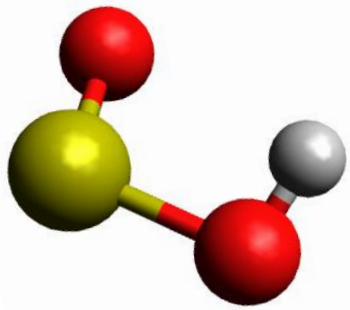
A Combined QM/ML Approach to Accelerate Photodynamics Simulation

Researcher: Pin-Han Chen
Supervisor: Dr. Daniel Roca-Sanjuán
Co-supervisosor: Lambert Schomaker

Photodynamics with Quantum Mechanics

Photodynamics trajectory over time



Coordinate: r_1

Velocity: V_1

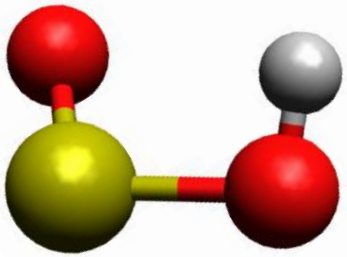
State: ST_1

Energy: E_1

Forces: F_1

1. Calculate next coordinate r_2 with velocity verlet

2. Predict surface hopping with overlap matrix or non-adiabatic coupling vector (NAC)



Coordinate: r_2

Velocity: V_2

State: ST_2

Energy: E_2

Forces: F_2

Machine Learning Photodynamics

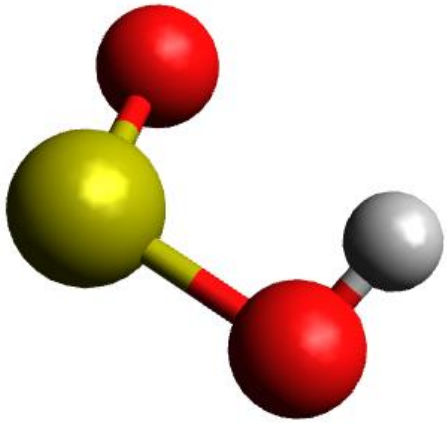
- QM calculation is expensive ➔ accelerate with ML model
- ML model trained on QM dataset
- Simulate PD trajectories with trained ML model
- Input: molecular coordinates (descriptor)
- Output: energies, forces, overlap, NAC

Local descriptor:
Describe the whole molecule

Global descriptor:
Describe atoms and the interaction between atoms

Reference QM Method

- Molecule: Hydroxysulfinyl radical (HOSO)
- Method: MS-CASPT2 (13,10) / ANO-S-VDZP (OpenMolcas / SHARC)
- Time: 100 ~ 520 fs (0.5 fs / step)
- Hopping: FSSH + overlap matrix



