Machine Learned Interatomic Potentials for Photophysics

U2203761

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

13/03/2025

This project focuses on extending the usage of Machine-learned interatomic potentials (MLIPs) to model excited state properties in photophysical systems with near-quantum accuracy and significant computational efficiency. Foundational model, MACE-OFF23, was fine-tuned to create multi-headed models using Spin-Flip (SF) Time-Dependent Density Functional Theory (TDDFT) data for Furan, Oxirane and Di-Thymine. For Furan, the fine-tuned model achieves validation RMSEs of 2.2 meV/atom (for ground state) and 3.1 meV/atom (for first-excited state), demonstrating accuracy comparable to quantum mechanical methods with speed-ups of 10^2 to 10^5. This project's results highlight the potential of MLIPs for applications in photophysics, including sunscreen design and Organic Light-Emitting Diode (OLED) development. However, limitations arise in modelling non-adiabatic transitions, with models failing to capture conical intersections in molecular dynamics (MD) simulations. Future improvements include using Mixed-Reference Spin-Flip (MRSF) TDDFT, and extending to higher excited states.

I. Introduction

A. Machine Learned Interatomic Potentials

Since the 1990s [1], Machine Learned Interatomic Potentials (MLIPs) have emerged as transformative tools in computational sciences, promising to bridge the divide between computationally efficient but less precise empirical interatomic potentials and highly accurate but computationally expensive quantum mechanical methods [2, 3, 4]. With the significant advancements within the field of machine learning in the 21st century [5], MLIPs, which can predict atomic interactions and potential energy surfaces (PES) with near quantum ac-

curacy at a fraction of the computational cost, have become mainstays in computational sciences.

Thus far, MLIPs have mostly been employed to model ground-state properties, achieving significant success in areas such as materials science and molecular dynamics [6, 7, 8, 9, 10]. However, their potential in predicting excited-state properties, critical for research in photophysics, remains largely untapped and unexplored territory [11]. It is this application of MLIPs in predicting excited-state properties that forms the foundation of this project. MACE-OFF [12, 13] is a foundational model that uses equivariant neural networks to ensure predictions remain invariant to rotations, translations, or permutations of atoms. It also employs multi-head learning, allowing simultaneous predictions of multiple properties through separate output heads. This model is fine-tuned to predict excited-state potential energies and internal forces with precision comparable to quantum mechanical data generated using Time-Dependent Density Functional Theory (TDDFT). Hence offering a computationally efficient alternative to quantum mechanical methods for excited states.

B. Photophysics

This project is motivated by the need for accurate and efficient simulations of complex systems where excited states govern key behaviours. These systems are prevalent in Photophysics, the study of light-matter interactions which underpins many scientific fields such as materials science, chemical biology, and environmental chemistry. Photophysical processes are central to the development of organic photovoltaics [14], the design of fluorescent probes for biological imaging [15], and the study of atmospheric photochemistry [16].

A major challenge in photophysics is accurately understanding molecules in their excited states and the key phenomena they exhibit, such as fluorescence, phosphorescence, and non-radiative decay. Greater understanding of these phenomena could enable advancements in applications ranging from Organic Light-Emitting Diode (OLED) development to light-stable drug creation and designing sunscreens with tailored photostability [17, 18, 19].

Quantum mechanical methods such as Linear Response (LR) and Spin-Flip (SF) TDDFT have been the standard tools for simulating excited-state properties in the 21st century. However, despite offering high accuracy, they come with significant computational costs, which often become prohibitive for systems requiring extensive PES sampling or large-scale simulations [20, 21]. Machine-learned interatomic potentials (MLIPs) present a promising alternative, providing computational speed-ups

of 10^2 to 10^5 compared to TDDFT without sacrificing accuracy [22].

This project leverages MLIPs to model excited-state properties, focusing on small molecular systems such as furan, oxirane, and thymine dimers, which exhibit photophysical behaviors including ring-opening reactions, conical intersections, and UV-induced DNA damage [23]. Conical intersections, points of degeneracy between electronic states, are critical for ultrafast photophysical processes but challenging to model due to strong coupling between nuclear and electronic motion [24, 25]. By fine-tuning the foundational MACE-OFF model with high-accuracy quantum mechanical data, the goal is to predict excited-state potential energy surfaces (PES) with precision comparable to TDDFT at speeds orders of magnitude faster.

II. Background Theory

A. Quantum Mechanical Methods Background Theory

i. Density Functional Theory (DFT)

Density Functional Theory (DFT) is a quantum mechanical method used to calculate ground-state properties of many-electron systems. At its foundation, DFT rests on the Hohenberg-Kohn theorems [26], which assert that the ground-state properties of many-electron systems are uniquely determined by their electron densities, $n(\mathbf{r})$. The complex many-body Schrödinger equation is thus simplified by expressing the system's total energy as a function of the electron density, $E[n]$:

$$E[n] = T_s[n] + V_{\text{ext}}[n] + J[n] + E_{\text{xc}}[n] \quad (1)$$

where $T_s[n]$ represents the kinetic energy of non-interacting electrons; $V_{\text{ext}}[n]$ represents the interaction with external potentials; $J[n]$ represents the classical Coulomb energy; and $E_{\text{xc}}[n]$ represents the exchange-correlation energy.

The practical implementation of DFT, however, was revolutionized by the Kohn-Sham formalism [27], which introduces an auxiliary system of non-interacting electrons that reproduces the ground-state density of the true interacting system. The Kohn-Sham equations take the form:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \quad (2)$$

where $V_{\text{eff}}(\mathbf{r})$ is the effective potential, which incorporates the external potential, Coulomb interactions, and exchange-correlation contributions. The electron density is then reconstructed as:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \quad (3)$$

where $\psi_i(\mathbf{r})$ are the Kohn-Sham orbitals.

The Born-Oppenheimer approximation separates the motion of electrons and nuclei by assuming that nuclei are stationary relative to the fast-moving electrons. This approximation arises from the large mass difference between nuclei and electrons, allowing the electronic Schrödinger equation to be solved for a fixed nuclear configuration, while the nuclei move on a potential energy surface determined by the electrons [28, 29]. This separation of electronic and nuclear motion forms the basis for methods like DFT, enabling efficient calculations of molecular properties.

However, the Born-Oppenheimer approximation breaks down in systems where electronic and nuclear motions are strongly coupled, such as near conical intersections or during non-adiabatic processes [30]. In these cases, the coupling between electronic states and nuclear motion cannot be neglected, therefore more advanced methods that explicitly account for non-adiabatic effects, such as TDDFT are needed.

