsymbol{Z}) = 0, \tag{4.17}$$

where the last equality in Equation 4.17 holds because the diagonal splitting matrix $\boldsymbol{Z}$ is by construction traceless. We now note that the trace of a tensor product of the $n \times n$ diagonal matrix $\boldsymbol{Z}$ with itself can be written as a sum of the form

$$Tr(\boldsymbol{Z} \otimes \boldsymbol{Z}) = \sum_{i}^{n} Z_{ii}^2 + 2\sum_{i<j}^{n} Z_{ii}Z_{jj} \tag{4.18}$$

Setting Equation 4.18 to zero from the result of Equation 4.17 gives

$$\sum_{i}^{n} Z_{ii}^2 = -2 \sum_{i<j}^{n} Z_{ii} Z_{jj} \tag{4.19}$$

Substituting Equation 4.19 into the right-hand side of Equation 4.16, we obtain the relationship

$$\sum_{i<j}^{n} (Z_{ii}^2 + Z_{jj}^2) = -2(n-1) \sum_{i<j}^{n} Z_{ii} Z_{jj} \tag{4.20}$$

Equation 4.20 can now be substituted into Equation 4.15. and after simplifying we get the result

$$\sum_{i<j}^{n} (E_i - E_j)^2 = -2n \sum_{i<j}^{n} Z_{ii} Z_{jj} \tag{4.21}$$

The sum term on the right-hand side of Equation 4.21 is precisely the definition of the $c_{n-2}^{Z}$ CP coefficient,

$$c_{n-2}^{Z} = \sum_{i<j}^{n} Z_{ii} Z_{jj}. \tag{4.22}$$

Thus, the desired result is obtained:

$$c_{n-2}^{Z} = -\frac{1}{2n} \sum_{i<j}^{n} (E_i - E_j)^2 \tag{4.23}$$

## 4.5 Reproduction of Conical Intersections

Before examining $\omega$-CP model's ability to characterize the branching and seam spaces, we first evaluate the model's ability to reproduce the adiabatic energies in the vicinity of the MECI by comparing it to the direct learning of the adiabatic PESs, which we refer to as the direct-energy model. Specifically, we compare the mean absolute errors (MAEs) [4] of the adiabatic energies in the vicinity of the MECI geometries as a function of training set size. Figure 4.2 shows the log MAE as a function of the training set size.

For all the molecules studied, sub-chemical accuracy [5] can be achieved with around 300 training points in both the direct-energy and $\omega$-CP model. An important difference, however, is that the PESs obtained from the $\omega$-CP model are consistently more accurate at a fixed training size compared to the direct-energy model. The consistently better performance is accentuated at ConIns as the cusp cannot be well modeled in the direct-energy model. This is evident when plotting the energies along the BS vectors for both the $\omega$-CP and direct-energy models, which is shown in Fig.4.3. It can be seen that the cusp as well as the energies are well produced at the locus of a ConIns for the $\omega$-CP model. On the other hand, not only does the direct-energy model fail reproduce the cusp, it learns an avoided crossing instead, where the degeneracy is completely absent. As a result of the avoided crossing, large errors accumulate in the vicinity of the ConIns. It should be noted here that since the energies are tightly sampled (evaluated from LHS geometries

[4]The MAE is the error metric given by $\frac{1}{n}\sum_i^n|\tilde{E}_i - E_i|$, where $\tilde{E}_i$ and $E_i$ is the predicted and actual energies, respectively

[5]Sub-chemical accuracy is defined as 0.04 eV. In Fig.4.2 it is taken as log(0.04)

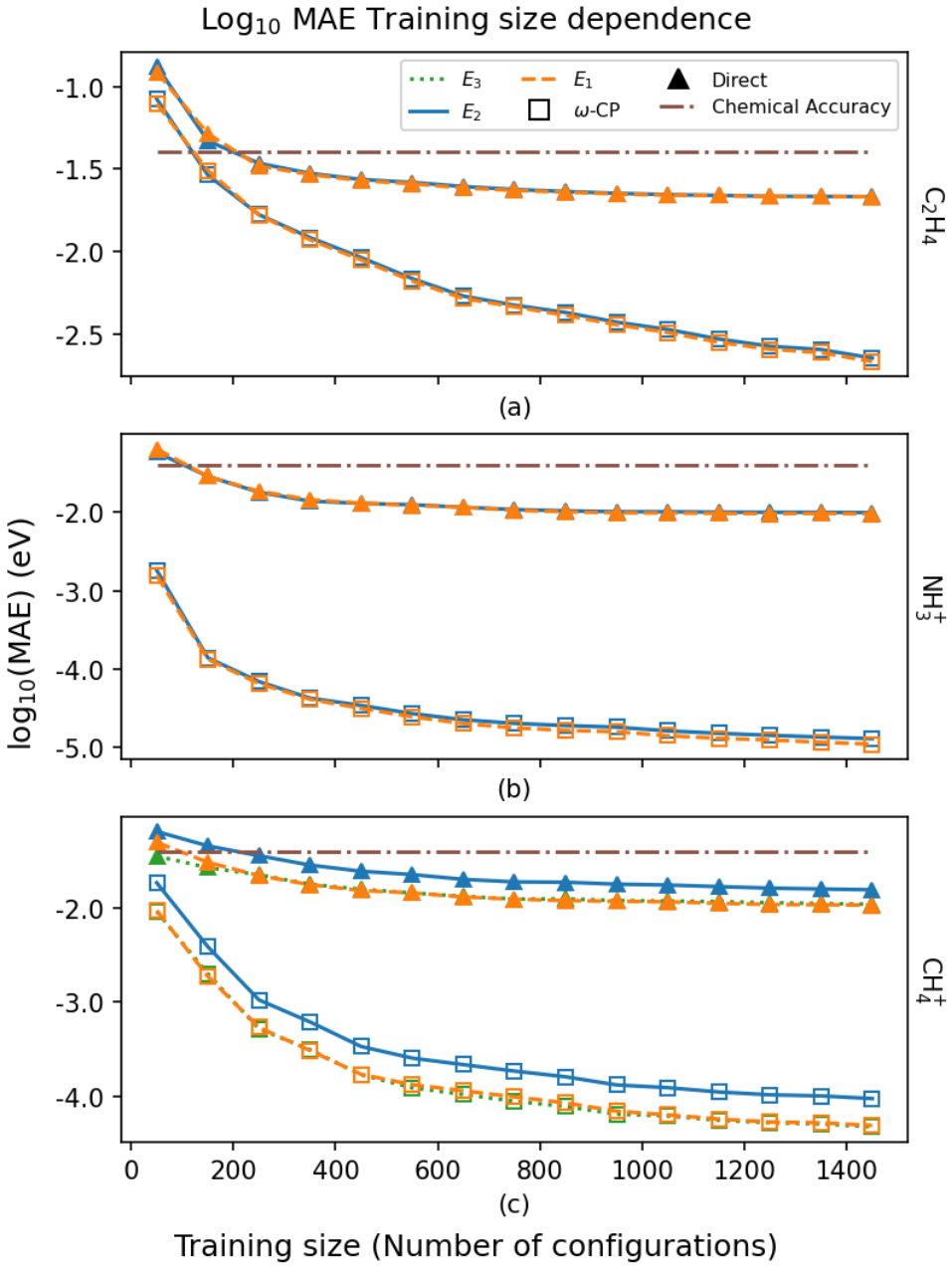


Figure 4.2: log of Mean absolute error as a function of training set size, in terms of number of nuclear configurations, when fitting the characteristic polynomial parameters ($\omega$-CP) or the adiabatic energies (Direct).

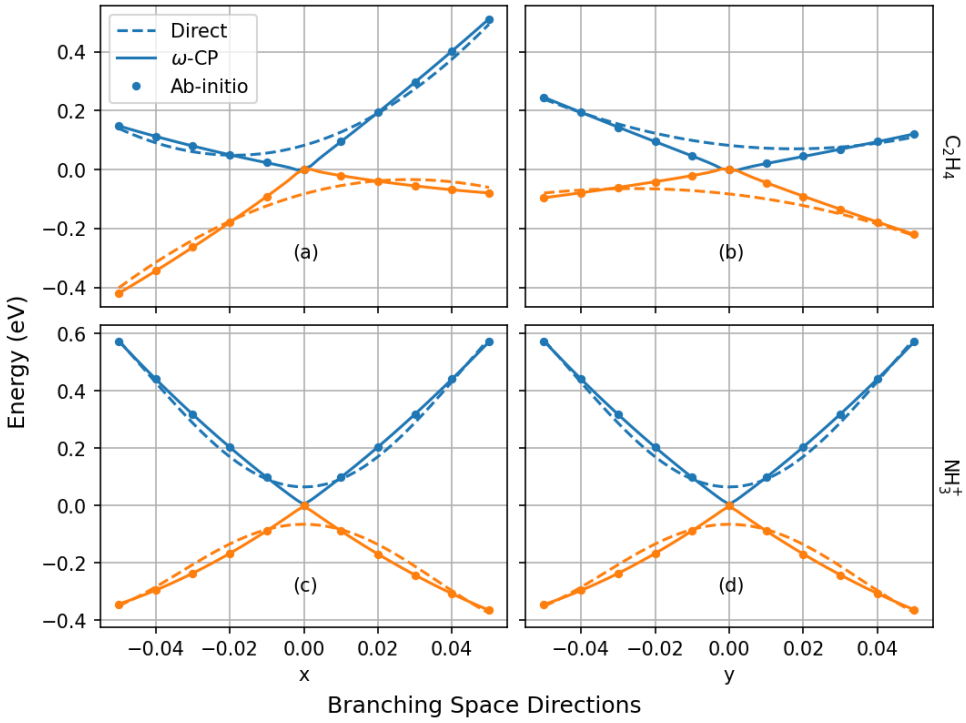


Figure 4.3: Adiabatic potentials in the immediate vicinity of a minimum energy conical intersection as determined from *ab initio* computations (isolated points) and the $\omega$-CP and direct energy surrogate potentials. The latter fail to capture the degeneracy and instead show an avoided crossing.

with bounds of $\pm 0.05$), it is not indicative from Fig.4.2 alone either method's ability to extrapolate away from the training region. This will be shown and discussed in Section 4.6.

As mentioned in Section 3.2.3, the smoothness and differentiability of a learned GPR surrogate is determined by the choice of kernel. To ensure that the avoided crossing in the direct-energy model is not an artifact, since the SE kernel only learns an infinitely-differentiable function, a kernel of the same differentiability class as the seams of ConIns

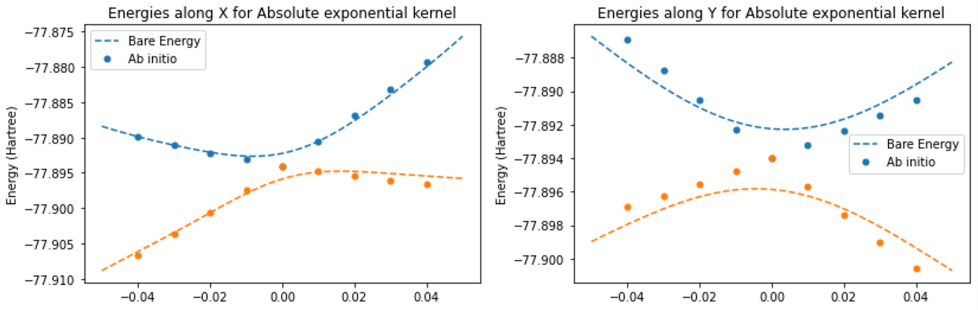


Figure 4.4: Adiabatic potential in the vicinity of a minimum energy conical intersection as determined from *ab initio* computations and the direct-energy model. The direct-energy model is learned with the absolute exponential kernel here instead. Even with a kernel belonging to the same differentiability class, the avoided crossing is reproduced in the direct energy model

is also used, namely the absolute exponential kernel

$$k_{abs}(x_1, x_2) = \sigma^2 e^{-\frac{|x_1 - x_2|}{l}} \tag{4.24}$$

which is non-differentiable.

The energies plotted along the BS cuts using the absolute exponential kernel is shown in Figure 4.4, which shows that even with a non-differentiable kernel, the cusp cannot be reproduced. While more sophisticated kernels such as additive or combined kernels can be conceived, which may enable the direct-energy model to better approximate or fully capture the seams of ConIns, the simplicity of the $\omega$-CP model is seen as a strength and not a weakness. Namely, a simple isotropic SE kernel can quantitatively capture the cusp.