(a) 5 doublets starting from first excited state

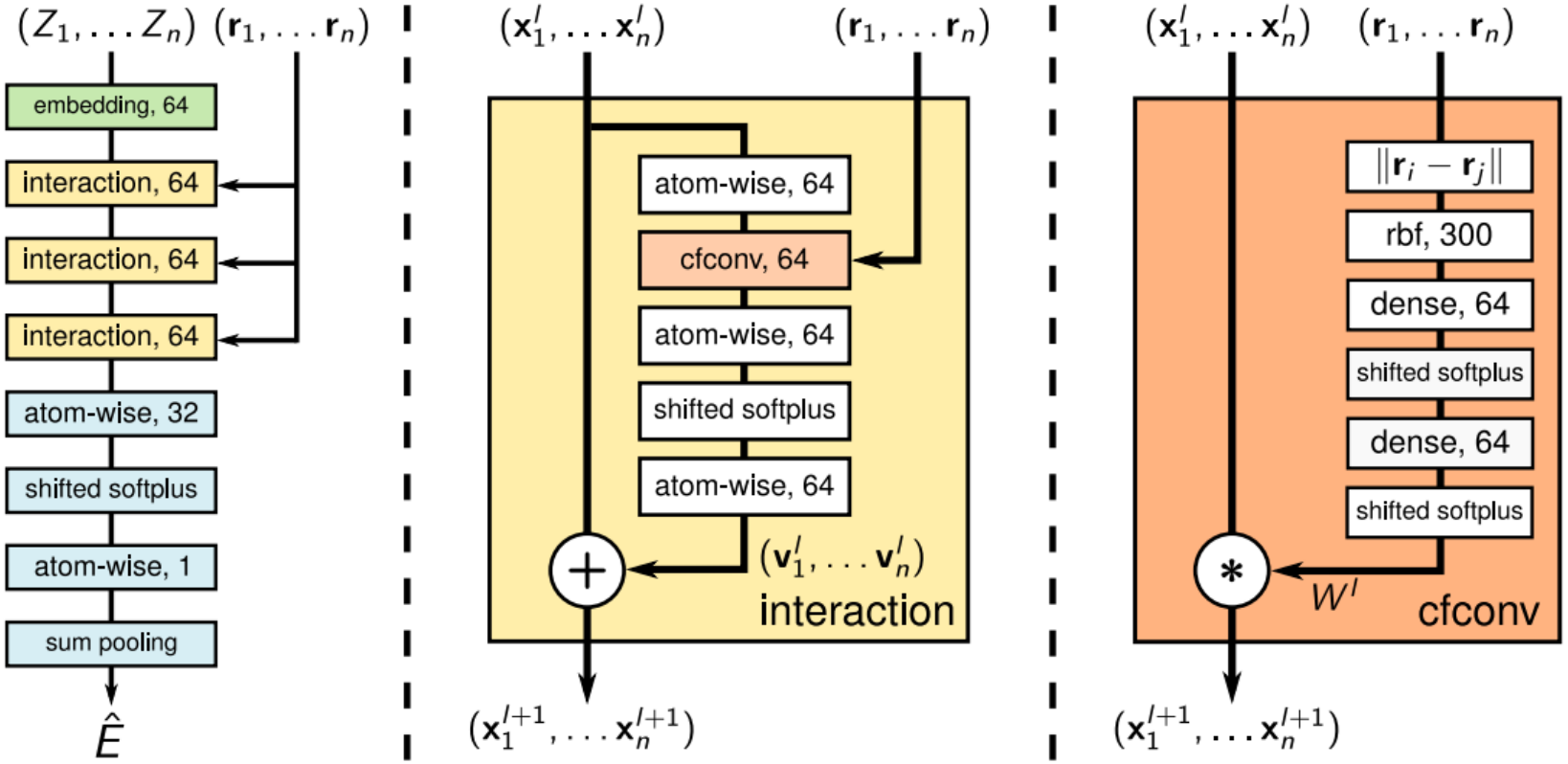
(b) 8 doublets starting from different states

Data provided by
Javier Carmona-García

J. Carmona-García, et. al, *J. Am. Chem. Soc.* **143**, 10836−10841, 2021.

Molecular Descriptor

SchNet: convolutional network taking interatomic distance

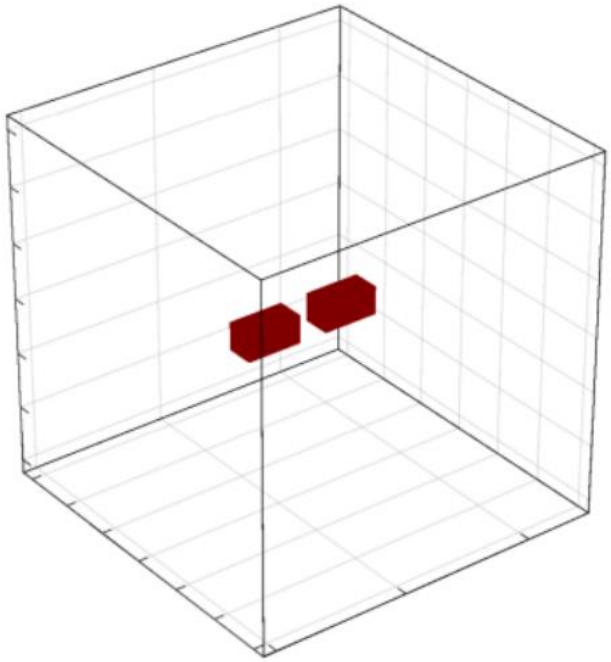


Kristof T. Schüttm et. al. *Advances in Neural Information Processing Systems*. **30** 992-1002, 2017.
J. Westermayr, et. al.*The Journal of Physical Chemistry Letters*, **11**(10):3828–3834, 2020.