DFT's accuracy is further limited by the quality of the exchange-correlation functional, $E_{\mathrm{xc}}[n]$. Various approximations, such as the Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA), have been developed, balancing computational feasibility with predictive capability, and hybrid functionals, which incorporate a fraction of exact Hartree-Fock exchange, further improve accuracy but come at an increased computational cost [31, 32].

ii. Time Dependent Density Functional Theory(TDDFT)

DFT is traditionally employed only for ground-state calculations. To describe excited-state properties, TDDFT extends the capabilities of DFT, addressing the dynamics of electronic systems under time-dependent perturbations, such as those induced by electromagnetic fields. Its foundational principle is the Runge-Gross theorem [33], which establishes a one-to-one correspondence between the time-dependent external potential $V_{\mathrm{ext}}(\mathbf{r}, t)$ and the time-dependent electron density $n(\mathbf{r}, t)$. This ensures that the time-dependent properties of a many-body system can be uniquely determined by $n(\mathbf{r}, t)$ [34].

iii. Linear Response TDDFT (LR TDDFT)

TDDFT is most commonly used in its Linear Response form to calculate excitation energies and other properties of excited states. In this approach, small external perturbations are applied to the system, and the resulting changes in the electron density are analysed [35]. In linear-response (LR) TDDFT, the density-density response function, $\chi(\mathbf{r}, \mathbf{r}'; \omega)$, de-

scribes how the electron density at position $\mathbf{r}$ changes in response to a perturbation applied at position $\mathbf{r}'$ with frequency ω. This response function is related to its non-interacting counterpart, $\chi_0(\mathbf{r}, \mathbf{r}'; \omega)$, via an equation similar to the Dyson equation:

$$\chi(\mathbf{r}, \mathbf{r}'; \omega) = \chi_0(\mathbf{r}, \mathbf{r}'; \omega) + \int \chi_0(\mathbf{r}, \mathbf{r}_1; \omega) K_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2; \omega) \times \chi(\mathbf{r}_2, \mathbf{r}'; \omega) \, d\mathbf{r}_1 d\mathbf{r}_2 \tag{4}$$

where K_{xc} is the exchange-correlation kernel.

In LR-TDDFT, excitation energies are calculated by analysing the response function. The excitation energies correspond to the frequencies at which the system undergoes a resonance, and these frequencies are found by solving the following eigenvalue equation:

$$AX - BY = \omega X \tag{5}$$

where the matrices A and B contain information about orbital energy differences and Coulomb integrals. The eigenvalues ω correspond to excitation energies, while the eigenvectors X and Y provide transition amplitudes.

LR-TDDFT has been widely used for calculating singlet-singlet excitations in closed-shell systems, often used in simulating UV-Vis absorption spectra, fluorescence, and other phenomena. However, its reliance on single excitations can limit its accuracy for systems with significant multi-reference character.

iv. Spin Flip TDDFT (SF TDDFT)

Spin-Flip (SF) TDDFT is designed to handle excited states with significant spin contributions, such as singlet-to-triplet transitions or systems with open-shell ground states. SF-TDDFT incorporates spin-flipping operators into the response framework, enabling the description of both spin-conserving and spin-changing transitions [36]. In the case of a triplet ground state, a spin-flip excitation corresponds to a transition where an electron is promoted between orbitals with opposite spin. The resulting excited states can exhibit singlet, triplet, or even multi-reference character, which SF-TDDFT captures with remarkable efficiency.

A key advantage of SF-TDDFT is that both ground and excited states are treated as excitations from the same reference configuration. This ensures that the potential energy surfaces (PESs) of these states intersect reliably, making SF-TDDFT particularly well-suited for modelling conical intersections and non-adiabatic processes (transitions or reactions in quantum systems where the Born-Oppenheimer approx-

imation breaks down, meaning the electronic and nuclear motions are strongly coupled). In contrast, LR-TDDFT, which treats excited states as single excitations from the ground state, often fails to accurately describe such intersections due to its reliance on a single-reference framework. This makes SF-TDDFT particularly valuable for systems such as diradicals, transition metal complexes, and photochemical processes where spin-state dynamics and conical intersections play a critical role. This is the method this project uses to produce quantum-mechanical (QM) training data.

v. Mixed Reference SF TDDFT (MRSF TDDFT)

While SF-TDDFT offers significant advantages for modelling systems with open-shell ground states and spin-state dynamics, it is not without limitations. One key drawback of SF-TDDFT is its tendency to produce impure states, where the resulting wavefunctions are no longer eigenstates of the spin operator $\hat{S}^2$. This can lead to inaccuracies in the description of spin multiplicities and complicate the interpretation of results, particularly for systems with strong multi-reference character [37]. Additionally, SF-TDDFT can struggle with state mixing, where transitions between states of different spin multiplicities are not cleanly separated, further limiting its reliability for certain photophysical processes. These problems in SF are encountered in this project during the Furan dataset curation.

To address these limitations, Mixed-Reference Spin-Flip TDDFT (MRSF-TDDFT) has been developed as an extension of SF-TDDFT. MRSF-TDDFT combines the advantages of SF-TDDFT—such as its ability to describe conical intersections and non-adiabatic processes—with a mixed-reference approach that improves the accuracy of spin-state descriptions and reduces issues with state mixing [38]. By incorporating multiple reference configurations, MRSF-TDDFT better captures the multi-reference character of systems like diradicals and diradicaloids, while maintaining the computational efficiency of linear-response methods.

MRSF-TDDFT has been shown to provide more reliable results for vertical excitation energies, singlet-triplet energy gaps, and conical intersections compared to traditional SF-TDDFT [39]. However, like SF-TDDFT, MRSF-TDDFT remains computationally demanding, and its performance depends on the choice of exchange-correlation functional and reference configurations.

B. Molecular Systems

i. Furan

Furan (C_4H_4O) is a five-membered heterocyclic molecule that exhibits fascinating photophysical behaviour, making it an ideal system for studying excited-state dynamics. Upon photo-excitation, furan undergoes two primary processes: ring-opening and ring-puckering. The ring-opening mechanism involves the breaking of the C–O bond, leading to the formation of a bi-radical intermediate, while the ring-puckering mechanism involves out-of-plane distortions of the ring structure [40, 41]. Conical intersections (see Fig. 1), where the potential energy surfaces (PESs) of the ground and excited states intersect, facilitate ultrafast non-radiative transitions.

The competition between ring-opening and ring-puckering pathways is a key feature of Furan's photophysics. Recent studies have shown that the ring-puckering pathway is favoured in the excited state due to lower energy barriers, while the ring-opening pathway dominates in the ground state [42]. This duality makes furan an excellent model system to explore.

ii. Oxirane

Oxirane (C_2H_4O), a three-membered cyclic ether, is a classic model system for studying photochemically driven ring-opening reactions. Upon photon absorption, oxirane is excited to a higher electronic state, leading to the breaking of the C–C bond. This process involves non-adiabatic transitions through conical intersections (see Fig. 2) [44, 45].

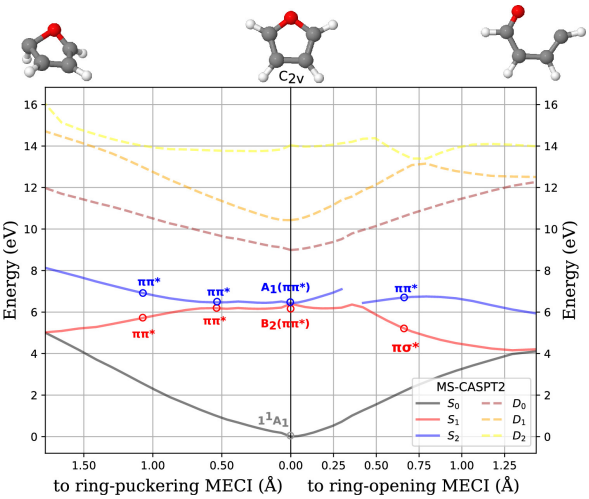

Figure 1: Potential energy surface (PES) for Furan, showing the ring-opening and ring-puckering pathways. The ground, first-excited, and second-excited singlet states are represented in grey, red, and blue, respectively. Conical intersections (where different state PES lines intersect) are at the ends of both pathways. Figure taken from [43].