## 4.6 Ability to learn Branching Spaces

In the previous section, the figure showing the energies along the BS cuts (Fig.4.3) was slightly putting the cart before the horse, as the direct-energy and $\omega$-CP model's ability to reproduce the BS vectors have not been examined. In this section, we first compare the ability for the direct-energy and $\omega$-CP model to reproduce the branching and seam space dimensions, which is achieved by examining the eigenvalues of the $c_0^z$ Hessian at a MECI, shown in Figure 4.5.

It is evident from Fig.4.5 that aside from a worse MAE and learning an avoided crossing, the direct-energy model has a more fundamental issue of getting the incorrect branching space dimensions. That is, the number of non-zero eigenvalues should be strictly two and five for the two-state ($C_2H_4$ and $NH_3^+$) and three-state ($CH_4^+$) ConIns. The presence of small nonzero eigenvalues in the direct-energy model implies that there are more than two linearly independent vectors that lifts degeneracy. This could not only lead to incorrect photodynamics due to qualitatively incorrect PESs, but also result in erroneous geometric phases if the wavefunction encircles additional artificial ConIns, thus changing the parity of the phase [2].

While Fig.4.5 is indicative of the model quality, it does not give information on the learned BS vectors themselves. Recall that the normalized BS vectors is obtained as the eigenvectors associated with the non-zero eigenvalues. To check that the GPR learned and *ab initio* BS vectors coincide, the angle between the $x$ ($\theta_x$) and $y$ ($\theta_y$) BS vectors

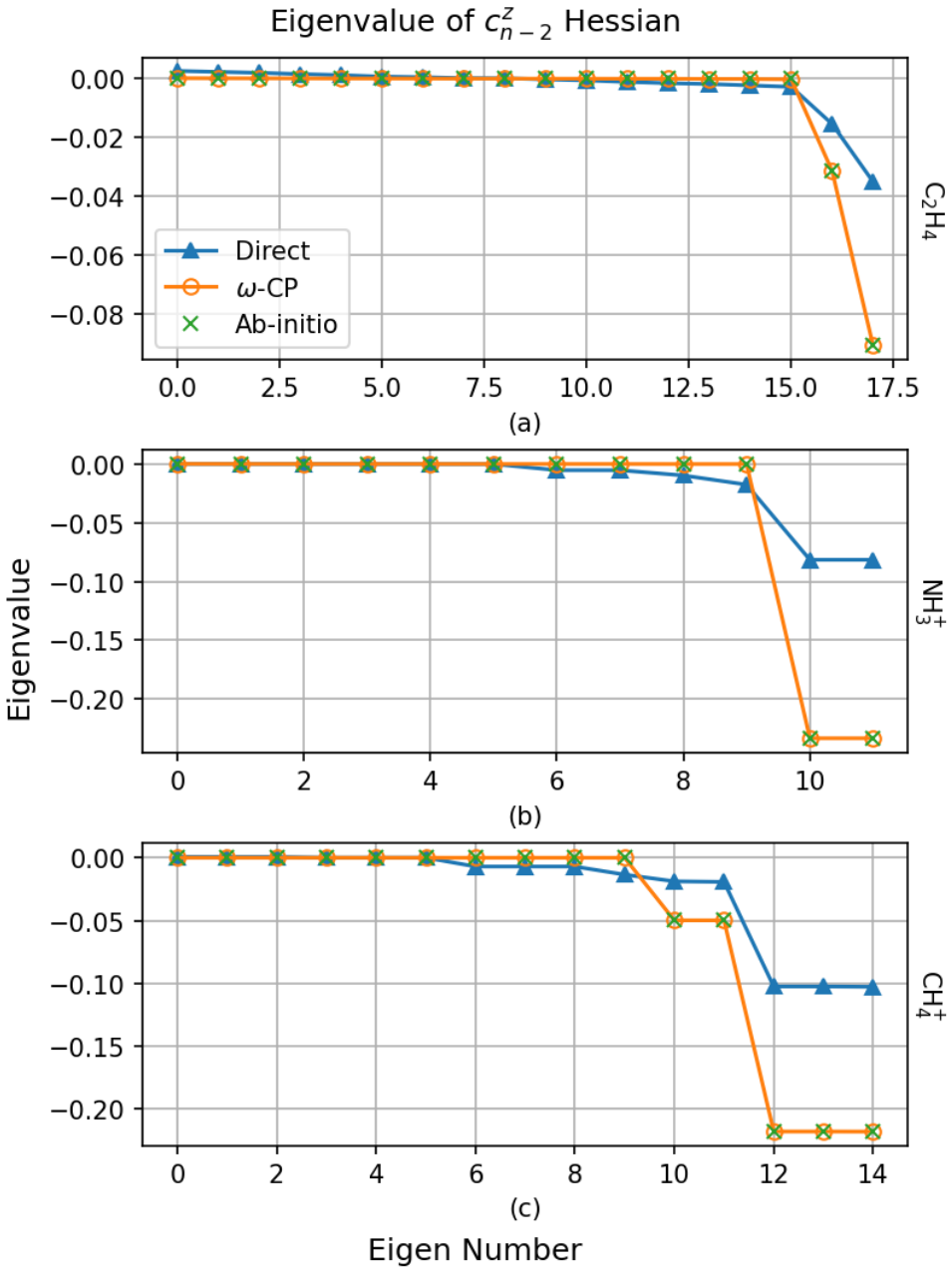


Figure 4.5: Eigenvalues of the $c^z_{n-2}$ Hessian matrix as computed from *ab initio* data and using the $\omega$-CP and direct-energy models. The number of non-zero eigenvalues corresponds to the dimension of the branching space.

| | | $\theta_x$ | $\theta_y$ | $\tilde{g}$ | $\tilde{h}$ | $s_x$ | $s_y$ |
|---|---|---|---|---|---|---|---|
| $C_2H_4$ | *ab initio* | 0.0 | 0.0 | 0.213 | 0.125 | -0.130 | -0.046 |
| | $\omega$-CP | 0.0 | 0.3 | 0.214 | 0.126 | -0.130 | -0.046 |
| | direct-energy | 3.7 | 8.0 | 0.130 | 0.084 | 0.128 | -0.050 |
| $NH_3^+$ | *ab initio* | 0.0 | 0.0 | 0.342 | 0.342 | 0.000 | 0.000 |
| | $\omega$-CP | 0.3 | 0.3 | 0.342 | 0.342 | 0.000 | 0.000 |
| | direct-energy | 5.3 | 5.3 | 0.202 | 0.202 | 0.000 | 0.001 |

Table 4.1: Comparison of the branching space parameters determined from *ab initio* data and using the surrogate potentials. The parameters $\tilde{g}$ and $\tilde{h}$ are the norms of the orthogonalized gradient difference and non-adiabatic coupling vectors, respectively, and the $\theta_i$, $i = x, y$ are the angles between the *ab initio* and $\omega$-CP and direct-energy GPR model branching space vector vectors $x$ and $y$. Angles are given in units of degrees. All other values are given in units of $E_h/Å$.

are computed. In addition, the intersection topography $g, h, s_x, s_y$ defined in Eq.2.21 and Eq.2.20 for both method is also computed, where the BS vector norms can be computed from the eigenvalue as $g, h = \sqrt{\frac{\lambda_{1,2}}{8}}$ with $\lambda_{1,2}$ being the two non-zero eigenvalues, which was shown in Eq.2.28. The intersection topographies and BS vector angles are summarized in Table 4.1.

The $\omega$-CP models are found to quantitatively reproduce the ConIns topography in both cases, with branching vector angles $\theta_{x,y}$ of less than 0.5° and parameters $\tilde{g}$, $\tilde{h}$ and $s_{x,y}$ in almost perfect agreement with the *ab initio* MRCI values. Additionally, it was found that the CP-coefficients are able to reproduce the BS with very modest training set sizes, as shown in Table 4.2, achieving reasonable accuracy with training set sizes as small as 500 *ab initio* points. The direct-energy models, on the other hand, fare less well, with maximum branching vector angles $\theta_{x,y}$ of 8.0° and parameters $\tilde{g}$, $\tilde{h}$ and $s_{x,y}$ that fail to correctly describe the ConIns topography, even with large training sets containing many thousands

| $C_2H_4$ | | $\theta_x$ | $\theta_y$ | $\tilde{g}$ | $\tilde{h}$ | $s_x$ | $s_y$ |
|---|---|---|---|---|---|---|---|
| *Ab initio* | | 0.0 | 0.0 | 0.213 | 0.125 | -0.130 | -0.046 |
| N=200 | $\omega$-cp | 2.4 | 8.9 | 0.216 | 0.125 | -0.130 | -0.041 |
| N=500 | $\omega$-cp | 0.9 | 1.6 | 0.216 | 0.127 | -0.130 | -0.045 |
| N=1500 | $\omega$-cp | 0.3 | 0.5 | 0.214 | 0.126 | -0.130 | -0.046 |

Table 4.2: Dependence of branching space topography parameters with respect to training set size (N). N=200, N=500 and N=1500 is shown. Reasonable accuracy is obtained at N=500 and approaches ab-initio values at N=1500

of points.

Finally, the ability for both the $\omega$-CP and direct-energy model to extrapolate the BS away from the training region is examined. An extended displacement of the adiabatic energies along the BS vectors is shown in Figure 4.6.

An interesting result is the $\omega$-CP model can accurately extrapolate the PESs to geometries outside of the training sets used to construct them. The $\omega$-CP model for both $C_2H_4$ and $NH_3^+$ remain accurate out to displacements of around $x = \pm 0.4$ and $y = \pm 0.4$, which correspond to geometries outside the span of the training sets. The difference in the training and testing regions is made clear from Figures4.7 and 4.8, which shows a superposition of the training set geometries as well as the geometries corresponding to $x$=0.4, respectively.

Somewhat remarkably, for ethylene, this geometry corresponds to ethylidene, a different structural isomer to the geometries present in the training set. Similarly, the ammonia geometry at $x$=0.4 corresponds to a near-dissociated N-H bond, which is not represented in the training set. The improved ability of the $\omega$-CP models to extrapolate is a direct

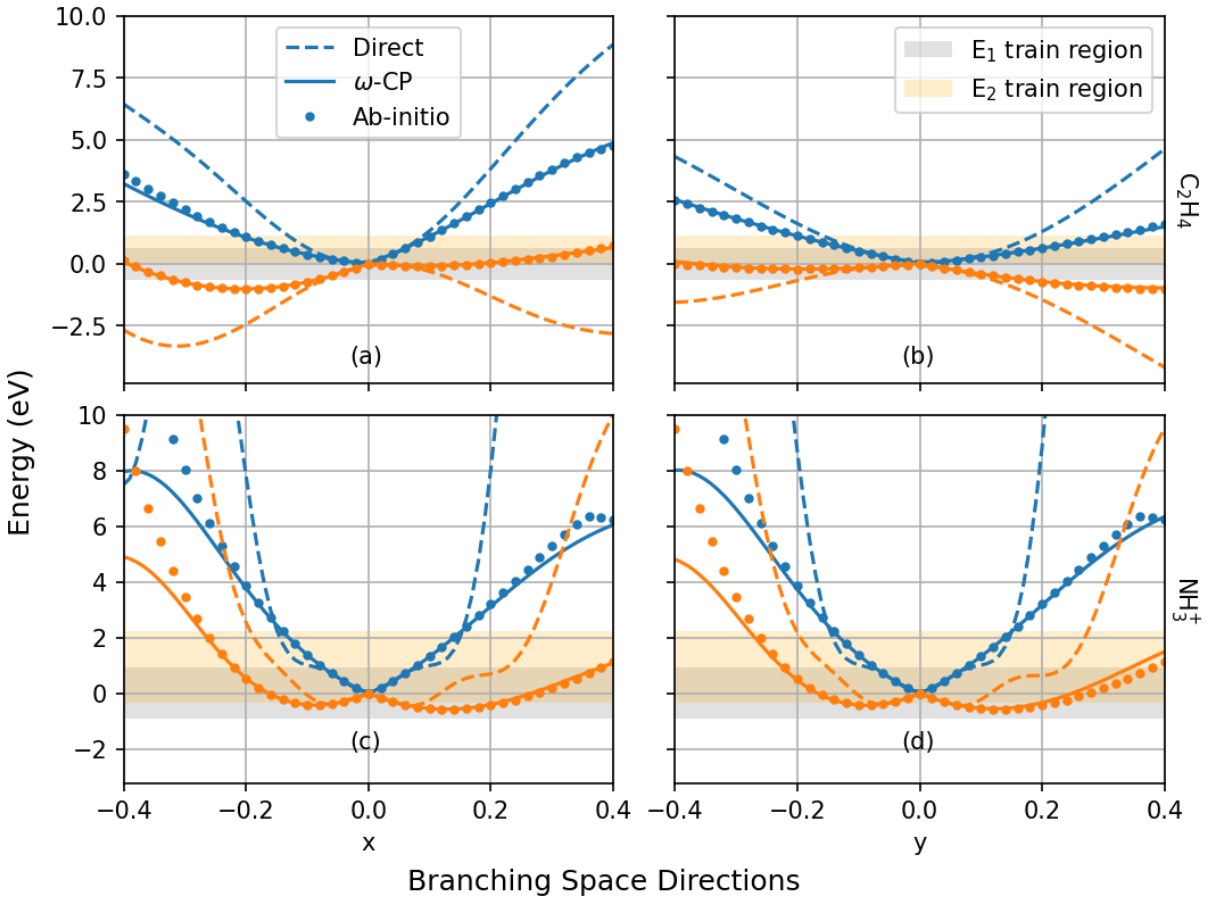


Figure 4.6: Adiabatic potential in an extended region along the BS, with the origin corresponding to the minimum energy conical intersection. The gray and beige shaded areas correspond to the $S_0$ and $S_1$ training energies, respectively. The $\omega$-CP model shows the ability to extrapolate far from the training region compared to the direct energy model.