Molecular Descriptor

Localized Orbital Locator (LOL):

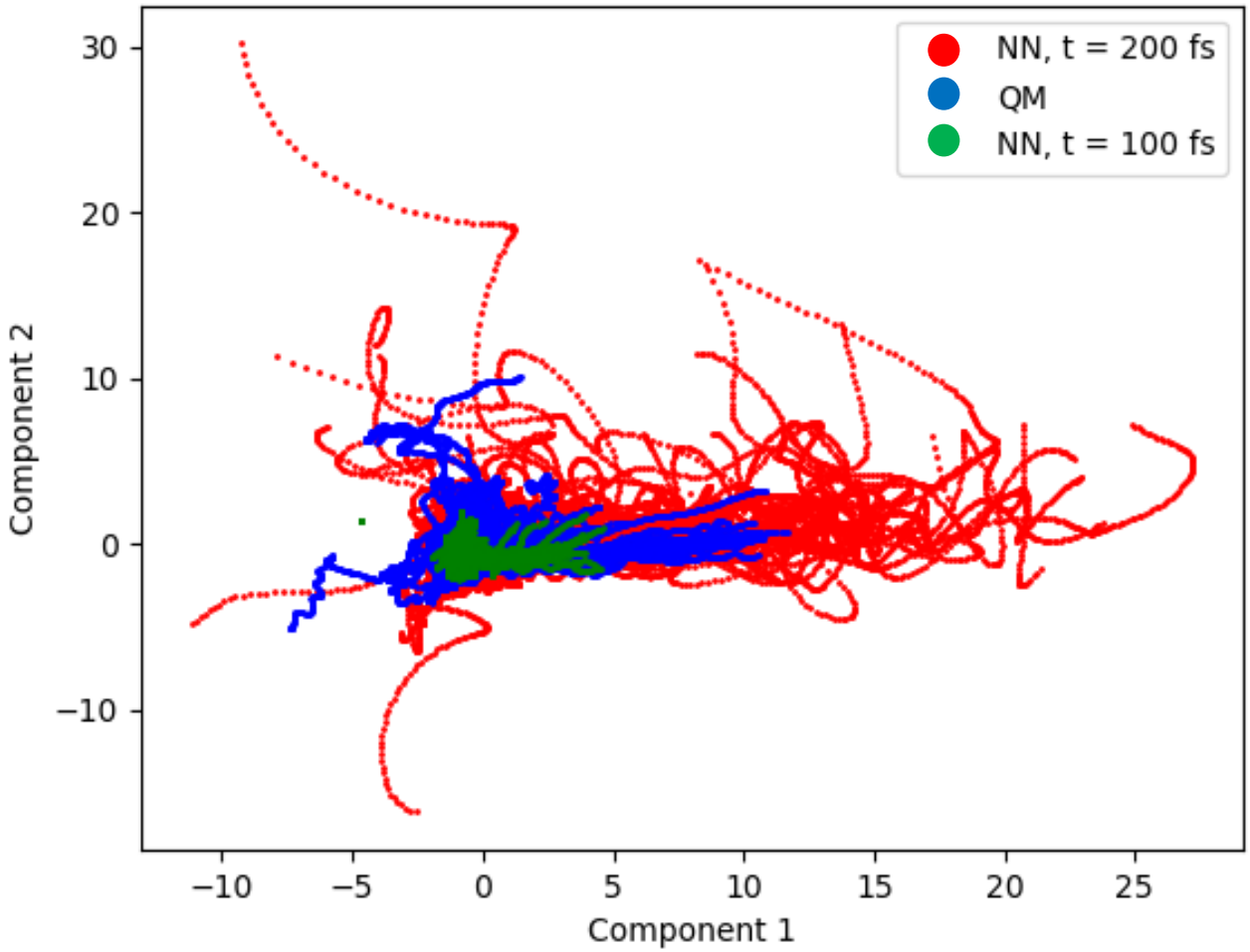
3D voxelized descriptor describing electron density



- The space is represented as a 14*14*14 grid
- Each voxel contains features describing electron density
- DenseNet with 3D convolution layers
- DFT (UB3LYP/6-31G) is required → ground state energy utilized for Δ-learning

Ankur Kumar Gupta and Krishnan Raghavachari, *Journal of Chemical Theory and Computation,* **18**(4), 2132-2143, 2022.