The ring-opening reaction of oxirane [46] is highly sensitive to the choice of computational method, as it involves significant multi-reference character and strong non-adiabatic couplings. SF-TDDFT and MRSF-TDDFT have been shown to accurately capture the excited-state dynamics of oxirane, including the formation of bi-radical intermediates and the role of conical intersections [47].

iii. Di-Thymine

Di-Thymine, formed by the covalent bonding of two thymine bases under UV irradi-

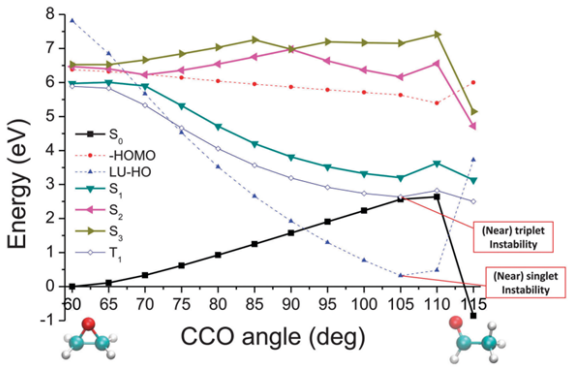

Figure 2: Potential energy surface (PES) for Oxirane, showing the ring-opening pathway with the ground, first-excited and second-excited singlet states in black, blue and pink respectively. Figure taken from [48] created using constrained optimisation scans with varied CCO angles.

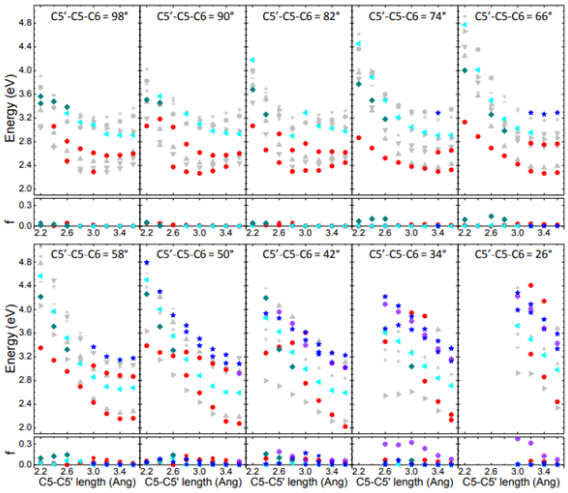

Figure 3: Potential energy surfaces (PESs) for Di-Thymine, plotted with energies against C5'-C5 bond lengths for various C5'-C5-C6 bond angles. The ground, first-excited and second-excited states given by red, light-blue and green respectively. Figure taken from [52].

ation, is a critical system for understanding UV-induced DNA damage [49, 50]. The photochemical pathway involves excitation to an excited state, followed by non-radiative relaxation through conical intersections, leading to the formation of cyclobutane pyrimidine dimers (CPDs) [51].

The repair of thymine dimers is also a photophysical process, involving the cleavage of the C–C bonds between the thymine bases. This process is mediated by excited-state dynamics and conical intersections, making it a challenging system to model accurately.

C. Machine Learning Models

i. Interatomic Potentials & MLIPs

Traditional interatomic potentials rely on predefined mathematical expressions to approximate potential energy surfaces (PESs), making them computationally efficient but limited in generality. These models often fail to capture complex bonding environments or electronic effects, particularly in photophysical contexts [53, 54]. Machine learning interatomic potentials (MLIPs), however, address these shortcomings by learning from quantum mechanical reference data rather than using predefined mathematical expressions. MLIPs approximate the total energy of a system as a sum of atomic contributions:

$$E_{\text{total}} = \sum_i E_i \tag{6}$$

where E_i represents the energy of the i-th atom's local environment.

Symmetry-preserving descriptors, such as Behler-Parrinello functions, Smooth Overlap of Atomic Positions (SOAP), and graph-based methods, encode atomic en-

vironments in a rotationally, translationally, and permutationally invariant manner [55]. This ensures equivariance, allowing the model to generalize across configurations and maintain consistent predictions regardless of the system's orientation.

MLIPs use algorithms like neural networks (NNs) and Gaussian process regression (GPR) to model relationships between descriptors and properties. NNs excel at handling large datasets and complex mappings, while GPR provides uncertainty estimates. These capabilities make MLIPs particularly suited for photophysics, where they can accurately capture critical features like conical intersections and steep PES gradients, essential for modelling excited-state dynamics.

ii. The MACE Framework

The MACE framework is a state-of-the-art MLIP that combines the accuracy of quantum mechanical methods with the computational efficiency of machine learning. MACE employs equivariant neural networks, ensuring accurate predictions across diverse molecular configurations. A key innovation of MACE is its use of mutli-head learning, which allows a single model to simultaneously predict multiple properties, such as ground-state and excited-state energies, using separate output heads. This approach enhances the model's versatility and transferability, making it well-suited for complex photophysical systems.

Additionally, MACE's architecture is designed to handle large datasets efficiently, making it a powerful tool for high-throughput simulations and active learning workflows.

A notable implementation of MACE is MACE-OFF23, a transferable organic force field trained on the SPICE dataset [56, 57]. MACE-OFF23 is designed for organic chemistry applications, covering 10 chemical elements (H, C, N, O, F, P, S, Cl, Br, I). The model is trained to reproduce energies and forces computed at the ωB97M-D3(BJ)/def2-TZVPPD level of theory.

This project utilizes the MACE-OFF23 Medium model as a foundational model, finetuning it using the mutli-head replay protocol to predict excited-state properties for Oxirane, Furan, and Di-Thymine.

iii. Training MLIPs

The accuracy of MLIPs depends on their training process and minimizing the error between predictions and quantum mechanical (QM) reference data. In the MACE framework, training is achieved by optimizing a weighted loss function:

$$L = \sum_k w_k L_k \tag{7}$$

where L_k is the loss for the k-th head and w_k is its weight. By balancing these con-

tributions, the model reconciles discrepancies between datasets, enhancing robustness and transferability. In this project, one head predicts ground-state energies, while another predicts excited-state properties, allowing MACE to simultaneously model multiple electronic states with high accuracy.

MACE-OFF23 exemplifies this approach, as it is trained on the SPICE dataset to reproduce ωB97M-D3(BJ)/def2-TZVPPD energies and forces. The training process involves partitioning the dataset into training (95%) and testing (5%) subsets, ensuring that the model generalizes well to unseen configurations. This rigorous training protocol enables MACE-OFF23 to achieve high accuracy for organic molecules, making it a reliable tool for photophysical applications. This protocol prevents catastrophic forgetting by replaying a portion of the original training data during finetuning, ensuring that the model retains its generalizability while adapting to new datasets.