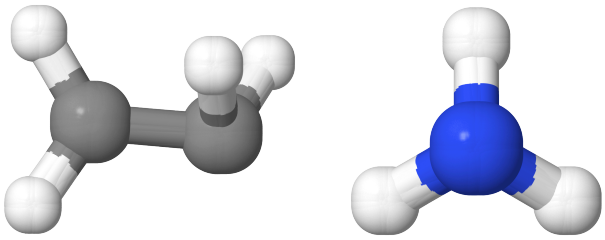

Figure 4.7: An overlap of all the training geometries for ethylene and ammonia cation. The training geometry is generated with atom centered Cartesian Latin hypercube sampling with ±0.05 Å bounds, which results in very tightly sampled configurations

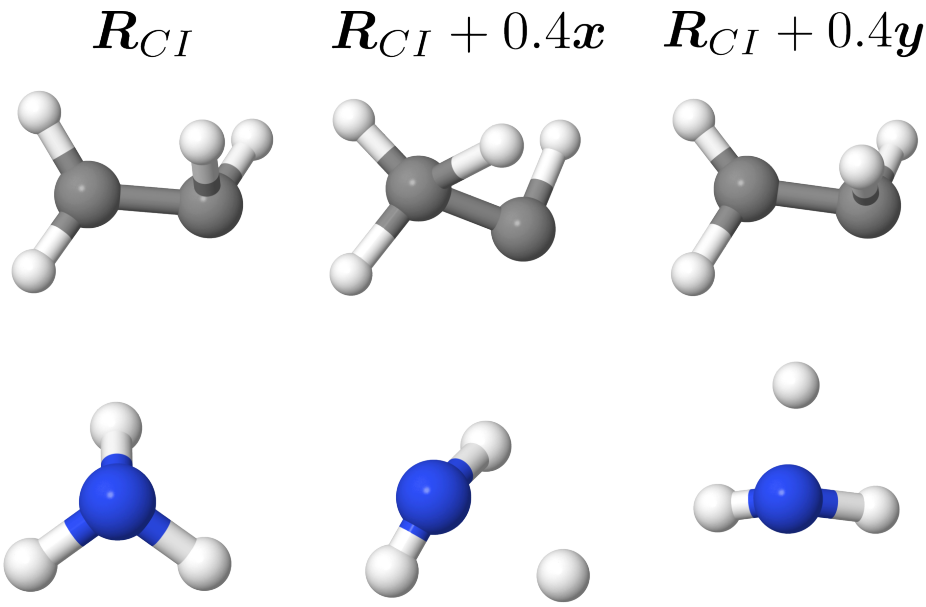


Figure 4.8: The ethylene (top row) and ammonia cation (bottom row) MECI geometries $\boldsymbol{R}_{CI}$, and its displacements along the $x$ and $y$ BS vectors at 0.4 length. The displaced ethylene MECI corresponds to a different structural isomer, and the ammonia cation resulting in a near-dissociation N-H bond.

result of the significantly longer length scales on which the surfaces $\omega(\boldsymbol{R})$ and $c_i^Z(\boldsymbol{R})$ vary compared to the PESs $E_i(\boldsymbol{R})$. That is, the $\omega(\boldsymbol{R})$ and $c_i^Z(\boldsymbol{R})$ are more slowly varying than the adiabatic energies $E_i(\boldsymbol{R})$. This, in turn results in an increased distance from which knowledge may be transferred from the training to prediction points. On the other hand, the rapidly-varying adiabatic PESs result in direct-energy GPR models that rapidly lose all predictive power when moving away from elements of the training set.

## 4.7 Conclusion

In this chapter, we have demonstrated the ability to construct quantitatively accurate GPR models of seams of ConIns via an indirect learning of the involved adiabatic PESs based on a CP formalism. The advocated $\omega$-CP approach yields a correct description of the branching and seam spaces, a feat that is harder to achieve for models based on the direct learning of adiabatic PESs. Furthermore, it is found that a single CP coefficient, $c^{Z}_{n-2}(\boldsymbol{R})$, contains all branching space information, irrespective of the number of intersecting states. The use of machine learned adiabatic PESs and non-adiabatic couplings in excited-state dynamics simulations is starting to gain traction [48, 49]. Here, there seems to be no reason to continue using directly-learned adiabatic PESs, given the unambiguous advantages of the indirect $\omega$-CP approach. We thus anticipate that the results presented here shall be of great use in directing future developments in this nascent, yet important, field of work. In terms of practical application, the $\omega$-CP formalism will be of use in a number of situations. For example, in MECI optimisation using quantum chemistry methods for which analytical gradients are not available; in such cases $\omega$-CP based surrogate potentials may straightforwardly be utilized. This is explored in the following Chapter

# 5

# DFT/MRCI Structure Optimization

## 5.1 Introduction

This Chapter is based on the results in the non peer-reviewed work (as of the writing of the thesis) Minimum Energy Conical Intersection Optimization Using DFT/MRCI(2).

Wang TY, Neville SP, Schuurman MS. Minimum Energy Conical Intersection Optimization Using DFT/MRCI(2). ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-3h77f-v2 This content is a preprint and has not been peer-reviewed.

Having established the ability for the $\omega$-CP model to quantitatively reproduce the energies and intersection topography at a ConIns, in this chapter, we apply the $\omega$-CP model to the optimization and characterization of DFT/MRCI(2) MECIs. The methodologies used in this chapter are largely similar to those developed in Chapter 4, with only a few minor modifications. However, a key difference lies in the role of the GPR surrogate.

As outlined in Section 2.3.4, the combined density functional theory and multi-reference configuration interaction (DFT/MRCI) method employs a selected-CI algorithm for the generation of the reference space, which when combined with DFT-specific corrections to the Hamiltonian matrix, results in highly compact wave function expansions. In addition, our group has developed a perturbative approximate variant of the parent DFT/MRCI method, which is termed DFT/MRCI(2) [16] and drastically reduces the computational cost while maintaining the accuracy of the original method. In this work, we will primarily employ DFT/MRCI(2) and expect the results and conclusions to be equally applicable to DFT/MRCI.

However, since the reference space is generated through a selected-CI procedure, it cannot be guaranteed that the DFT/MRCI(2) energies for a set of closely spaced structures form a reliably differentiable PES [15]. To overcome this, we treat the local discontinuities as noise within the GPR framework, and learn the surfaces by explicitly incorporating and optimizing a white noise kernel. The main difference in the application of GPR to DFT/MRCI(2) compared to MR-CIS in the previous chapter is the explicit use and optimization of noise as a smoothing device to an inherent electronic structure problem. As in Chapter 4, the $\omega$-CP model is used to optimize the DFT/MRCI(2) MECI geometries for representative intersection motifs in the molecules ethylene, butadiene, and fulvene

## 5.2 Minimum Energy Conical Intersection Optimization and Characterization

In this section, additional MECI characterization metrics are defined, which is useful to compare the results of two different electronic structure methods. The MECI optimization method employed to obtain the DFT/MRCI(2) MECIs is also introduced.

### 5.2.1 Characterization of two-state Conical Intersection

As mentioned in Section 2.2.1, a two-state ConIns can be fully characterized in a first-order model by the set of parameters $\{s_x, s_y, \tilde{g}, \tilde{h}\}$ (Eq.2.21 and Eq.2.22), which describes the tilt and asymmetry of the cones. While the set of parameters fully characterize the intersection topography, they do not clearly link to the coupled electronic and nuclear dynamics that may result when passing through this region of the PES. Recent work by Lindh *et al.* [50] introduced composite parameters that correlate with potential reactive outcomes after traversing through the ConIns region. The authors defined the following two parameters:

$$\begin{aligned} \mathcal{P} &= \frac{L}{1-\Delta_{gh}^2}\left(1-\Delta_{gh}\cos(2\theta_s)\right) \\ \mathcal{B} &= \sqrt[3]{\frac{L}{4\Delta_{gh}^2}}\left(\sqrt[3]{(1+\Delta_{gh})\cos^2\theta_s}+\sqrt[3]{(1-\Delta_{gh})\sin^2\theta_s}\right), \end{aligned} \tag{5.1}$$

where the intersection asymmetry, $\Delta_{gh}$, and the parameters $L$ and $\theta_s$ are given by:

$$\begin{aligned} \Delta_{gh} &= \frac{\tilde{g}\cdot\tilde{g} - \tilde{h}\cdot\tilde{h}}{\tilde{g}\cdot\tilde{g} + \tilde{h}\cdot\tilde{h}} \\ L &= \sqrt{s_x^2 + s_y^2} \\ \theta_s &= \tan^{-1}\left(\frac{s_y}{s_x}\right). \end{aligned} \tag{5.2}$$

For $\mathcal{P} < (>)\ 1$, the ConIns corresponds to a peaked (sloped) intersection, and for $\mathcal{B} < (>)$ 1, the ConIns has a bifurcating (single-path) intersection. In short, a ConIns is peaked (sloped) if the upper surface is (is not) at a minimum in the branching plane, and it is single-path (bifurcating) if there is only one minimum (multiple minima) on the lower surface in the branching plane.

While the above parameters $\mathcal{P}$ and $\mathcal{B}$ enable the connection of intersection topography to preliminary dynamical outcomes, it does not provide a direct measure for comparing the branching spaces of two different electronic structure methods. A metric to directly compare two branching spaces is by projecting the branching space vectors of one method onto the branching plane of the other [50, 51]. More precisely, let $\{\boldsymbol{x}, \boldsymbol{y}\}$ and $\{\boldsymbol{x}', \boldsymbol{y}'\}$ be the sets of orthonormalized branching space vectors for method $I$ and $J$, respectively, then the area enclosed by the unit branching vectors of method $I$ projected onto the branching plane of method $J$ is given by:

$$r_{IJ} = |(\boldsymbol{x}\cdot\boldsymbol{x}')(\boldsymbol{y}\cdot\boldsymbol{y}') - (\boldsymbol{x}\cdot\boldsymbol{y}')(\boldsymbol{y}\cdot\boldsymbol{x}')| \tag{5.3}$$

The value $r_{IJ}$ is bounded between 0 for orthogonal branching planes and 1 for parallel

branching planes.

Together, the MECI can be characterized with the parameters $\tilde{g}$, $\tilde{h}$, $s_x$, $s_y$, $\mathcal{P}$ and $\mathcal{B}$ defined in Eqs. 2.21 and 5.1, and the branching spaces of different methods can be compared with $r_{IJ}$ defined in Eq. 5.3, which we use to compare the optimized MECIs of DFT/MRCI(2) to those of *ab initio* MRCI.