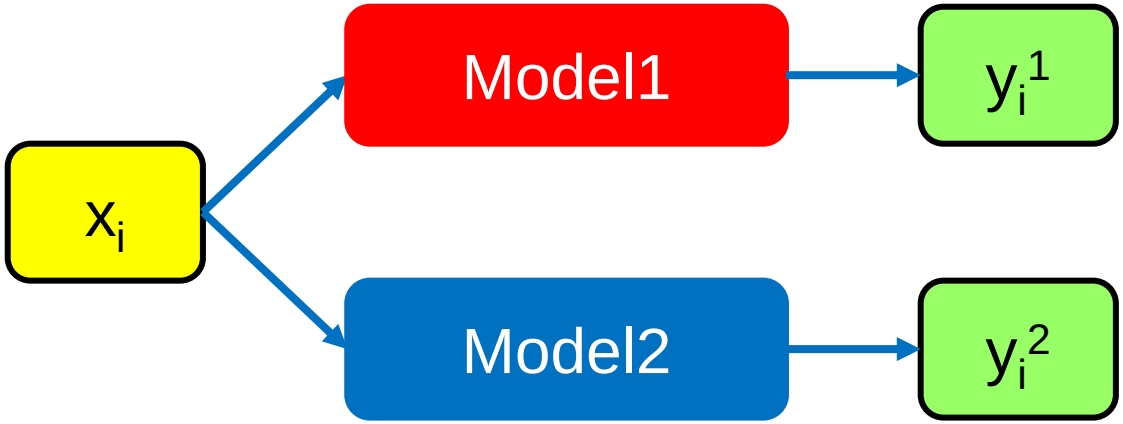
Applicable Domain



PCA of coordinates for QM and ML trajectories

Last 100 fs exceeds the training set domain

Active Learning

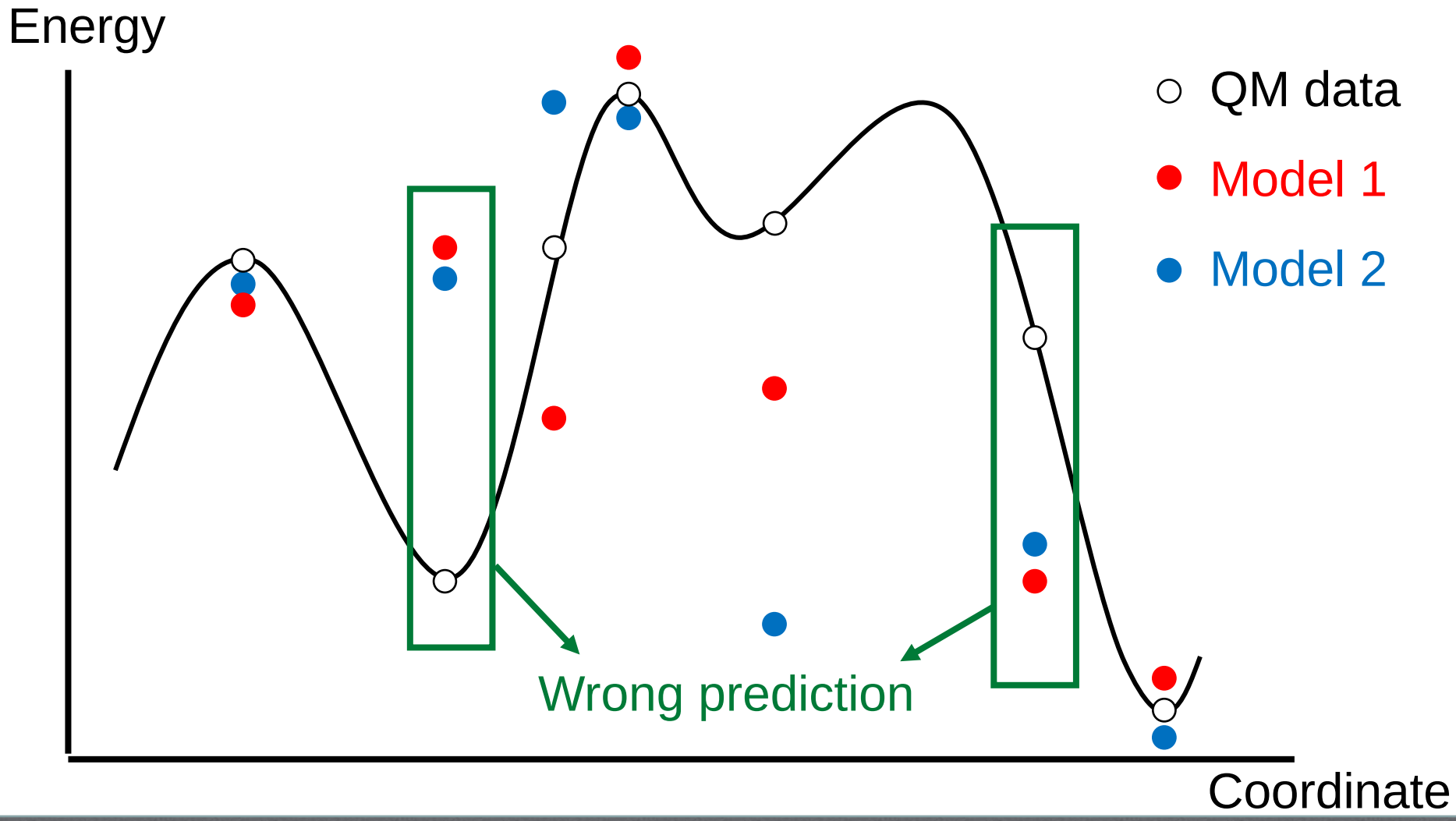


Previous work by
J. Westermayr, et. al.