Other techniques, such as transfer learning and active learning, further refine MLIPs. Transfer learning adapts pretrained models to new datasets, reducing computational costs and improving performance on systems with limited QM data. Active learning identifies the most informative data points for additional QM calculations, ensuring that the training dataset is both comprehensive and efficient [58].

III. Methodology

This project was conducted collaboratively, with each partner focusing on distinct molecular systems. My work centred on furan, encompassing dataset creation, model training, and analysis of excited-state property predictions, while my partner, Lucas, focused on oxirane and di-thymine. Additional systems, such as fulvene, retinal, and azobenzene, were considered but not explored further due to time constraints. All calculations were performed using the def2-TZVPPD basis set and the wB97M-D3BJ functional, implemented in the ORCA quantum chemistry package [59]. The MACE-OFF23 model was fine-tuned using mutli-head replay protocol to predict excited-state properties, with the aim for results to be comparable to high-level quantum mechanical (QM) data.

A. Computational Environment

All calculations were performed on Warwick's Scientific Computing Research Technology Platform (SCRTP), a Linux-based high-performance computing cluster. The workflow was managed using Unix command-line operations, with Jupyter Notebooks employed for scripting, data handling, and visualization. The ASE (Atomic Simulation Environment) library

was used in conjunction with ESTEEM [60, 61] to interface ORCA and MACE.

B. Furan

The study of furan involved the generation of geometries, quantum mechanical calculations, dataset creation, model training, and evaluation, as outlined below.

i. Geometry Generation

For the ring-opening pathway, constrained geometry optimizations were performed using ORCA, systematically varying the CCO bond angle. For the ring-puckering pathway, geometries were generated by linearly interpolating between an initial geometry (optimized ground-state structure) and a final geometry obtained from literature [40]. This approach allowed for efficient sampling of the puckering pathway.

ii. Quantum Mechanical Calculations

Ground-state energies and forces were computed using DFT, while excited-state properties were calculated using LR-TDDFT and SF-TDDFT. SF-TDDFT was chosen for its ability to capture spin-state mixing and conical intersections. Single-point energy calculations (engrad runs) were performed for each geometry to obtain energies and forces for both ground and excited states.

iii. Dataset Creation

The QM data (energies and forces) were compiled into datasets for the ground state and first excited state. These datasets were used to train and fine-tune separate heads in the MACE-OFF23 model.

iv. Model Training and Evaluation

The foundational MACE-OFF23 model was finetuned using the mutli-head replay protocol, which prevents catastrophic forgetting by replaying a portion of the original training data during finetuning. Separate output heads were used for the ground state and first excited state. The model was trained with varying hyper-parameters to find a balance between computational efficiency and accuracy. Active learning was employed to identify the most informative geometries for additional QM calculations, improving dataset quality.

The finetuned MACE models were compared to ORCA SF-TDDFT ground-truth data and the un-finetuned raw MACE-OFF23 model. Comparative plots of PESs were generated to evaluate model performance. Molecular dynamics (MD) simulations were attempted starting from the Franck-Condon point in the excited state.

C. Oxirane and Di-Thymine

Lucas' work focused on oxirane and di-thymine, employing similar methodologies

to those used for furan.

i. Oxirane

For oxirane, constrained geometry optimizations were performed using ORCA, varying the CCO bond angle to explore the ring-opening pathway. Ground-state and excited-state properties were calculated using SF-TDDFT, and the resulting data was used to create datasets for model training.

ii. Di-Thymine

Similarly, for di-thymine, constrained geometry optimizations were performed, varying both the C5–C5' bond length and the C5–C5–C6 bond angle to explore the dimerisation pathway. Ground-state and excited-state properties were calculated using SF-TDDFT, with the resulting data compiled into datasets for model training.

IV. Results

A. Quantum Mechanical Training Data

The generation of high-quality quantum mechanical (QM) data is critical for training MLIPs.

i. Furan

For Furan, Spin-Flip (SF) and Linear-Response (LR) TDDFT calculations were performed on the ring-opening and ring-puckering geometries created via ORCA geometry optimisations and linear interpolation. In Fig. 4, SF-TDDFT results demonstrate smoother and more physically consistent PESs compared to LR-TDDFT. For the ring-opening mechanism, the Franck–Condon point is located at approximately 37°, while for the ring-puckering mechanism, it corresponds to Geometry number 0, consistent with previous studies [43]."

ii. Oxirane

For Oxirane, SF-TDDFT calculations were performed on geometries generated using constrained optimizations, varying the CCO bond angle. Fig. 5 shows the resulting PES, with the Franck-Condon point at 60°. The two states converge at approximately 107°, corresponding to the S0/S1 conical intersection, in agreement with previous research [48]. The ground state PES exhibits a significant drop after this point, reflecting the reduction in ring strain as the three-membered ring opens.

iii. Di-Thymine

For Di-Thymine, SF-TDDFT calculations were performed on geometries generated by varying both the C5-C5' bond length and the C5'-C5-C6 bond angle (see Fig. 6). Fig. 7 shows the resulting PES plots, which