### 5.2.2 Minimum Energy Conical Intersection Optimization

Since this work focuses on the GPR surrogate representation of DFT/MRCI(2) seams of ConInss and the subsequent characterization of MECIs, we start from geometries close to the MECI and build a static surface on which a ConIns optimization is performed. We thus remove the requirement for active learning; namely the optimal approach for updating the surrogate and hyperparameters during the course of the MECI optimization. The employment of an active learning process will be important for the practical deployment of the approach for the optimization of structural minima, but would add additional complexity to the present proof-of-principles study and is instead reserved for future work.

There are multiple approaches for optimizing MECIs [52–55], but most, including the techniques employed here, require the identification and utilization of the BS vectors, which is defined by the energy difference gradient $\boldsymbol{g}$ and the derivative coupling vector $\boldsymbol{h}$ (Eq.2.18)

The direct optimization [52] approach, combined with the updated branching space method [56] is employed for the MECI optimizations here. The direct optimization ap-

proach in essence turns a constrained optimization problem into an unconstrained one by utilizing the *a priori* information on the constrained subspaces, which is the branching and seam space. The minimizing gradient $\boldsymbol{d}$ in the direct optimization method is given by

$$\boldsymbol{d} = c_1 \left( (1 - c_2) P \nabla \omega + 2 c_2 \Delta E \frac{\nabla(\Delta E)}{||\nabla(\Delta E)||} \right), \tag{5.4}$$

where $\omega(\boldsymbol{R})$ is the average energy, and $\Delta E$ is the adiabatic energy difference. We find that the parameters used in Ref. [57] worked well, namely $c_1 = 0.2$ and $c_2 = 0.9$, and in this work no Hessian is incorporated in the MECI optimization. $P$ is the projector onto the seam space,

$$P = \mathbf{1} - \boldsymbol{x} \otimes \boldsymbol{x} - \boldsymbol{y} \otimes \boldsymbol{y}, \tag{5.5}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the unit BS vectors defined in Eq.2.20 and $\otimes$ refers to the outer product.

In the updated branching space method the derivative coupling unit vector $\boldsymbol{y}$ is not explicitly computed, but rather updated iteratively using the following recursion relation,

$$\boldsymbol{y}_{k+1} = \frac{\langle \boldsymbol{y}_k, \boldsymbol{x}_{k+1} \rangle \boldsymbol{x}_k - \langle \boldsymbol{x}_k, \boldsymbol{x}_{k+1} \rangle \boldsymbol{y}_k}{\sqrt{\langle \boldsymbol{y}_k, \boldsymbol{x}_{k+1} \rangle^2 + \langle \boldsymbol{x}_k, \boldsymbol{x}_{k+1} \rangle^2}} \tag{5.6}$$

where $<,>$ refers to an inner product, and the initial guess used in this work is given by $\boldsymbol{y}_0 = \nabla\omega$. For clarity, since the adiabatic energies have discontinuous derivatives along seams of ConInss, the $\omega$ and $\Delta E$ in Eq.5.4 are obtained by learning the $\omega$-CP model.

## 5.3 Computational Details

### 5.3.1 Quantum Chemistry Calculations

The DFT/MRCI(2) method was employed to compute single point electronic energies in the vicinity of seams of conical intersection. Previous work has shown the (approximate) DFT/MRCI(2) approach to furnish excitation energies in near-quantitative agreement with the parent DFT/MRCI method. [17, 18] The density fitted [58, 59] cc-pVDZ basis set was used in all cases as implemented in the PySCF package, and 3 roots were computed for all the molecules studied. The initial, guess reference space was generated using the previously discussed AutoRAS procedure. [60] The General Reference Configuration Interaction (GRaCI) package was used for all DFT/MRCI(2) computations [61], where the DFT portion of the calculation was computed using PySCF [62–64].

For comparison, the MECIs and corresponding branching spaces were also determined using *ab initio* MRCI methods. MR-CI singles (MR-CIS) and MR-CI singles and doubles (MR-CISD) MECIs were optimized, employing state-averaged complete active space self-consistent field (SA-CASSCF) reference wave functions, where $m$-electron, $n$-orbital active spaces, and number of roots in the averaging procedure, $S$, are specified $S$-($m$, $n$) and vary by molecule. For ethylene, butadiene, and fulvene, these correspond to 2-(2,2), 2-(4,4), and 2-(6,6), respectively. The COLUMBUS electronic structure package was used for all *ab initio* MRCI computations. [43].

### 5.3.2 Sampling

For the DFT/MRCI(2) MECI optimizations, we consider a distribution of the parameters $\tilde{g}$, $\tilde{h}$, $s_x$, $s_y$, $\mathcal{P}$, $\mathcal{B}$ and $r_{IJ}$ computed over a set of 50 surrogates, each constructed with different LHS centered at the same geometry. The reason for this is we found that using atom-centered Cartesian LHS to generate the training data can result in deviations between different learned surrogate surfaces. This can then manifest itself in slightly different optimized MECI geometries and branching space topographies. This dependence on the training data is expected due to GPR being a non-parametric method. In contrast to parametric fitting, GPR does not assume a functional form, but is based solely on the correlation between each training point, given by the choice of kernel, thus its sensitivity to the training set. We have found this to be a more general problem, and believe LHS generated nuclear configurations are not ideally suited to general, on-the-fly construction of PESs when the kernel hyperparameters are obtained from MLE, in terms of both consistency and scalability. In addition to the curse of dimensionality from which LHS suffers [65], it has been shown that GPR hyperparameters optimized with marginalized log-likelihood from data generated with LHS can perform poorly due to irregular pairwise distance distribution [44, 66]. The question of optimal sampling and fitting, however, is not addressed in this work and we stick to atom-centered Cartesian LHS. To confirm the observed sensitivity does not instead originate from the underlying DFT/MRCI(2) calculations, however, the same procedure of constructing multiple different surrogates was employed using *ab initio* MRCI energies, and the same sensitivity of the parameters on the

training data was found. This will be discussed in more Section 5.6.

### 5.3.3 Descriptor and Kernels

In this chapter, the SOAP parameters were not explicitly optimized. A default set of SOAP parameters were found to work well and subsequently used for all calculations. Namely, $\{r = 4, n = 6, l = 6, \sigma = 0.1\}$. The original SOAP work [19] showed that $l_{max} = 6$ is able to adequately reconstruct their test structures. Using this as a starting point, the $n_{max}$ term is matched to the same value, and the radial cutoff $r$ is picked from the rounded average of the GA optimized values from Chapter 4. We found this set of value able to adequately reproduce the surfaces in the vicinity of the MECI for the molecules studied in the previous chapter, i.e. at the MRCI level of theory without including the whitenoise, and is thus used here.

A simple isotropic SE kernel combined whitenoise kernel is used to to model the discontinuties as i.i.d. homoscedastic noise.

$$\begin{aligned} k(x_i, x_j) &= k_{RBF} + k_{white} \\ k_{RBF}(x_i, x_j) &= \sigma^2 \exp - \frac{||X_i - X_j||^2}{2l^2} \\ k_{white} &= \sigma_n^2 \delta_{ij} \end{aligned} \tag{5.7}$$

As noise is explicitly considered here, the kernel hyperparameters $\{\sigma^2, l, \sigma_n^2\}$ are optimized via the maximization of the marginal log-likelihood of Eq 3.19.

## 5.4 Minimum Energy Conical Intersections of DFT/MRCI(2)

As mentioned in the Introduction, the nature of DFT/MRCI(2) renders the surfaces locally non-smooth. More specifically, the DFT/MRCI(2) PESs resemble underlying smooth surfaces contaminated by deterministic noise that arises from the native selection procedure of the method. We have chosen to approximate the local discontinuous variations of the DFT/MRCI(2) potentials as independent and identically distributed (i.i.d.) Gaussian noise within the GPR framework, which has a simple and closed form method of determination via the optimization of a whitenoise kernel. It is important at this point to reinforce the point that DFT/MRCI(2) PESs exhibit deterministic noise, but treating it as a i.i.d. Gaussian noise within the GPR framework leads to significant formal simplifications. GPR-DFT/MRCI(2) herein refers to the GPR surrogate fitted using DFT/MRCI(2) single point energies, and is used to make clear the distinction between the surrogate DFT/MRCI(2) and the actual DFT/MRCI(2) quantities. Additionally, the learned GPR surrogate of a function $f(x)$ is denoted with an overbar, as $\overline{f}(x)$.

In this section, the MECIs obtained via optimization on GPR-DFT/MRCI(2) is presented and compared to the MECIs computed using analytical *ab initio* MRCI gradients and non-adiabatic couplings. The representative examples discussed below have been previously studied in detail: the accidental $S_0/S_1$ ConIns of "twisted-pyramidalized" ethylene, [67–69] "transoid" butadiene, [70, 71] and "twisted" fulvene, [72].

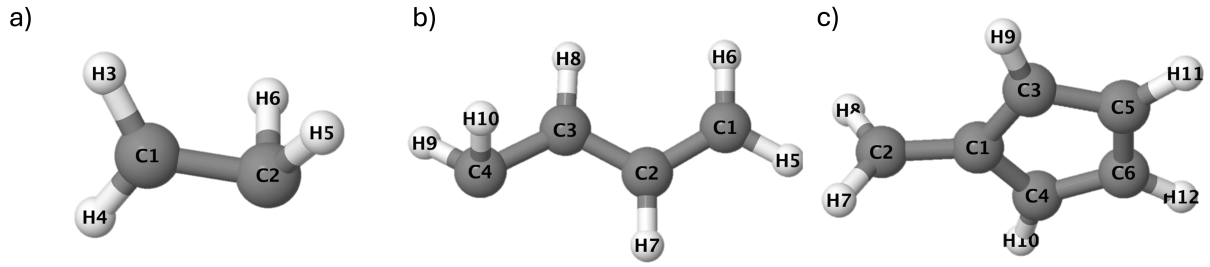


Figure 5.1: MECI geometries optimized using GPR surrogates of DFT/MRCI(2) potentials for a) ethylene, b) butadience and c) fulvene. The atom numbers are also labeled, which is used to define the internal coordinates of interest.

A comparison of the key internal coordinates[1] between the mean DFT/MRCI(2) structures, computed from the optimized MECI over the 50 surrogates constructed outlined Section 5.3.2, and those obtained using *ab initio* MRCI is shown in Table 5.1. The nuclear configurations for each of the structures used to define the internal coordinates are shown in Fig. 5.1, where the atom numbering employed will be used throughout the subsequent discussion.

The optimized GPR-DFT/MRCI(2) and *ab initio* MRCI MECI intersection topography parameters $\{s_x, s_y, \tilde{g}, \tilde{h}, \mathbf{P}, \mathcal{B}\}$ for ethylene, butadiene and fulvene is summarized in Table 5.2, and Table 5.3 shows the overlap of the branching spaces $r_{IJ}$ for each of the methods. For rows and columns corresponding to GPR-DFT/MRCI(2), the presented value is the arithmetic mean $\pm$ the standard deviation over the values obtained from the set of 50 surrogates.