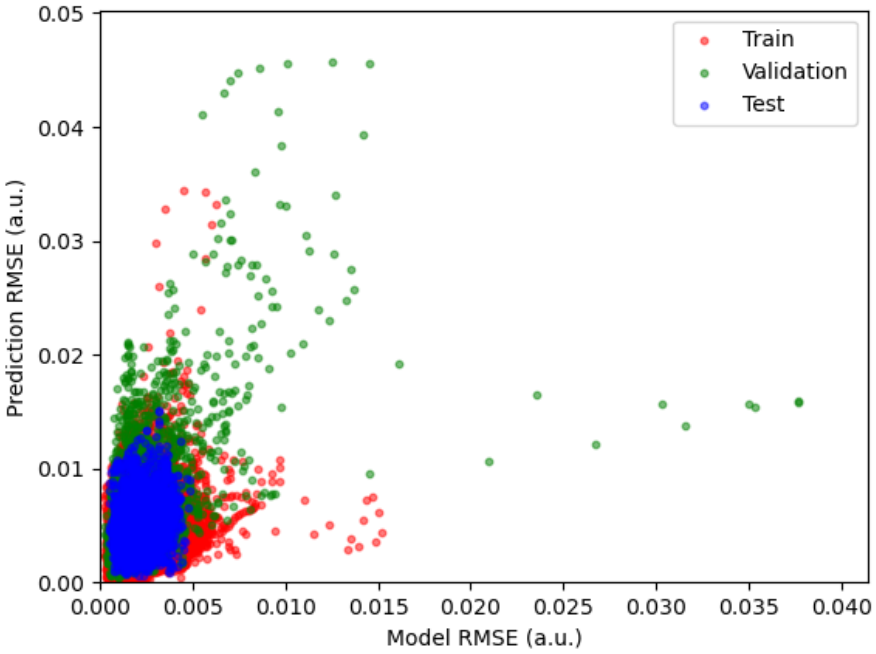
- 2 distinct models trained on the same dataset
- x_i is predicted by 2 models respectively
- RMSE(y_i^1 - y_i^2) < threshold: x_i is in applicable domain
- Reliability of this method was never discussed