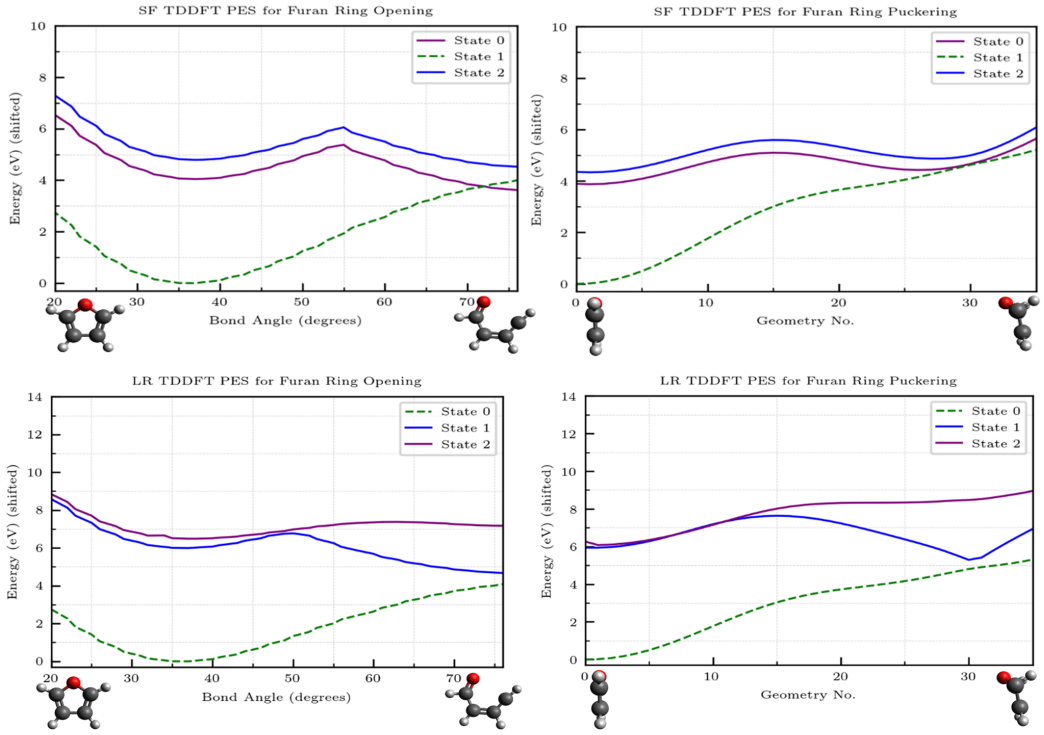

Figure 4: Furan PES plots for [Top] Spin-Flip (SF) and [Bottom] Linear-Response (LR) TDDFT. [Left] Displays the ring-opening mechanism plotted with energies against CCO bond angles, and [Right] displays the ring-puckering mechanism plotted with energies against geometry indexes for a series of linear interpolated geometries. All plots were created using ORCA engrad runs with iroots = [0,1,2]. For SF, the ground and first-excited singlet states are given by green and blue respectively, with purple representing a triplet first-excited state provided by ORCA iroot 0. For LR, the 0,1, and 2 iroots are plotted with green, blue and purple respectively.

reveal numerous conical intersections between the ground and excited states. A notable feature is the degeneracy at a bond angle of 43.14°, where a stretch of bond lengths exhibit degeneracy between the two states. These results show some yet sufficient agreement with past research [52], demonstrating the utility of SF-TDDFT for modelling complex photophysical processes.

B. Training and Hyperparameter Selection

The finetuning of the MACE-OFF model involved training on the QM datasets with varying hyper-parameters, including the number of epochs (complete passes through the training dataset) and batch sizes (the number of training examples processed before the model updates its weights). Learning curves, shown in Fig. 8, illustrate the relationship between these hyper-parameters and the model's performance. At a fixed batch size, the validation Root

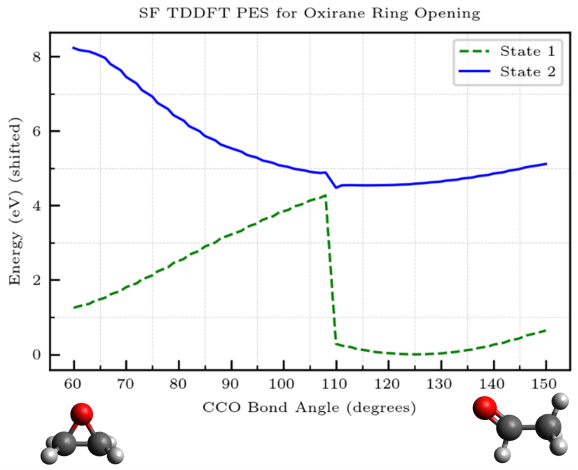

Figure 5: PES for Oxirane over a range of CCO angles developed using Spin-Flip (SF) TDDFT in ORCA. States 1 and 2 are represented by green and blue lines respectively.

Mean Square Error (RMSE) decreases as the number of epochs increases, indicating improved model accuracy with more training iterations. Conversely, at a fixed number of epochs, the validation RMSE fluctuates with changes in batch size, suggesting that batch sizes too small or too large may lead to more stable training.

The optimal model was determined to have 300 epochs and a batch size of 5, as it achieved a balance between accuracy and computational efficiency. This model exhibited a validation RMSE of 2.2

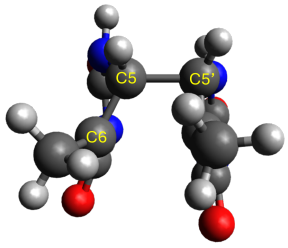

Figure 6: Di-Thymine structure with atoms C5, C5', and C6 labelled in yellow.

meV/atom for the ground state (es1 head) and 3.1 meV/atom for the first excited state (es2 head), with corresponding force RMSEs of 21.7 meV/Å and 19.0 meV/Å, respectively. These results demonstrate the model's ability to accurately reproduce both energies and forces for the training datasets.

Fig. 9 plots validation RMSE against training RMSE, revealing a trend where validation errors are lower than training errors. This is unusual, as models typically fit closer to the training data. The observed trend suggests that the training dataset contains a significant number of challenging geometries, such as those near conical intersections, which are more difficult for the model to accurately represent.

C. Performance of the finetuned Model

In Fig. 10, it can bee seen that the finetuned MACE-OFF model significantly outperforms the raw foundational model. For the ground state, the finetuned model accurately reproduces the PES for both ring-opening and ring-puckering mechanisms, closely matching the ORCA "ground-truth" curves. In contrast, the raw model struggles, particularly for the ring-opening mechanism, where it overshoots the expected energy changes. For the excited state, the finetuned model's es2

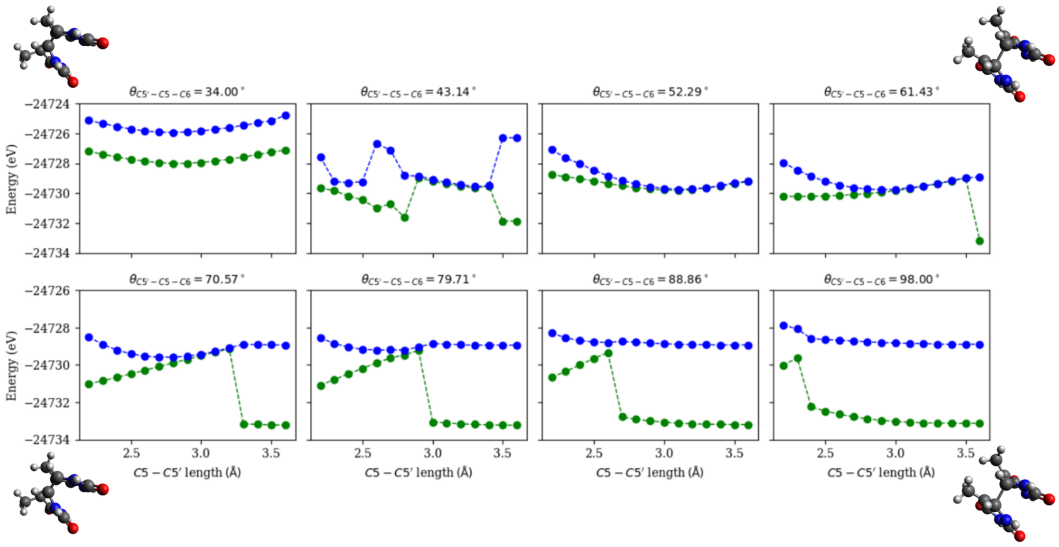

Figure 7: Potential energy surface (PES) plots for Di-Thymine with energies against C5'-C5 bond lengths for various C5'-C5-C6 bond angles created using ORCA SF TDDDFT engrad runs on geometries created with ORCA constrained optimisations. The ground and first-excited states are given by green and blue respectively.

head demonstrates remarkable accuracy, with validation and training RMSEs of 3.1 meV/atom and 5.0 meV/atom, respectively. This demonstrates the viability of MLIP usage in excited-state physics and a significant improvement over the raw model, which lacks the capability to even model excited states.