Turning first to the comparison of the structural parameters, we find that the DFT/MRCI(2)

[1]Internal coordinates here refer to general coordinates that is translational and rotational invariant.

| | | $r(C_1C_2)$ | $\theta(H_4C_1H_3)$ | $\theta(H_5C_2H_6)$ | $\gamma(C_1H_5H_6C_2)$ | |
|---|---|---|---|---|---|---|
| Ethylene | DFT/MRCI(2) | 1.43 ±0.01 | 114.2 ±1.424 | 93.70 ±1.08 | 62.76 ±1.124 | |
| | MR-CIS | 1.42 | 113.7 | 93.00 | 59.97 | |
| | MR-CISD | 1.41 | 112.4 | 94.3 | 61.39 | |
| | | $r(C_1C_2)$ | $r(C_2C_3)$ | $r(C_3C_4)$ | $\theta(H_5C_1H_6)$ | $\gamma(C_3H_9H_{10}C_4)$ |
| Butadiene | DFT/MRCI(2) | 1.37 ±0.002 | 1.40 ±0.002 | 1.44 ±0.004 | 117.63 ±0.19 | 62.77 ±0.26 |
| | MR-CIS | 1.33 | 1.45 | 1.43 | 116.84 | 65.00 |
| | MR-CISD | 1.35 | 1.45 | 1.42 | 117.11 | 65.74 |
| | | $r(C_1C_2)$ | $r(C_1C_3)$ | $r(C_3C_5)$ | $\theta(H_7C_2H_8)$ | $d(H_7C_2C_1C_3)$ |
| Fulvene | DFT/MRCI(2) | 1.45 ±0.009 | 1.41 ±0.007 | 1.41 ±0.004 | 116.5 ±1.269 | 89.99 ±0.007 |
| | MR-CIS | 1.48 | 1.42 | 1.43 | 118.2 | 89.99 |
| | MR-CISD | 1.48 | 1.43 | 1.43 | 118.05 | 90.03 |

Table 5.1: Comparison of optimized internal coordinates for the $S_0/S_1$ MECI of ethylene, butadiene and fulvene. $r(xy)$ corresponds to a bond stretch between atoms $xy$ given in Angstroms, $\theta(xyz)$ corresponds to a bend angle between atoms $xyz$ and $\gamma(xyzw)$ corresponds to the out of plane angle of atom $w$, and $d(xyzw)$ corresponds to the dihedral (torsion) angle of atoms $xyzw$. The stretch coordinate $r$ is given in Angstroms, and the $\theta$, $\gamma$ and $d$ angles are given in degrees.

| | | $\tilde{g}$ | $\tilde{h}$ | $s_x$ | $s_y$ | $\mathcal{P}$ | $\mathcal{B}$ |
|---|---|---|---|---|---|---|---|
| Ethylene | DFT/MRCI(2) | 0.209 ±0.002 | 0.181 ±0.004 | 0.056 ±0.011 | 0.019 ±0.005 | 0.087 ±0.026 | 1.526 ±0.198 |
| | MR-CIS | 0.213 | 0.125 | 0.130 | 0.046 | 0.507 | 1.289 |
| | MR-CISD | 0.204 | 0.129 | 0.102 | 0.029 | 0.303 | 1.174 |
| Butadiene | DFT/MRCI(2) | 0.224 ±0.002 | 0.129 ±0.001 | 0.032 ±0.007 | 0.032 ±0.001 | 0.082 ±0.015 | 0.602 ±0.062 |
| | MR-CIS | 0.207 | 0.110 | 0.133 | 0.023 | 0.452 | 1.114 |
| | MR-CISD | 0.183 | 0.109 | 0.123 | 0.023 | 0.492 | 1.268 |
| Fulvene | DFT/MRCI(2) | 0.187 ±0.056 | 0.172 ±0.005 | 0.036 ±0.032 | 4.90e-4 ±3.2e-3 | 0.052 ±0.116 | 1.902 ±0.742 |
| | MR-CIS | 0.187 | 0.180 | 0.047 | 5.89e-5 | 0.063 | 2.414 |
| | MR-CISD | 0.202 | 0.199 | 0.032 | 1.74e-3 | 0.025 | 3.637 |

Table 5.2: Comparison of DFT/MRCI(2), MR-CIS and MR-CISD derived ConIns topographies at the respective MECI geometries.

[2]

MECIs are generally in good agreement with the structures optimized using *ab initio* MRCI methods, as shown by the similarity in the internal coordinates. Deviations in bond lengths and bond angles are typically within 0.05 Angstroms and 3 degrees, respectively. Since they are not the same level of theory, this deviation is well within reasonable agreement.

The good agreement between the GPR-DFT/MRCI(2) and *ab initio* MRCI structural parameters is further mirrored in the comparison of the branching space parameters summarized in Table 5.2, but with some caveats. The norms of the BS vectors, $\tilde{g}$ and $\tilde{h}$, are agreeable between all methods and levels of theory except for ethylene, which exhibits a slight difference in the norm $\tilde{h}$. The deviation, however, does not give rise to a qualitatively different ConIns region, as both methods furnish an asymmetric double cone. However, there appears to be a consistent difference between the GPR-DFT/MRCI(2) and MRCI

| | $r_{IJ}$ | DFT/MRCI(2) | MR-CIS | MR-CISD |
|---|---|---|---|---|
| Ethylene | DFT/MRCI(2) | 1.000 | 0.950 | 0.988 |
| | MR-CIS | 0.950 | 1.000 | 0.978 |
| | MR-CISD | 0.988 | 0.978 | 1.000 |
| Butadiene | DFT/MRCI(2) | 1.000 | 0.975 | 0.962 |
| | MR-CIS | 0.975 | 1.000 | 0.981 |
| | MR-CISD | 0.962 | 0.981 | 1.000 |
| Fulvene | DFT/MRCI(2) | 1.000 | 0.975 | 0.976 |
| | MR-CIS | 0.975 | 1.000 | 0.999 |
| | MR-CISD | 0.976 | 0.999 | 1.000 |

Table 5.3: Comparison of DFT/MRCI(2), MR-CIS and MR-CISD branching spaces at the respective MECI geometries.

[3]

tilt parameters $s_x$ and $s_y$. Specifically, these parameters are consistently smaller for the GPR-DFT/MRCI(2) branching spaces in comparison to the *ab initio* MRCI results for the examples considered here, resulting in the former yielding intersections with vertical non-sloped cones. A more detailed analysis of this observation will be presented in Section 5.5.

While the absolute magnitude of the tilt parameters may exhibit large differences, we see that if the ratio of $s_x/s_y$ is similar, then the composite parameters $\mathcal{P}$ and $\mathcal{B}$, defined in Eq. 5.1, may be similar. For example, although the DFT/MRCI(2) tilt parameters for ethylene are significantly smaller than their *ab initio* MRCI counterparts, the predicted $\mathcal{P}$ and $\mathcal{B}$ values fall under the same category. That is, they both predict peaked and single-path intersections. On the other hand, for butadiene, not only are the GPR-DFT/MRCI(2) tilt parameters a lot smaller, the ratio is also significantly different. As a result, GPR-DFT/MRCI(2) predicts a bifurcating intersection, whereas MRCI predicts a single-path

intersection.

However, these differences in intersection topography, as discussed above, do not result in significant differences in the corresponding branching spaces. As shown in Table 5.3, the GPR-DFT/MRCI(2) and *ab initio* MRCI branching planes are all nearly-parallel ($r_{IJ}$ clost to 1). For comparison, a previous study by Filatov *et al.* and Lindh *et al.* compared the branching spaces resulting from different electronic structure methods to MR-CISD for a range of molecules. [50, 51] The methods examined included SA-CASSCF, ensemble DFT approaches [73, 74], and the orthogonalization-corrected (OM2) semi-empirical configuration interaction method [75, 76]. As that study demonstrated, the branching spaces furnished by different electronic structure approaches can vary widely, with, for example, spin-flip TDDFT (SF-TDDFT) yielding $r_{IJ}$ overlaps with the benchmark MRCI ranging from 0.64 for a twisted-bond-alternating methylimine MECI to near quantitative agreement and a 0.986 overlap for methylimine MECI. Thus, the average overlap between the GPR-DFT/MRCI(2) and MR-CISD branching spaces of 0.98 for the examples considered should be considered near quantitative agreement.

Finally, to verify that the GPR-DFT/MRCI(2) potentials accurately reproduce the DFT/MRCI(2) potentials, Fig. 5.2 displays the GPR-DFT/MRCI(2) energies calculated along the BS directions, obtained from the non-zero $c_0^z$ Hessian eigenvectors, alongside the DFT/MRCI(2) single-point energies, with the origin set at the MECI geometry. Fig. 5.2 shows that the GPR surrogates reproduce well the DFT/MRCI(2) energies in the vicinity of the MECI along both branching space directions.

At this point it is important to discuss a challenge that arises when learning the $\overline{c_0^z}(\boldsymbol{R})$ surface. In the direct MECI optimization approach used here, the minimizing gradient (Eq.5.4) has two components, which can be thought as performing two operations. The first minimizes the average energy projected onto the seam space, and the second minimizes a quantity proportional to $\Delta E^2$, with its minimizing gradient as $2\Delta E(\nabla \hat{\boldsymbol{x}})$. For the $\omega-$CP model, $c_0^z(\boldsymbol{R})$ is a strictly negative and real quantity for any $\boldsymbol{R}$. Problems will arise whenever the solution gets near the degeneracy and the $\overline{c_0^z}(\boldsymbol{R})$ surrogate returns a positive, albeit small value, which renders the predicted $\Delta E$, and subsequently the gradient component responsible for minimizing the degeneracy, to be zero, since we only take the real parts of the gradient. Thus the degeneracy cannot be made arbitrarily small with increased iterations. To ensure that the branching space cuts in the vicinity of the origin do not exhibit spurious cusps due to positive $\overline{c_0^z}(\boldsymbol{R})$ values, all the $\overline{c_0^z}(\boldsymbol{R})$ values were shifted with $\overline{c_0^z}(\boldsymbol{R}_{MECI})$ in the branching space cut plots (Fig. 5.2), which also forces the origin and $\overline{\Delta E}(\boldsymbol{R}_{MECI})$ to be zero.

Additionally, Fig. 5.3 shows the eigenvalues of the $c_0^z$ Hessian, with the number of non-zero (zero) eigenvalues corresponding to the branching (seam) space dimension. The first point of note is that the eigenvalues of the seam space coordinates should be strictly zero. For GPR-DFT/MRCI(2), the seam space eigenvalues show small deviations from zero, but is within the uncertainty of the GPR surrogate model, which will be discussed in Section 5.6. In practice, Fig. 5.3 shows that GPR-DFT/MRCI(2) is able to correctly capture the branching and seam space dimensions.

Summarizing these observations, we find that the optimized MECI structures, as well as

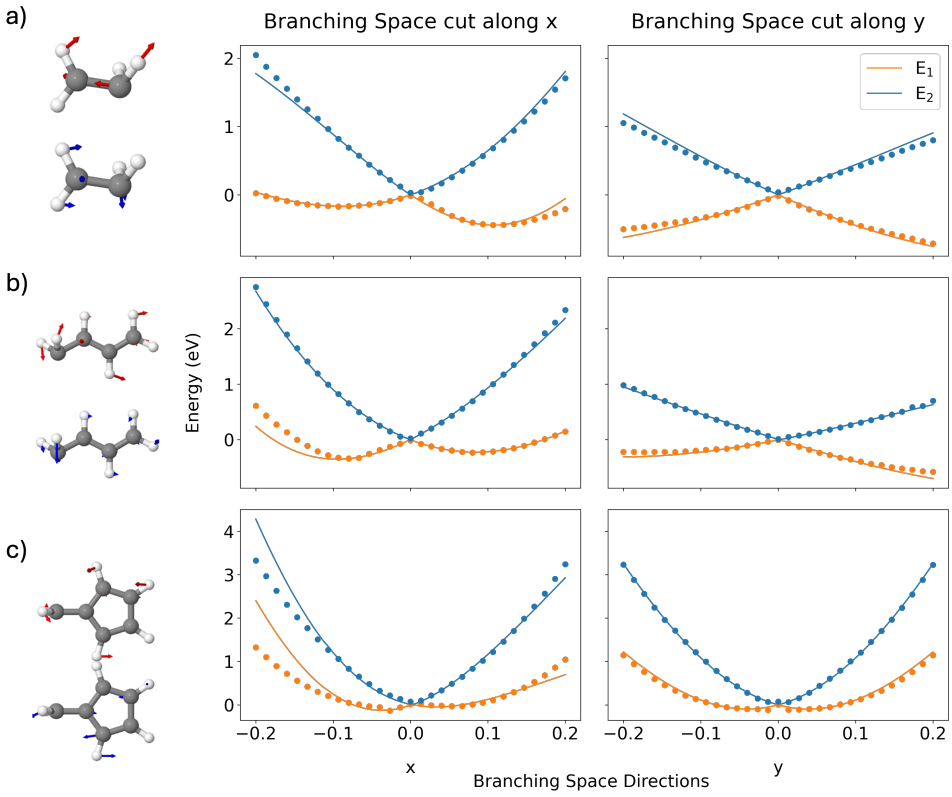


Figure 5.2: Adiabatic energies in the vicinity of the MECI along the branching vectors for: a) Ethylene, b) Butadiene, c) fulvene. Left panel shows the optimized MECIs and their corresponding x (red) and y (blue) branching vectors. The middle and right panel shows the Adiabatic energies along the x and y branching vectors, where the dotted points correspond to the DFT/MRCI(2) energies and the solid lines correspond to the GPR predicted energies.