J. Westermayr, et. al.*The Journal of Physical Chemistry Letters*, **11**(10):3828–3834, 2020.

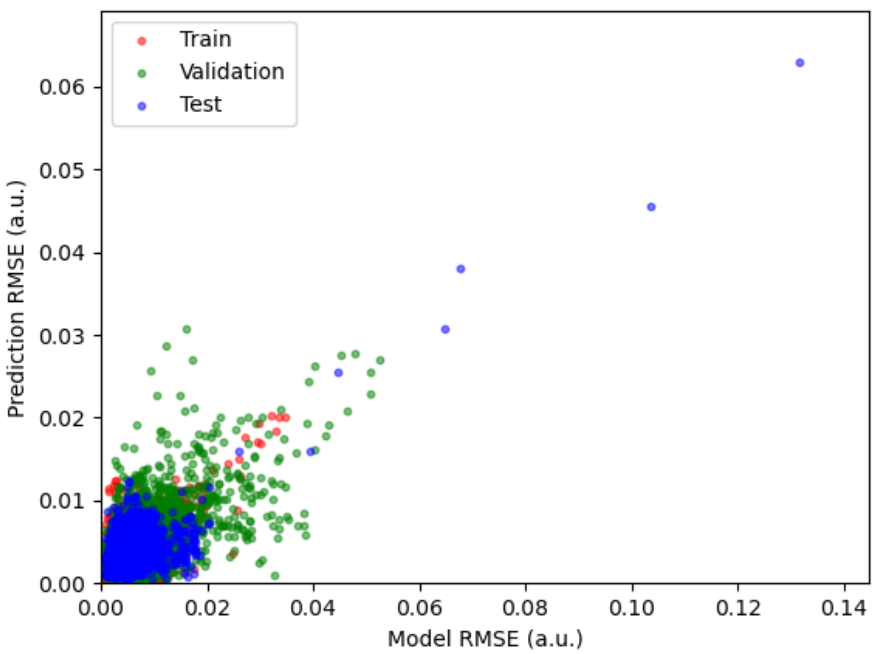
Active Learning



Active Learning



SchNet1 + SchNet2

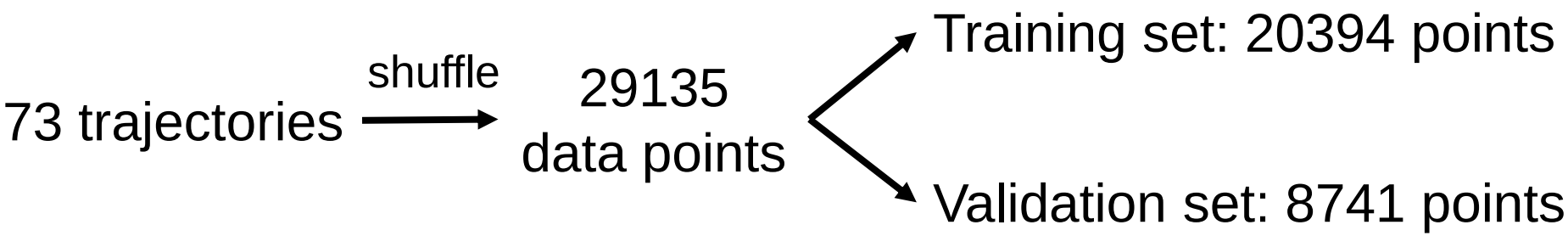


LOL + SchNet

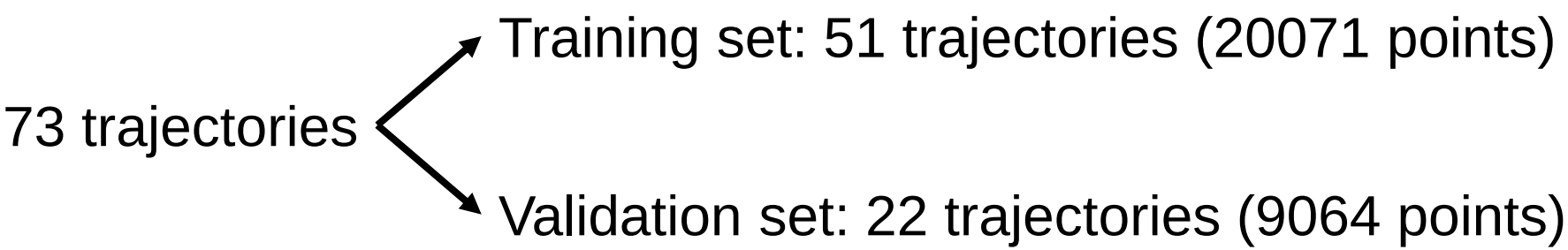
The error between LOL and SchNet better represents the real error

Data Processing

Uniform dataset:



Non-uniform dataset:



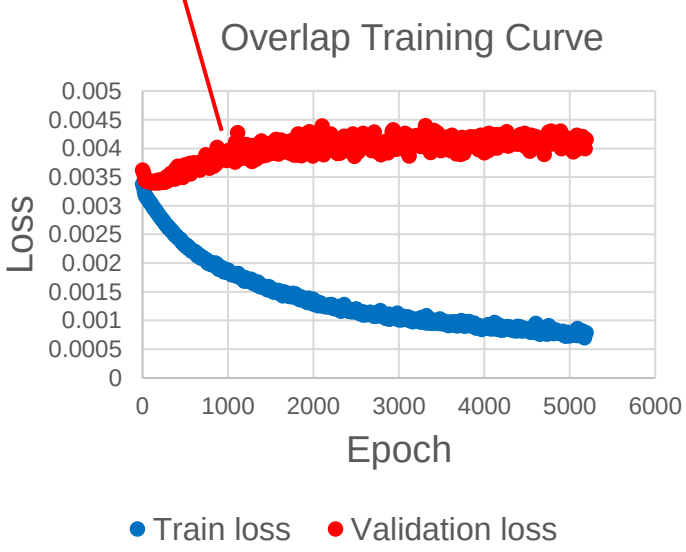
Training Results: Energy and Force

Validation (Test) RMSE		Energy (eV)	Force (eV / Å)
LOL	Uniform	0.103 (0.117)	1.35 (1.32)
	Non-uniform	0.158 (0.189)	1.36 (1.43)
SchNet	Uniform	0.105 (0.105)	1.34 (1.19)
	Non-uniform	0.163 (0.183)	1.26 (1.29)
LOL+SchNet	Uniform	0.073 (0.082)	1.23 (1.07)
	Non-uniform	0.125 (0.147)	1.13 (1.18)