D. Excited-State Molecular Dynamics

To explore the potential of the finetuned model for excited-state molecular dynamics (MD), Fig. 11 shows an attempt at running excited-state MD using finetuned MACE-OFF. It is expected that from the Franck-Condon geometries for both the opening and puckering mechanisms, the excited-state trajectories will approach the S0/S1 conical intersections, at which point the energy gap between the ground and excited states will be small enough (i.e. <0.2 eV), such that a trajectory hop to ground-state trajectories from the S0/S1 conical intersection can be performed and the new trajectory followed until the it reaches the Franck-Condon geometry again in the ground state. However, the states never converge as seen in the figure, and a trajectory hop is not possible. This highlights a limitation of the current approach for modelling non-adiabatic transitions.

V. Conclusions

This project demonstrates the potential of machine-learned interatomic potentials (MLIPs) modelling excited-state proper-

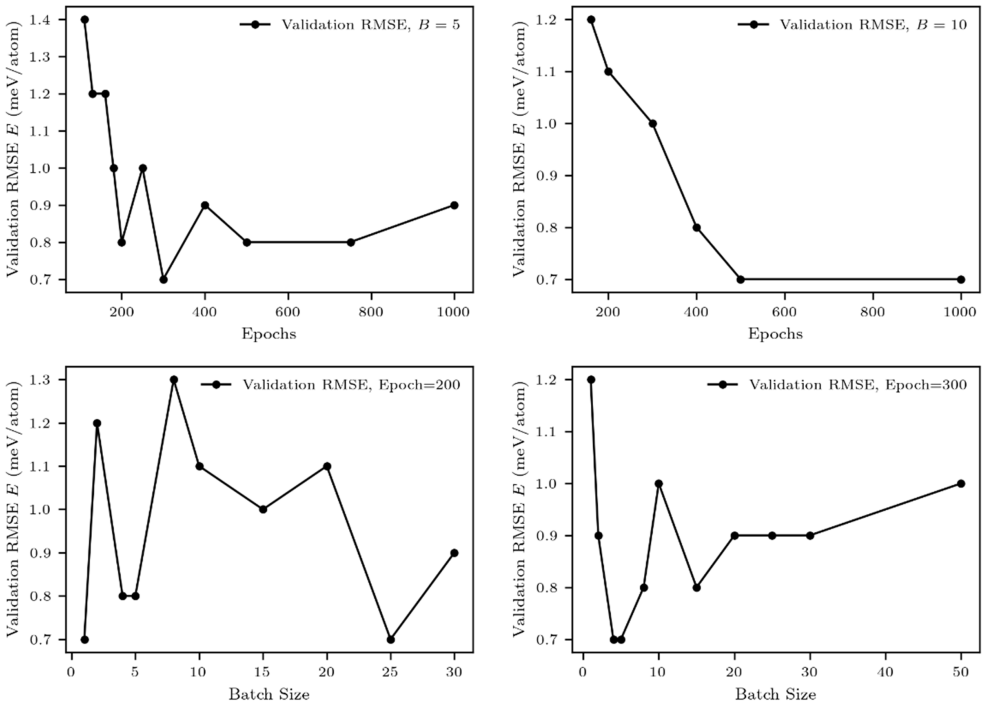

Figure 8: Learning curves showing validation RMSE (Root Mean Square Error) for trained models. [Top] RMSE vs. epochs at fixed batch sizes: [Left] 5 and [Right] 10. [Bottom] RMSE vs. batch sizes at fixed epochs: [Left] 200 and [Right] 300.

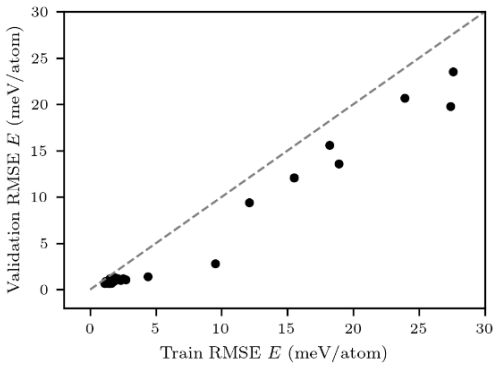

Figure 9: Validation RMSE against Training RMSE for many fine-tuned MACE-OFF models. The dotted line represents Training RMSE = Validation RMSE.

ties for photophysical systems. By fine-tuning the MACE-OFF23 model using high-quality quantum mechanical (QM) data generated from Spin-Flip (SF) Time-Dependent Density Functional Theory (TDDFT), we achieved accurate predictions of potential energy surfaces (PES) for Furan, Oxirane, and Di-Thymine. The finetuned model exhibited validation RMSEs of 2.2 meV/atom for the ground state and 3.1 meV/atom for the first excited state, demonstrating its ability to reproduce both energies and forces with near-quantum accuracy. These results highlight the promise of MLIPs as a computationally efficient alternative to traditional quantum mechanical methods, offering speed-ups of 10^2 to 10^5 while maintaining high precision.

The success of this project underscores the applicability of MLIPs in photophysics, with potential applications in fields such as sunscreen production, OLED development, and light-stable drug design. By accurately