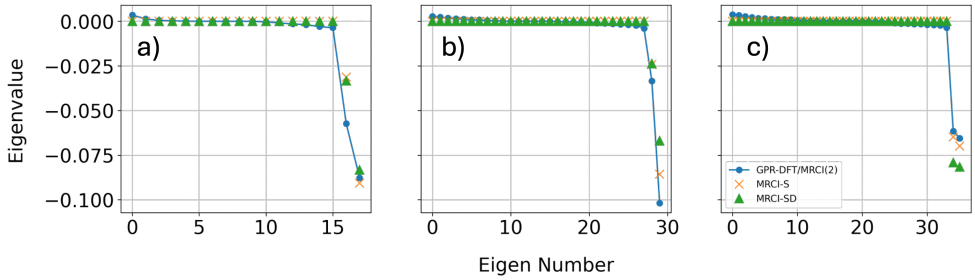


Figure 5.3: Eigenvalues of the $c_0^z$ Hessian for: a) Ethylene, b) Butadiene, c) Fulvene. The dotted values are the GPR predicted values for DFT/MRCI(2), and the cross and triangle points are the *ab initio* MR-CIS and MR-CISD values, respectively. The number of non-zero (zero) elements correspond to the branching (seam) space dimension.

the branching spaces of GPR-DFT/MRCI(2) and *ab initio* MRCI are in near-quantitative agreement. The ConIns topography, however, exhibits systematic differences from *ab initio* MRCI approaches in the form of smaller GPR-DFT/MRCI(2) tilt parameters $s_x$ and $s_y$.

## 5.5 Comparison of DFT/MRCI(2) and *ab initio* MRCI Surrogate Surfaces

As shown in Section 5.4, the main qualitative difference between the GPR-DFT/MRCI(2) and *ab initio* MRCI MECI topographies is the tilt of the double cones. Here, we explore this finding in more detail and confirm that these deviations represent real differences in how the electronic structure methods describe the seam of intersection, rather than artifacts of the surrogate generation process.

To this end, GPR surrogates for the ethylene MECI is constructed using MR-CIS single point energies, and the resulting optimized kernel hyperparameters for the $\tilde{\omega}$ and $\tilde{c_0^z}$ GPR surrogates between MR-CIS and DFT/MRCI(2) is compared. These values are given in Table 5.4. The difference in the tilt parameters can then be directly related to the learned hyperparameters. It can be seen that apart from the noise term, the variance $\sigma^2$ and lengthscale $l$ of $\overline{c_0^z}(\boldsymbol{R})$ is similar between DFT/MRCI(2) and MR-CIS. The straight-forward interpretation of these results is that the two learned surfaces are similar in terms of how large and how rapidly the $\overline{c_0^z}(\boldsymbol{R})$ surface is varying around a MECI, which are reflected in the lengthscale and variance, respectively. The GPR-DFT/MRCI(2) surrogates have a larger noise term to account for the local energetic variations. The two $\overline{\omega}(\boldsymbol{R})$ surfaces, however, are different, with the GPR-DFT/MRCI(2) having appreciably smaller $\sigma^2$ and $l$ hyperparameters. Examining Eq. 2.21 and taking the gradient of $\overline{\omega}(\boldsymbol{R})$, we can see that the tilt parameters are proportional to $\sigma^2$ and inversely proportional to $l$,

$$\nabla\overline{\omega}(X_*) = \sum_i^N \alpha_i \exp\left(-\frac{||X_i - X_*||^2}{2l^2}\right)(X_i - X_*)\frac{\sigma^2}{l^2} \tag{5.8}$$

Additionally, from the computed $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{h}}$ norms as well as the large overlap between the DFT/MRCI(2) and MR-CIS branching spaces, we can exclude differences in the branching vectors as the reason for the difference in the tilt parameters. That is, the norms of $\boldsymbol{x}$ and $\boldsymbol{y}$ in Eq. 2.21 is not the dominant term that results in smaller values of $s_x$ and $s_y$ for DFT/MRCI(2) compared to MR-CIS. The significantly smaller value of $\sigma^2$ for DFT/MRCI(2) is thus found to be responsible for giving rise to smaller values of $s_x$ and $s_y$.

There is a simple rationalization of the difference between the DFT/MRCI(2) and MR-CIS $\overline{c_0^z}(\boldsymbol{R})$ surfaces, which differ only in the noise term. Namely, that the MR-CIS potential surfaces do not suffer from the presence of deterministic noise that is inherent to the DFT/MRCI(2) method. However, the same conclusion cannot be drawn for the differences in $\overline{\omega}(\boldsymbol{R})$. Since we cannot provide physically motivated reasons for the difference in the log-likelihood landscape that arise due to different electronic structure methods, the MLE optimized noise term does not offer insight into the physics of the DFT/MRCI(2) local energetic variation. When the set of optimized hyperparameters $\{\sigma^2, l, \sigma_n^2\}$ differ significantly, they may correspond to extrema of drastically different log-likelihood landscapes, thus frustrating a simple interpretation. Furthermore, it is also important to remember that the whitenoise kernel hyperparameter $\sigma_n^2$ is not an absolute measure of noise. Depending on the kernel composition, the log-likelihood surface can differ, even employing the same

| | | DFT/MRCI(2) | MR-CIS |
|---|---|---|---|
| $c_0^z$ | $\sigma^2$ | $316^2 \pm 0.0$ | $316^2 \pm 0.0$ |
| | $l$ | $0.095 \pm 0.002$ | $0.111 \pm 0.001$ |
| | $\sigma_n^2$ | $0.011 \pm 0.001$ | $2.52 \times 10^{-6} \pm 5.36 \times 10^{-7}$ |
| $\omega$ | $\sigma^2$ | $40^2 \pm 20^2$ | $277^2 \pm 98^2$ |
| | $l$ | $0.088 \pm 0.007$ | $0.323 \pm 0.010$ |
| | $\sigma_n^2$ | $0.0023 \pm 0.0001$ | $3.27 \times 10^{-8} \pm 4.54 \times 10^{-9}$ |

Table 5.4: MLE optimized GPR hyperparameters of ethylene for DFT/MRCI(2) and MR-CIS. The first and second row corresponds to the learned hyperparameters for the $\overline{c}_0^z(\boldsymbol{R})$ and $\overline{\omega}(\boldsymbol{R})$ surface, respectively. The values presented is the arithmetic mean $\pm$ standard deviation over 50 surrogates.

training data, thus using different kernels changes the learned $\sigma_n^2$ value significantly. More importantly, different metric spaces or cost functions used in the training can result in drastically different sets of optimized hyperparameters, and subsequently change the role and interpretation of the noise term in the model. For the analysis performed here, we assume that the PESs in the vicinity of the MECI computed using DFT/MRCI(2) and *ab initio* MR-CIS do not have significant qualitative differences, which would otherwise make the use of the same kernel inappropriate. Therefore, although the i.i.d. noise model provides a simple and intuitive picture, it should only be interpreted as a smoothing device in the construction of PESs. The main consequence from all the above is, the DFT/MRCI(2) $\omega(\boldsymbol{R})$ surface is responsible for the smaller tilt parameters, but we cannot conclude from the hyperparameters alone that the DFT/MRCI(2) and MR-CIS $\omega(\boldsymbol{R})$ surfaces are similar with added deterministic noise for DFT/MRCI(2), as in the case of the $c_0^z(\boldsymbol{R})$ surfaces, because the optimized hyperparameters are drastically different.

| | DFT/MRCI(2) | MR-CIS |
|---|---|---|
| $\lvert s_x \rvert$ | $\pm 0.011$ | $\pm 0.010$ |
| $\lvert s_y \rvert$ | $\pm 0.005$ | $\pm 0.005$ |
| $\lvert g \rvert$ | $\pm 0.002$ | $\pm 0.002$ |
| $\lvert h \rvert$ | $\pm 0.004$ | $\pm 0.002$ |
| $\mathcal{P}$ | $\pm 0.026$ | $\pm 0.048$ |
| $\mathcal{B}$ | $\pm 0.198$ | $\pm 0.056$ |

Table 5.5: GPR surrogate hyperparameter dependence on LHS for MR-CIS and DFT/MRCI(2). The magnitude of error that arise due to sampling is largely equivalent between the two methods.

## 5.6 Fundamental Uncertainty in DFT/MRCI(2) Derived GPR-Surrogates

Before turning to the main challenges that arise from the present framework, we first address the uncertainty corresponding to the surrogate dependence on different LHS distributions. To study the sampling dependence of the GPR surrogates, 50 surrogates were also constructed at the MR-CIS level of theory by LHS about the MECI geometries. MECI optimization was performed on each of the 50 GPR surrogate surfaces, and the mean and standard deviation of the parameters reported in Table 5.2 were also determined for this dataset. For each MECI optimization, the geometry with the largest $\Delta E$ in the training set was chosen as the starting geometry in order to keep the initial conditions consistent. As expected, the surrogate dependence on LHS is also present in the MR-CIS derived surfaces, and the standard deviations on the branching space parameters is of the same order of magnitude. These results are summarized in Table 5.5.

These results demonstrate that there is a non-negligible sampling dependence to the resulting surrogates, regardless of the electronic structure method employed to generate the electronic energies. This can also be seen more directly from Table 5.4, which shows the distribution of learned hyperparameters over the set of 50 surrogates. This error is clearly always present when fitting surfaces using training data generated from LHS distributions, and although usually small, it is particularly relevant when points of degeneracy are sought.

While this sampling dependence is present in both the MR-CIS and DFT/MRCI(2) derived surrogates, a major difference is the value of the learned noise hyperparameter $\sigma_n^2$: it is significantly larger for DFT/MRCI(2) compared to MR-CIS. Note that since these hyperparameters are learned from data transformed to have zero mean and unit variance, it does not correspond to a physically interpretable value. Nevertheless, by comparing the learned DFT/MRCI(2) and MR-CIS $\sigma_n^2$ values, shown in Table 5.4, it is clear that the learned DFT/MRCI(2) $\omega(\boldsymbol{R})$ and $c_0^z(\boldsymbol{R})$ surface is a lot noisier compared to MR-CIS due to the local energetic variations. This is not unexpected, since each nuclear configuration results in the generation of a new reference space, but also highlights the need for surrogate representations of the nascent DFT/MRCI(2) surfaces in order for them to be meaningfully differentiable.

An important consequence of the larger noise in the DFT/MRCI(2) potentials is that it translates into a limitation on the level of degeneracy that can be achieved. The GPR surrogate returns a mean $\overline{y}(X_*)$ and standard deviation $\delta_{\overline{y}}(X_*)$ by treating the target $y$ as a Gaussian random variable. Within the GPR framework, we can then say that the real value $y(X_*)$ falls between the interval $\overline{y}(X_*) \pm 2\delta_{\overline{y}}(X_*)$ with a 95% confidence. For a

noiseless GPR model, ignoring the presence of a small regularization term, the predicted variance of a training point is zero:

$$\begin{aligned}\delta_f^2(X_*) \equiv \mathbb{V}[f_*] &= K(X_*, X_*) \\ &\quad - K(X_*, X)K(X, X)^{-1}K(X, X_*) \\ &= K_{**} - \sum_j^N \sum_i^N K_{*i} K_{ij}^{-1} K_{j*} \\ &= K_{**} - \sum_j^N \delta_{*j} K_{j*} \\ &= K_{**} - K_{**} = 0,\end{aligned} \tag{5.9}$$

where we have made use of the fact that $K$ is symmetric and that for $X_* \in \mathcal{D}$, where $\mathcal{D}$ is the training set, the sum over $i$ on the second line becomes a Kronecker-$\delta$ function. In a noiseless model, adding a training point collapses the predicted variance (and standard deviation) at that point to zero. Thus a model can be made arbitrarily accurate within a region by adding training points, provided the covariance matrix does not become ill-conditioned. The lower bound on the predicted standard deviation for a noiseless model is therefore zero, such that $\delta_f(X_*) = \delta_f^{min} = 0$ for $X_* \in \mathcal{D}$, where $\delta_f^{min}$ corresponds to a minimum uncertainty given in terms of standard deviation. In our model, however, a non-zero lower bound on the predicted variance exists at any point due to the learned noise term, even if the point is in the training set. That is, Eq. 5.9 does not hold in the presence of noise because the sum over $i$ does not become a Kronecker-$\delta$ function when the inverse takes on the form of $(K_{ij} + \sigma_n^2 \mathbf{1})^{-1}$. As a result, $\delta_f(X_*) \geq \delta_f^{min} > 0$ for $X_* \in \mathcal{D}$.