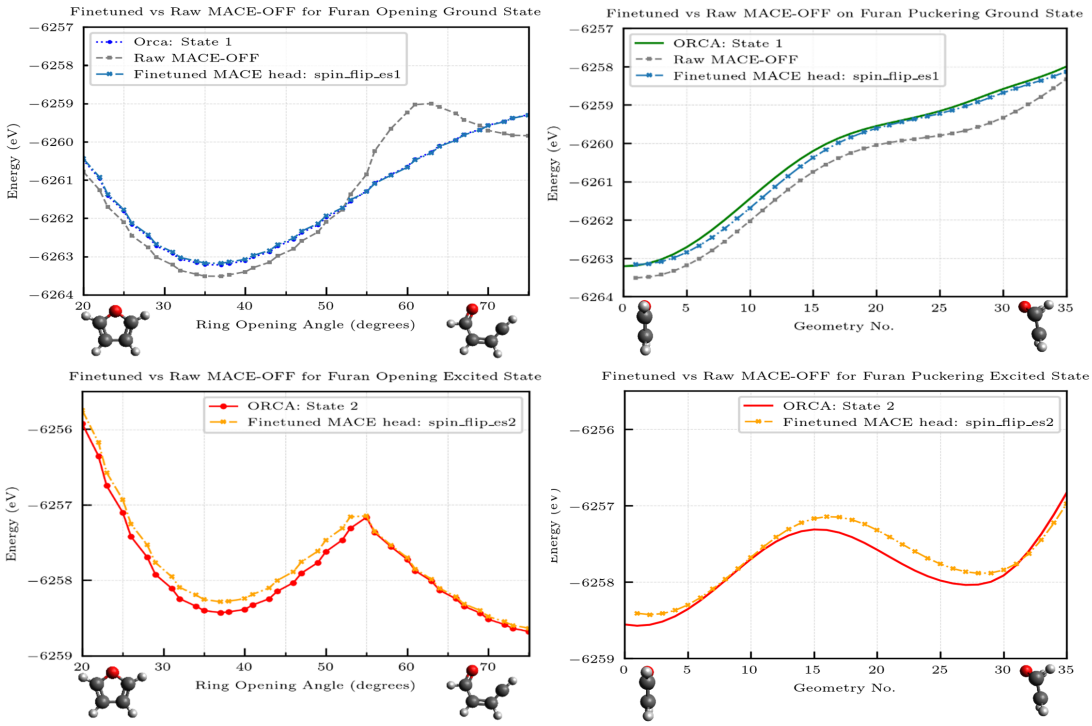

Figure 10: PES plots comparing finetuned MACE-OFF against its raw foundational model's performance at reproducing [Left] ring-opening and [Right] ring-puckering potential energy surfaces (PES). [Top] Plots Display the `"ground truth"` SF TDDFT ORCA ground state in dark blue/green and Raw and finetuned MACE-OFF using `head(spin_flip_es1)` in light blue and grey, respectively. [Bottom] Plots display the ORCA SF TDDFT first-excited state in red and finetuned MACE-OFF using `head(spin_flip_es2)` in orange. Ground and Excited States are plotted separately to allow clearer differentiation between ORCA and finetuned MACE-OFF lines.

modelling excited-state dynamics, MLIPs can provide insights into key photophysical processes, such as fluorescence, phosphorescence, and non-radiative decay, which are critical for understanding light-matter interactions. The ability to predict conical intersections and non-adiabatic transitions with high accuracy opens new avenues for studying complex systems where excited states govern key behaviours.

However, several limitations and areas for improvement were identified during this project. Firstly, the use of MRSF-TDDFT instead of SF-TDDFT could improve the accuracy of the QM datasets, particularly for systems with strong multi-reference character. MRSF-TDDFT addresses some of the limitations of SF-TDDFT, such as state mixing and impure spin states, making it a more robust choice for generating training data.

Secondly, the geometry generation process for Furan's ring-puckering mechanism could be improved/corrected by using the Nudged Elastic Band (NEB) method instead of linear interpolation. Linear in-

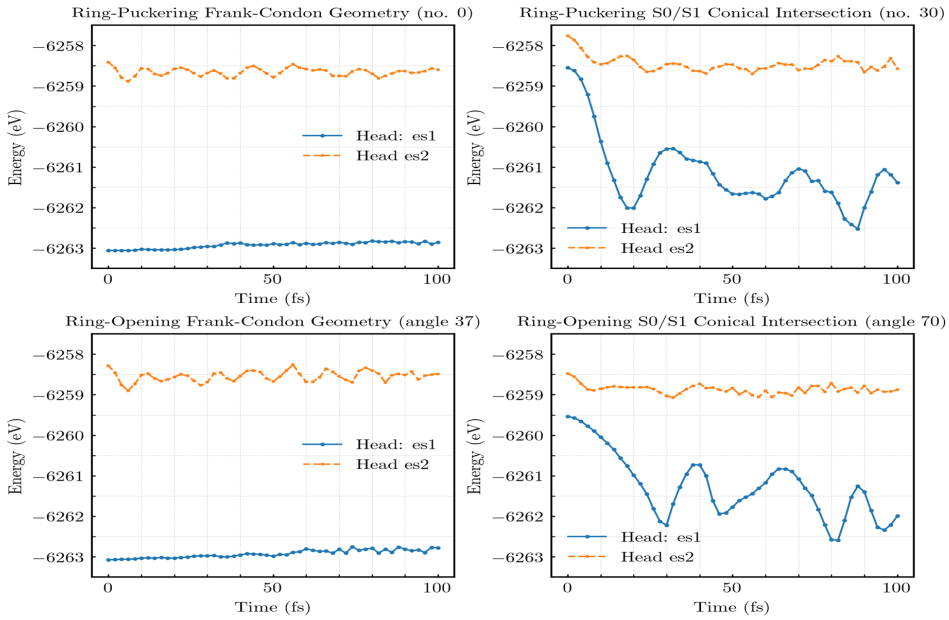

Figure 11: Molecular dynamics (MD) simulations for Furan using finetuned MACE-OFF. [Top] MD runs for Ring-Puckering from [Left] Franck-Condon Geometry (i.e. Geometry number 0), and [Right] S0/S1 Conical Intersection Geometry (i.e. Geometry number 30). [Bottom] MD runs for Ring-Opening from [Left] Franck-Condon Geometry (i.e. Angle 37), and [Right] S0/S1 Conical Intersection Geometry (i.e. Angle 70). Finetuned multi-head model heads es1 and es2 represented by blue and orange lines respectively.

terpolation does not account for the natural rotation of hydrogen atoms or fixed bond lengths, leading to less accurate geometries. NEB, which identifies minimum energy pathways, would provide a more realistic representation of the reaction pathway and improve the quality of the training data.

Additionally, this project focused on modelling only the first excited state. Expanding the scope to include additional heads corresponding to additional excited states would provide a more comprehensive understanding of MLIPs' ability to reproduce photophysical processes and the study of systems with complex electronic structures.

Finally, while the finetuned model performed well in reproducing PESs, it struggled to capture non-adiabatic transitions in molecular dynamics (MD) simulations. The model's inability to close the energy gap between the ground and excited states prevented trajectory hops at conical intersections, highlighting a limitation in its ability to model real-life photophysical simulations. Further refinement of the model, including the incorporation of more QM data and improved training protocols, could address this issue and enable more accurate MD simulations.

In conclusion, this project demonstrates that MLIPs are a powerful tool for predicting excited-state properties and mod-

elling photophysical processes. The fine-tuned MACE-OFF23 model achieved remarkable accuracy in reproducing PESs for multiple molecular systems, offering significant computational speed-ups compared to traditional quantum mechanical methods. While challenges remain, particularly in modelling non-adiabatic transitions, the results are highly promising and pave the way for future applications of MLIPs in photophysics and related fields. By addressing the identified limitations and exploring new methodologies, MLIPs can be further refined to unlock their full potential in computational chemistry and materials science.

References

[1] Blank, T. B., Brown, S. D., Calhoun, A. W., Doren, D. J., *J. Chem. Phys.* **103**, 4129 (1995)

[2] Behler, J., Csányi, G., *Eur. Phys. J. B* **94**, 142 (2021)

[3] Rosenbrock, C. W., Gubaev, K., Shapeev, A. V., et al., *npj Comput. Mater.* **7**, 24 (2021)

[4] Behler, J., Parrinello, M., *Phys. Rev. Lett.* **98**, 146401 (2007)

[5] Tufail, S., Riggs, H., Tariq, M., Sarwat, A. I., *Electronics* **12**, 1789 (2023)

[6] Yu, W., Ji, C., Wan, X., Zhang, Z., Robertson, J., Liu, S., Guo, Y., *Int. J. Mech. Syst. Dyn.* **1**, 159 (2021)

[7] Byggmästar, J., et al., *J. Phys.: Condens. Matter* **34**, 305402 (2022)

[8] Long, T., Li, J., Wang, C., Wang, H., Cheng, X., Lu, H., Zhang, Y., Zhou, C., *Polymer* **308**, 127416 (2024)

[9] Mishin, Y., *Acta Mater.* **214**, 116980 (2021)