The presence of noise thus renders the minimum uncertainty non-zero, which means that the degeneracy can only be meaningfully determined up to the lower bound before adding additional points becomes irrelevant. Note that this statement holds on the condition that re-optimizing the GP hyperparameters with additional training points does not decrease the learned $\sigma_n^2$, which we found to be the case for the systems under study. Since the minimum uncertainty is learned in terms of $c_0^z(\boldsymbol{R})$, it is propagated forward to $\Delta E(\boldsymbol{R})$ in the small value limit as [77]

$$\begin{aligned} \delta_{\Delta E} &= \left| \frac{d\Delta E}{dc_0^z} \right| \delta_{c_0^z} \\ &= \left| \frac{1}{\sqrt{-c_0^z}} \right| \delta_{c_0^z}, \end{aligned} \tag{5.10}$$

recalling that $\Delta E(\boldsymbol{R}) = \sqrt{-4c_0^z(\boldsymbol{R})}$. The result of this is illustrated in Fig. 5.4, where the DFT/MRCI(2) $\Delta E$ at the MECI is non-zero, and the degeneracy falls within the 95% confidence interval of the minimum uncertainty. That is,

$$0 \in \left[ \Delta E^{DFT/MRCI(2)} - 2\delta_{\Delta E}^{min}, \Delta E^{DFT/MRCI(2)} + 2\delta_{\Delta E}^{min} \right]. \tag{5.11}$$

For comparison, the MR-CIS minimum standard deviation $\delta_{\Delta E}^{min}$ for ethylene was found to be $6 \times 10^{-4}$ eV, which is consistent with the level of convergence employed in the calculation of the electronic energies. A major shift in interpretation resulting from treating the local energetic variations in DFT/MRCI(2) as noise is the role of the single point calculations in the construction of surrogate PESs. It is no longer taken as the ground truth and is only meaningful within an interval governed by the learned noise. A consequence of this

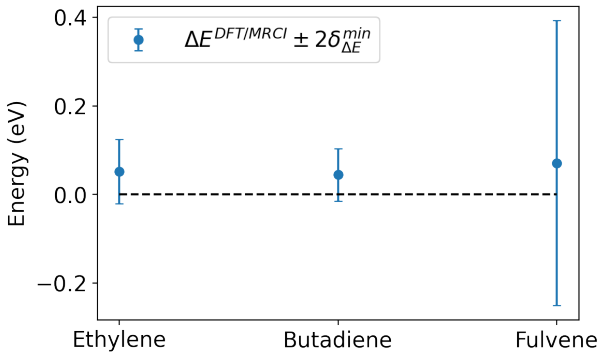


Figure 5.4: The effect of noise as a learned hyperparameter on the achievable accuracy of the optimized DFT/MRCI(2) MECI degeneracy. The blue dots shows the DFT/MRCI(2) energies at the optimized MECI, and the error bar shows the 95% confidence interval as a result of the learned noise. The true degeneracy (dashed line) is contained in the 95% confidence interval for all the molecules

is that the degeneracy may not exist simultaneously in the DFT/MRCI(2) PESs and the GPR surrogate, and if it does, it may not correspond to the same set of geometries unless the noise approaches zero.

In summary, there are two main challenges present in the GPR-DFT/MRCI(2) MECI optimization. The first is quantifiable by examining the minimum uncertainty of the degeneracy $\delta^{min}_{\Delta E}$, and is to be expected since we are essentially smoothing a noisy surface. The second challenge, maintaining a physical sign on the $c^z_0(\boldsymbol{R})$ surface, is related to the machine learning architecture employed. Whether the physical constraints of a positive $c^z_0(\boldsymbol{R})$ surface can be learned given enough data or requires a built-in constraint is an open question. There has been recent work addressing this issue, including designing neural network architectures [78] and employing different companion matrices [79]. In the case of GPR, this issue is intimately related to the Gaussian noise model used, where the noise

is symmetrical about the mean with infinite support and cannot guarantee a non-positive value. This issue can be naively remedied in a noiseless setting by adding training points where nonphysical $c_0^z$ is predicted, or more generally and robustly, even in the noised setting, by considering bounded GPR [80], which includes the use of warped-GPR, non-Gaussian likelihood or constrained kernel hyperparameter optimization, and may be worth exploring in future work.

## 5.7 Conclusion

In this work, we have evaluated the ability of DFT/MRCI(2) to describe seams of conical intersections using a GPR surrogate approach. To obviate the need to learn adiabatic surfaces, which exhibit cusps at points of degeneracy, we instead learned the coordinate dependant coefficients of the characteristic polynomial of the potential matrix, which are smooth functions of nuclear coordinates. We found that the explicit inclusion and optimization of a whitenoise kernel allows for an intuitive framework to learn smooth surrogate DFT/MRCI(2) surfaces that are otherwise discontinuous due to the selected CI aspects of the method. Using the GPR surrogate characteristic polynomial coefficients, we optimized the MECIs of ethylene, butadiene and fulvene and computed for the first time the corresponding branching space and topography using DFT/MRCI(2). The asymmetry of the double cones at the optimized MECIs agree well with *ab initio* MRCI, but the tilt parameters are smaller in magnitude, indicating that DFT/MRCI(2) preferentially gives vertical double-cones compared to *ab initio* MRCI for each of the molecules studied. The difference in the optimized hyperparameters of the average energy surrogates, combined with the good agreement between the branching spaces, shows that the difference in the tilt parameters is a result of the learned average energy surface.

The presence of deterministic noise in the DFT/MRCI(2) potentials was dealt with approximately via the inclusion of white noise in the GPR kernel. This leads to a fundamental change in interpretation of the optimised points of (near) degeneracy, and the

question of the true existence of conical intersections at the DFT/MRCI(2) level of theory. As a result of the noise term, the surrogate surfaces cannot be made arbitrarily accurate with the addition of more data points, thus affecting the level of degeneracy that can be obtained. Despite the issue with explicitly incorporating noise, however, we have shown that the GPR learning of characteristic polynomial coefficients to account for the deterministic noise present in the DFT/MRCI(2) potentials is a viable method that returns sensible surfaces that optimizes easily with direct optimization methods, and yields MECI geometries in excellent agreement with their *ab initio* MRCI counterparts.

# 6

# Conclusion

In this thesis, we have extended and applied a previously proposed method for the indirect learning of coupled potential energy surfaces via the characteristic polynomial coefficient of the potential matrix, which we term the $\omega$-CP model. Specifically, regions of conical intersections was studied as the non-differentiability renders the direct learning of adiabatic energies challenging for conventional ML methods.

In Chapter 4, the $\omega$-CP model was first validated at the *ab intio* MRCI level of theory for three different ConIns: the accidental $S_0/S_1$ ConIns of twisted-pyramidalized ethylene, the symmetry required $(E \otimes e)$ two-state ConIns of $NH_3^+$, and the symmetry-required $(T_2 \otimes (e \oplus t_2))$ three-state intersection of $CH_4^+$. By using a SOAP descriptor and a simple GPR framework, it was found for all three ConIns that the $\omega$-CP model is able to quantitatively reproduce the branching and seam spaces, as judged by the accuracy of the predicted intersection topographies as well as the adiabatic energies. The $\omega$-CP model also exhibited high extrapolation fidelity in regions far away from the training set due to the slowly varying nature of the $\omega$-CP surfaces. Additionally, it was found that all the branching space information is contained in a single CP coefficient.

In Chapter 5, the $\omega$-CP model was extended to the DFT/MRCI(2) method. Contrary to *ab initio* MRCI, DFT/MRCI(2) lacked analytical or reliable numerical solution to compare against due to the locally non-smooth PESs. Aside from being able to effectively learn the ConIns, GPR offered a natural mechanism to smoothen the locally non-smooth PESs via the inclusion and optimization of a whitenoise kernel. Using the $\omega$-CP model, the accidental $S_0/S_1$ MECI of "twisted-pyramidalized" ethylene, "transoid" butadiene, and "twisted" fulvene were optimized and characterized. It was found that the optimized DFT/MRCI(2) MECI structures are very similar to *ab initio* MRCI, while some of the intersection topography differed. Namely, the asymmetry of the cones were similar, but the tilt parameters were consistently smaller for DFT/MRCI(2). Despite the differences in the intersection topography, the branching spaces between DFT/MRCI(2) and MRCI were found to be highly similar as judged by their overlaps. This analysis paved way for

the first quantitative ConIns analysis for DFT/MRCI, and showed the validity of the $\omega$-CP model to fit DFT/MRCI(2) PESs.

# References


1. Szabo, A. & Ostlund, N. S. Modern quantum chemistry : introduction to advanced electronic structure theory eng. ISBN: 0486691861 (Dover Publications, Mineola, N.Y, 1996 - 1989).
2. Domcke, W., Yarkony, D. & Koppel, H. Conical intersections : electronic structure, dynamics spectroscopy eng. ISBN: 9812386726 (World Scientific, River Edge, NJ, 2004).
3. Mead, C. A. & Truhlar, D. G. Conditions for the definition of a strictly diabatic electronic basis for molecular systems. J. Chem. Phys. **77,** 6090–6098. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/77/12/6090/11149561/6090_1_online.pdf (Dec. 1982).
4. Yarkony, D. R. Diabolical conical intersections. Rev. Mod. Phys. **68,** 985–1013 (4 1996).
5. Matsika, S., Cundari, T. R. & Lipkowitz, K. B. eng. in Reviews in Computational Chemistry 83–124 (John Wiley Sons, Inc, Hoboken, NJ, USA, 2007). ISBN: 9780470082010.
6. Atchity, G. J., Xantheas, S. S. & Ruedenberg, K. Potential energy surfaces near intersections. The Journal of Chemical Physics **95,** 1862–1876. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/95/3/1862/18994151/1862_1_online.pdf (Aug. 1991).
7. Domcke, W., Yarkony, D. & Koppel, H. Conical intersections : theory, computation and experiment eng. ISBN: 9789814313445 (World Scientific, Singapore ; 2011).
8. Köppel, H. & Schubert, B. The concept of regularized diabatic states for a general conical intersection. Mol. Phys. **104,** 1069–1079. eprint: https://doi.org/10.1080/00268970500417937 (2006).
9. Kammeraad, J. A. & Zimmerman, P. M. Estimating the Derivative Coupling Vector Using Gradients. J. Phys. Chem. Lett. **7,** 5074–5079. eprint: https://doi.org/10.1021/acs.jpclett.6b02501 (2016).
10. Plasser, F. & Lischka, H. in Quantum Chemistry and Dynamics of Excited States 277–297 (John Wiley Sons, Ltd, 2020). ISBN: 9781119417774. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119417774.ch9.
11. Lindh, R. & Galván, I. F. in Quantum Chemistry and Dynamics of Excited States 299–353 (John Wiley Sons, Ltd, 2020). ISBN: 9781119417774. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119417774.ch10.
12. Li Manni, G., Guther, K., Ma, D. & Dobrautz, W. in Quantum Chemistry and Dynamics of Excited States 133–203 (John Wiley Sons, Ltd, 2020). ISBN: 9781119417774. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119417774.ch6.
13. Parr, R. G. & Weitao, Y. Density-Functional Theory of Atoms and Molecules ISBN: 9780195092769 (Oxford University Press, Jan. 1995).

14. Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. The Journal of Chemical Physic **140,** 18A301. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4869598/19900307/18a301_1_1.4869598.pdf (Apr. 2014).

15. Marian, C. M., Heil, A. & Kleinschmidt, M. The DFT/MRCI method. WIREs Computational Molecular Science **9,** e1394. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1394 (2019).

16. Neville, S. P. & Schuurman, M. S. A perturbative approximation to DFT/MRCI: DFT/MRCI(2). The Journal of Chemical Physics **157,** 164103. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0118285/16551884/164103_1_online.pdf (Oct. 2022).

17. Costain, T. S., Rolston, J. B., Neville, S. P. & Schuurman, M. S. A DFT/MRCI Hamiltonian parameterized using only ab initio data. II. Core-excited states. The Journal of Chemical Physics **161,** 114117. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0227385/20166310/114117_1_5.0227385.pdf (Sept. 2024).

18. Costain, T. S., Ogden, V., Neville, S. P. & Schuurman, M. S. A DFT/MRCI Hamiltonian parameterized using only ab initio data: I. valence excited states. The Journal of Chemical Physics **160,** 224106. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0210897/19985974/224106_1_5.0210897.pdf (June 2024).

19. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. Phys. Rev. B **87,** 184115 (18 2013).

20. Himanen, L. et al. DScribe: Library of descriptors for machine learning in materials science. Comput. Phys. Commun. **247,** 106949. ISSN: 0010-4655 (2020).

21. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. Phys. Rev. Lett. **108,** 058301 (5 2012).

22. Huo, H. & Rupp, M. Unified representation of molecules and crystals for machine learning. Machine Learning: Science **3,** 045017 (2022).

23. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. The Journal of Chemical Physics **134,** 074106. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3553717/15435271/074106_1_online.pdf (Feb. 2011).

24. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. Phys. Chem. Chem. Phys. **18,** 13754–13769 (20 2016).

25. Himanen, L. et al. DScribe: Library of descriptors for machine learning in materials science. Comput. Phys. Commun. **247,** 106949. ISSN: 0010-4655 (2020).

26. Raghunathan, S. & Priyakumar, U. D. Molecular representations for machine learning applications in chemistry. International Journal of Quantum Chemistry **122,** e26870. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.26870 (2022).

27. Kocer, E., Mason, J. K. & Erturk, H. Continuous and optimally complete description of chemical environments using Spherical Bessel descriptors. AIP Advances **10,** 015021. ISSN: 2158-3226. eprint: https://pubs.aip.org/aip/adv/article-pdf/doi/10.1063/1.5111045/12998426/015021_1_online.pdf (Jan. 2020).

28. Barnard, T. et al. Leveraging genetic algorithms to maximise the predictive capabilities of the SOAP descriptor. Mol. Syst. Des. Eng. **8,** 300–315 (3 2023).

29. Laakso, J. et al. Updates to the DScribe Library: New Descriptors and Derivatives 2023. arXiv: 2303.14046 [cond-mat.mtrl-sci].

30. Westermayr, J. & Marquetand, P. Machine Learning for Electronically Excited States of Molecules. Chemical Reviews **121,** 9873–9926. eprint: https://doi.org/10.1021/acs.chemrev.0c00749 (2021).

31. Abedsoltan, A., Belkin, M. & Pandit, P. Toward large kernel models in Proceedings of the 40th International Confere (JMLR.org, Honolulu, Hawaii, USA, 2023).

32. Rasmussen, C. E. & Williams, C. K. I. Gaussian Processes for Machine Learning (Adaptive Computation and Machin ISBN: 026218253X (The MIT Press, 2005).

33. Dral, P. O. Quantum Chemistry in the Age of Machine Learning. J. Phys. Chem. Lett. **11,** 2336–2347. eprint: https://doi.org/10.1021/acs.jpclett.9b03664 (2020).

34. Behler, J. Perspective: Machine learning potentials for atomistic simulations. J. Chem. Phys. **145.** 170901. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4966192/13889426/170901_1_online.pdf (Nov. 2016).

35. Wang, H., Zhang, L., Han, J. & E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. Comput. Phys. Commun. **228,** 178–184. ISSN: 0010-4655 (2018).

36. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. Nat. Commun. **8,** 13890 (2017).

37. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. J. Chem. Phys. **148.** 241722. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5019779/16655678/241722_1_online.pdf (Mar. 2018).

38. Fdez. Galván, I. & Lindh, R. Smooth Things Come in Threes: A Diabatic Surrogate Model for Conical Intersection Optimization. J. Chem. Theory Comput. **19,** 3418–3427. eprint: https://doi.org/10.1021/acs.jctc.3c00389 (2023).

39. Li, C., Hou, S. & Xie, C. Constructing Diabatic Potential Energy Matrices with Neural Networks Based on Adiabatic Energies and Physical Considerations: Toward Quantum Dynamic Accuracy. J. Chem. Theory Comput. **19,** 3063–3079. eprint: https://doi.org/10.1021/acs.jctc.2c01074 (2023).

40. Shu, Y. & Truhlar, D. G. Diabatization by Machine Intelligence. J. Chem. Theory Comput. **16,** 6456–6464. eprint: https://doi.org/10.1021/acs.jctc.0c00623 (2020).

41. Opalka, D. & Domcke, W. Interpolation of multi-sheeted multi-dimensional potential-energy surfaces via a linear optimization procedure. J. Chem. Phys. **138.** 224103. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4808358/13421625/224103\_1\_online.pdf` (June 2013).

42. Gutleb, T. S., Barrett, R., Westermayr, J. & Ortner, C. Parameterizing Intersecting Surfaces via Invariants 2024. arXiv: `2407.03731 [math.NA]`.

43. Lischka, H. et al. COLUMBUS, an ab initio electronic structure program, release 7.0 2012.

44. Gramacy, R. B. Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences `http://bobby.gramacy.com/surrogates/` (Chapman Hall/CRC, Boca Raton, Florida, 2020).

45. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. **12,** 2825–2830 (2011).

46. Deb, K., Sindhya, K. & Okabe, T. Self-adaptive simulated binary crossover for real-parameter optimization in Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (Association for Computing Machinery, London, England, 2007), 1187–1194. ISBN: 9781595936974.

47. Blank, J. & Deb, K. pymoo: Multi-Objective Optimization in Python. IEEE Access **8,** 89497–89509 (2020).

48. Westermayr, J. et al. Machine learning enables long time scale molecular photodynamics simulations. Chem. Sci. **10,** 8100–8107 (35 2019).

49. Westermayr, J. et al. Deep learning study of tyrosine reveals that roaming can lead to photodamage. Nat. Chem. **14,** 914–919. ISSN: 1755-4349 (2022).

50. Fdez. Galván, I., Delcey, M. G., Pedersen, T. B., Aquilante, F. & Lindh, R. Analytical State-Average Complete-Active-Space Self-Consistent Field Nonadiabatic Coupling Vectors: Implementation with Density-Fitted Two-Electron Integrals and Application to Conical Intersections. Journal of Chemical Theory and Cor **12,** 3636–3653. eprint: `https://doi.org/10.1021/acs.jctc.6b00384` (2016).

51. Nikiforov, A., Gamez, J. A., Thiel, W., Huix-Rotllant, M. & Filatov, M. Assessment of approximate computational methods for conical intersections and branching plane vectors in organic molecules. The Journal of Chemical Physics **141,** 124122. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4896372/15485465/124122\_1\_online.pdf` (Sept. 2014).

52. Bearpark, M. J., Robb, M. A. & Bernhard Schlegel, H. A direct method for the location of the lowest energy point on a potential surface crossing. Chemical Physics Letters **223,** 269–274. ISSN: 0009-2614 (1994).

53. Sicilia, F., Blancafort, L., Bearpark, M. J. & Robb, M. A. New Algorithms for Optimizing and Linking Conical Intersection Points. Journal of Chemical Theory and Computation **4,** 257–266. eprint: `https://doi.org/10.1021/ct7002435` (2008).

54. Levine, B. G., Coe, J. D. & Martínez, T. J. Optimizing Conical Intersections without Derivative Coupling Vectors: Application to Multistate Multireference Second-Order Perturbation Theory (MS-CASPT2). The Journal of Physical Chemistry B **112,** 405–413. eprint: `https://doi.org/10.1021/jp0761618` (2008).

55. Yarkony, D. R. Systematic determination of intersections of potential energy surfaces using a Lagrange multiplier constrained procedure. The Journal of Physical Chemistry **97,** 4407–4412. eprint: https://doi.org/10.1021/j100119a026 (1993).

56. Maeda, S., Ohno, K. & Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. Journal of Chemical Theory and Computation **6,** 1538–1545. eprint: https://doi.org/10.1021/ct1000268 (2010).

57. Sanz García, J., Maskri, R., Mitrushchenkov, A. & Joubert-Doriol, L. Optimizing Conical Intersections without Explicit Use of Non-Adiabatic Couplings. Journal of Chemical Theory and Computation **0,** null. eprint: https://doi.org/10.1021/acs.jctc.4c00326 (0).

58. Dunlap, B. I. Robust and variational fitting. Phys. Chem. Chem. Phys. **2,** 2113–2116 (10 2000).

59. Feyereisen, M., Fitzgerald, G. & Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. Chemical Physics Letters **208,** 359–363. ISSN: 0009-2614 (1993).

60. Neville, S. P. & Schuurman, M. S. Removing the Deadwood from DFT/MRCI Wave Functions: The *p*-DFT/MRCI Method. Journal of Chemical Theory and Computation **17,** 7657–7665. eprint: https://doi.org/10.1021/acs.jctc.1c00959 (2021).

61. Neville, S. & Schuurman, M. GRaCI: General Reference Configuration Interaction version 0.1. Oct. 2021.

62. Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. Journal of Computational Chemistry **36,** 1664–1671. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23981 (2015).

63. Sun, Q. et al. Recent developments in the PySCF program package. The Journal of Chemical Physics **153,** 024109. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0006074/16722275/024109_1_online.pdf (July 2020).

64. Sun, Q. et al. PySCF: the Python-based simulations of chemistry framework. WIREs Computational Molecular Scienc **8,** e1340. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1340 (2018).

65. Viana, F. Things you wanted to know about the Latin hypercube design and were afraid to ask in (May 2013).

66. Boya Zhang, D. A. C. & Gramacy, R. B. Distance-Distributed Design for Gaussian Process Surrogates. Technometrics **63,** 40–52. eprint: https://doi.org/10.1080/00401706.2019.1677269 (2021).

67. Bonačiýć-Koutecký, V. et al. Angewandte Chemie (International ed.) **14,** 575–576 (1975).

68. Brooks, B. R. & Schaefer, H. F. I. Sudden polarization: pyramidalization of twisted ethylene. Journal of the American Chemical Society **101,** 307–311. eprint: https://doi.org/10.1021/ja00496a005 (1979).

69. Ben-Nun, M & Martınez, T. J. Photodynamics of ethylene: ab initio studies of conical intersections. Chemical Physics **259,** 237–248. ISSN: 0301-0104 (2000).

70. Olivucci, M., Ragazos, I. N., Bernardi, F. & Robb, M. A. A conical intersection mechanism for the photochemistry of butadiene. A MC-SCF study. Journal of the American Chemical Society **115,** 3710–3721. eprint: https://doi.org/10.1021/ja00062a042 (1993).

71. Levine, B. G. & Martínez, T. J. Isomerization Through Conical Intersections. Annual Review of Physical Chemistry **58,** 613–634. ISSN: 1545-1593 (2007).

72. Bearpark, M. J., Bernardi, F., Olivucci, M., Robb, M. A. & Smith, B. R. Can Fulvene S1 Decay Be Controlled? A CASSCF Study with MMVB Dynamics. Journal of the American Chemical Society **118,** 5254–5260. eprint: https://doi.org/10.1021/ja9542799 (1996).

73. Filatov, M. & Shaik, S. A spin-restricted ensemble-referenced Kohn–Sham method and its application to diradicaloid situations. Chemical Physics Letters **304,** 429–437. ISSN: 0009-2614 (1999).

74. Filatov, M. & Shaik, S. Diradicaloids: Description by the Spin-Restricted, Ensemble-Referenced KohnSham Density Functional Method. The Journal of Physical Chemistry A **104,** 6628–6636. eprint: https://doi.org/10.1021/jp0002289 (2000).

75. Thiel, W. Semiempirical quantum–chemical methods. WIREs Computational Molecular Science **4,** 145–157. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1161 (2014).

76. Thiel, W. in Advances in Chemical Physics 703–757 (John Wiley & Sons, Ltd, 1996). ISBN: 9780470141526. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470141526.ch10.

77. Hand, D. J. Measurements and their Uncertainties: A Practical Guide to Modern Error Analysis by Ifan G. Hughes, Thomas P. A. Hase. International Statistical Review **79,** 280–280. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-5823.2011.00149_8.x (2011).

78. Shu, Y., Varga, Z., Parameswaran, A. M. & Truhlar, D. G. Fitting of Coupled Potential Energy Surfaces via Discovery of Companion Matrices by Machine Intelligence. Journal of Chemical Theory and Computation **0,** null. eprint: https://doi.org/10.1021/acs.jctc.4c00716 (0).

79. Gutleb, T. S., Barrett, R., Westermayr, J. & Ortner, C. Parameterizing Intersecting Surfaces via Invariants 2024. arXiv: 2407.03731 [math.NA].

80. Swiler, L. P., Gulian, M., Frankel, A. L., Safta, C. & Jakeman, J. D. A Survey of Constrained Gaussian Process Regression: Approaches and Implementation Challenges. Journal of Machine Learning for Modeling and **1,** 119–156. ISSN: 2689-3967 (2020).