[10] Liang, W., Lu, G., Yu, J., *Adv. Theory Simul.* **3**, 2000180 (2020)

[11] Westermayr, J., Marquetand, P., *Chem. Rev.* **121**, 9873 (2021)

[12] Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C., Csányi, G., *arXiv* 2206.07697 (2023)

[13] Kovács, D. P., Moore, J. H., Browning, N. J., Batatia, I., Horton, J. T., Kapil, V., Witt, W. C., Magdău, I.-B., Cole, D. J., Csányi, G., *arXiv* 2312.15211 (2023)

[14] Brédas, J. L., Norton, J. E., Cornil, J., Coropceanu, V., *Acc. Chem. Res.* **42**, 1691 (2009)

[15] Lavis, L. D., Raines, R. T., *ACS Chem. Biol.* **9**, 855 (2014)

[16] Finlayson-Pitts, B. J., Pitts, J. N., *Science* **276**, 1045 (1997)

[17] Rodrigues, N. D. N., Staniforth, M., Stavros, V. G., *Proc. R. Soc. A* **472**, 20160677 (2016)

[18] Baker, L. A., Marchetti, B., Karsili, T. N. V., Stavros, V. G., Ashfold, M. N. R., *Chem. Soc. Rev.* **46**, 3770 (2017)

[19] Rodrigues, N. D. N., Stavros, V. G., *Sci. Prog.* **101**, 8 (2018)

[20] Casida, M. E., Huix-Rotllant, M., *Annu. Rev. Phys. Chem.* **63**, 287 (2012)

[21] Dreuw, A., Head-Gordon, M., *Chem. Rev.* **105**, 4009 (2005)

[22] Bartók, A. P., Payne, M. C., Kondor, R., Csányi, G., *Phys. Rev. Lett.* **104**, 136403 (2010)

[23] Domcke, W., Yarkony, D. R., Köppel, H., *Conical Intersections* (World Scientific, 2004)

[24] Herzberg, G., Longuet-Higgins, H. C., *Discuss. Faraday Soc.* **35**, 77 (1963)

[25] Yarkony, D. R., *Rev. Mod. Phys.* **68**, 985 (1996)

[26] Hohenberg, P., Kohn, W., *Phys. Rev.* **136**, B864 (1964)

[27] Kohn, W., Sham, L. J., *Phys. Rev.* **140**, A1133 (1965)

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

[29] Born, M., Huang, K., *Dynamical Theory of Crystal Lattices* (Clarendon Press, 1988)

[30] Tully, J. C., *J. Chem. Phys.* **93**, 1061 (1990)

[31] Jones, R. O., *Rev. Mod. Phys.* **87**, 897 (2015), Hafner, J., *J. Comput. Chem.* **29**, 2044 (2008)

[32] Perdew, J. P., Burke, K., Ernzerhof, M., *Phys. Rev. Lett.* **78**, 1396 (1997)

[33] Runge, E., Gross, E. K. U., *Phys. Rev. Lett.* **52**, 997 (1984)

[34] *Phys. Chem. Chem. Phys.* **11**, 4436 (2009)

[35] Capaz, R. B., Ullrich, C. A., *Braz. J. Phys.* **44**, (2014)

[36] Casanova, D., Krylov, A. I., *Phys. Chem. Chem. Phys.* **22**, 4326 (2020)

[37] Park, W., Komarov, K., Lee, S., Choi, C. H., *J. Phys. Chem. Lett.* **14**, 8896 (2023)

[38] Horbatenko, Y., Sadiq, S., Lee, S., Filatov, M., Choi, C. H., *J. Chem. Theory Comput.* **17**, 848 (2021)

[39] Horbatenko, Y., Lee, S., Filatov, M., Choi, C. H., *J. Phys. Chem. A* **123**, 7991 (2019)

[40] Stenrup, M., Larson, Å., *Chem. Phys.* **379**, 6 (2011)

[41] Gromov, E., Trofimov, A., Gatti, F., Köppel, H., *J. Chem. Phys.* **133**, 164309 (2010)

[42] Schalk, O., Galiana, J., Geng, T., Larsson, T. L., Thomas, R. D., Fdez Galván, I., Hansson, T., Vacher, M., *J. Chem. Phys.* **152**, 064301 (2020)

[43] Uenishi, R., Boyer, A., Karashima, S., Humeniuk, A., Suzuki, T., *J. Phys. Chem. Lett.* **15**, 2222 (2024)

[44] Sandoski, A., BelBruno, J. J., *J. Phys. Org. Chem.* **12**, 681 (1999)

[45] Bin, X., Azizi, A., Xu, T., Kirk, S. R., Filatov, M., Jenkins, S., *Int. J. Quantum Chem.* **119**, e25957 (2019)

[46] Kalnin'sh, K. K., *Russ. J. Appl. Chem.* **83**, 858 (2010)

[47] Huix-Rotllant, M., Natarajan, B., Ipatov, A., Wawire, C. M., Deutsch, T., Casida, M. E., *Phys. Chem. Chem. Phys.* **12**, 12811 (2010)

[48] Li, J.-H., Zuehlsdorff, T. J., Payne, M. C., Hine, N. D. M., *Phys. Chem. Chem. Phys.* **17**, 12065 (2015)

[49] Yoshimura, Y., Taya, Y., Matsumura, H., Fujimoto, K., *J. Photopolym. Sci. Technol.* **21**, 525 (2008)

[50] Chen, Z., Liu, Y., Zhou, N., Zhang, Q., Han, K., *J. Anal. Chem.* **66**, 642 (2011)

[51] Huix-Rotllant, M., Schwinn, K., Pomogaev, V., Farmani, M., Ferre, N., Lee, S., Choi, C. H., *J. Chem. Theory Comput.* **19**, 147 (2023)

[52] Li, J.-H., Zuehlsdorff, T. J., Payne, M. C., Hine, N. D. M., *J. Phys. Chem. C* **122**, 11633 (2018)

[53] Müser, M. H., Sukhomlinov, S. V., Pastewka, L., *Adv. Phys.: X* **8**, 2093129 (2023)

[54] LeSar, R., *Introduction to Computational Materials Science: Fundamentals to Applications*, Cambridge University Press, Cambridge, 2013, pp. 62–95

[55] Kývala, L., Dellago, C., *J. Chem. Phys.* **159**, 094105 (2023)

[56] Eastman, P., P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, G. De Fabritiis, and T. E. Markland, *Sci. Data* **10**, 11 (2023).

[57] Eastman, P., B. P. Pritchard, J. D. Chodera, and T. E. Markland, *J. Chem. Theory Comput.* **20**, 8583 (2024).

[58] Butler, P. W. V., Hafizi, R., Day, G. M., *J. Phys. Chem. A* **128**, 945–957 (2024)

[59] Neese, F., *WIRES Comput. Mol. Sci.* **2**, 73 (2012)

[60] Turner, M. A. P., M. D. Horbury, V. G. Stavros, and N. D. M. Hine, *J. Phys. Chem. A* **123**, 873 (2019).

[61] Zuehlsdorff, T. J., P. D. Haynes, M. C. Payne, and N. D. M. Hine, *J. Chem. Phys.* **146**, 124504 (2017).