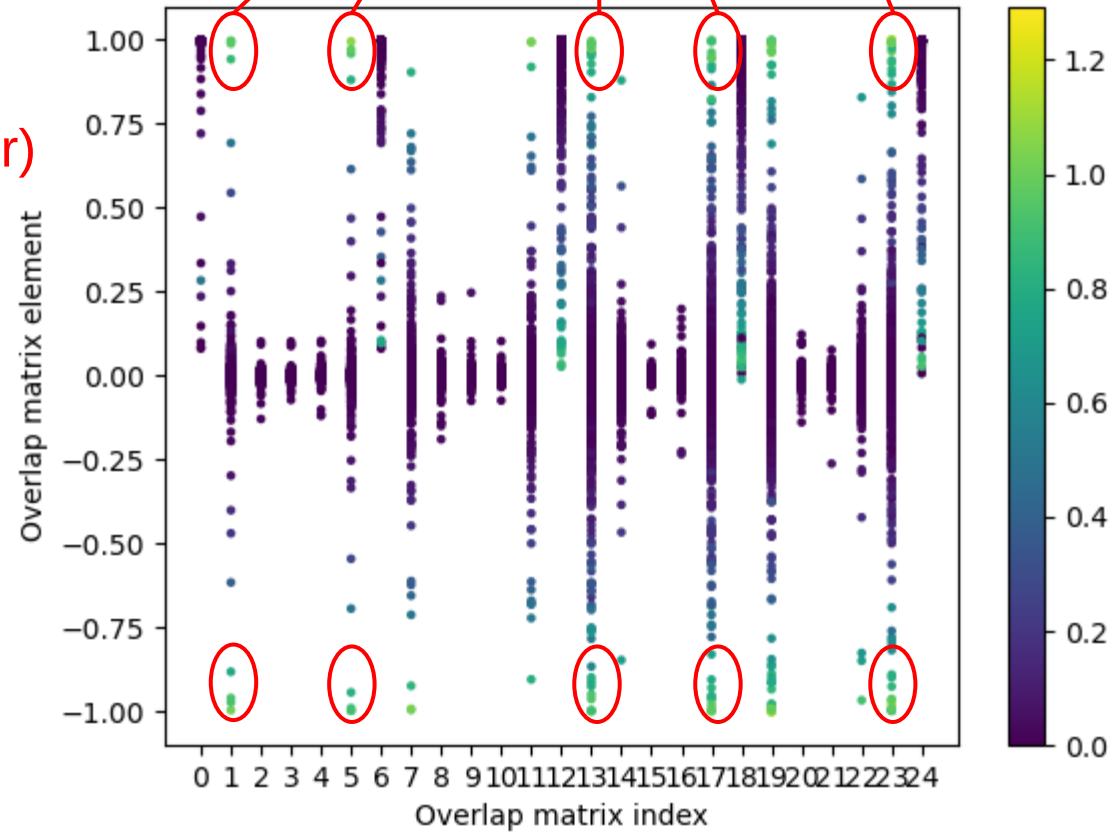
Training Results: Overlap Matrix

Lack of data in the CI region limits the performance of ML

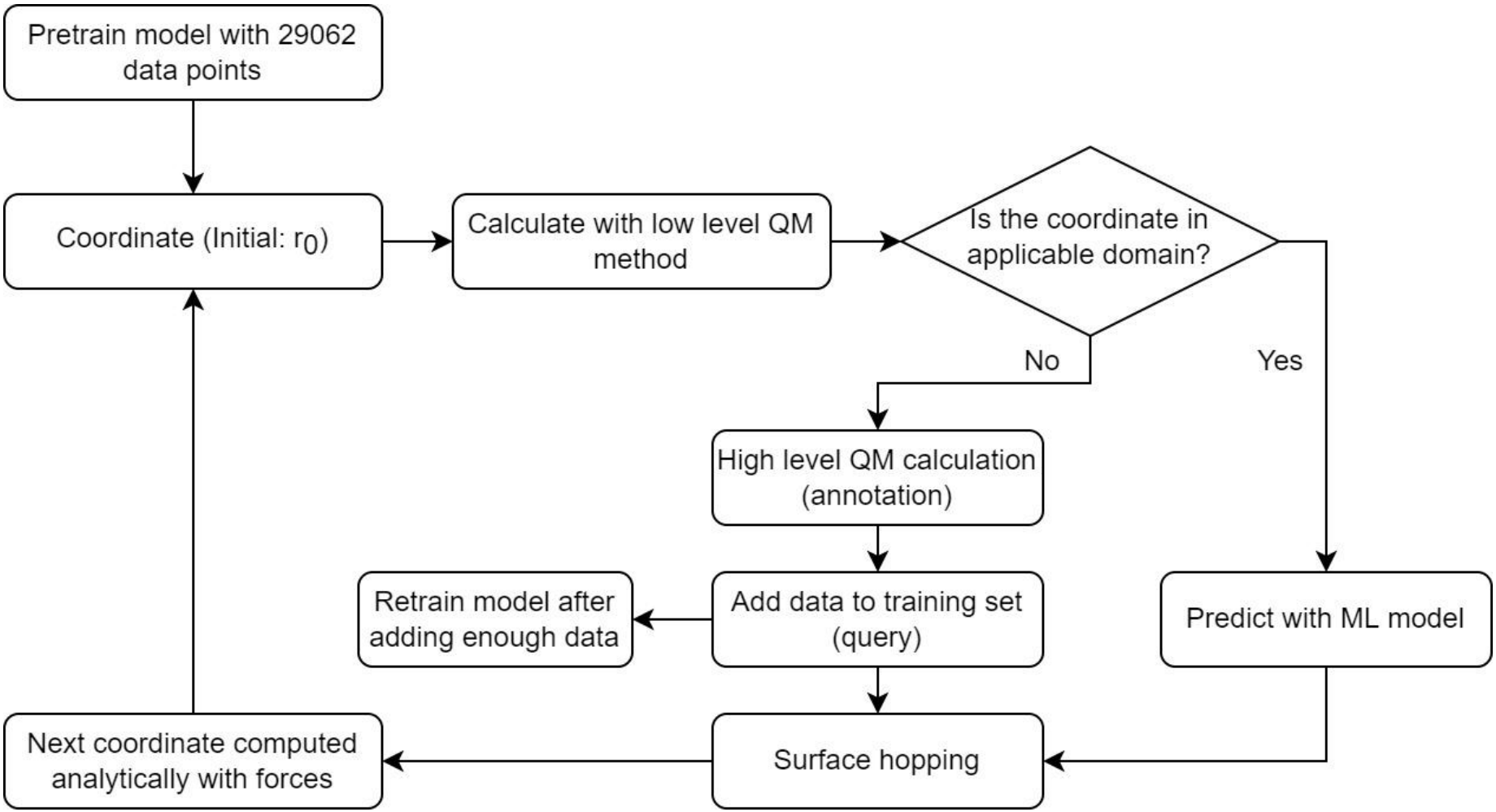
Overfitting
(Validation error >> Train error)



Off-diagonal elements with high absolute values are inaccurate



PD Simulation Workflow



C.Y. Zhu, K. Nobusada, H. Nakamura, *J. Chem. Phys.* **115**, 3031 (2001)

PD Results – Photoproducts

(a) 5 doublets starting from first excited state for 200 fs

Photoproduct	ML (N = 73)		ML (N = 1000)		QM/ML (N = 73)		QM (N = 73)	
	N°	%	N°	%	N°	%	N°	%
HO+SO	41	56.2	561	56.1	29	39.7	32	43.8
HOS+O	0	0	2	0.2	0	0	2	2.7
$H+SO_2$	0	0	10	1.0	0	0	0	0
No photolysis	32	43.8	427	42.7	44	60.3	39	53.4

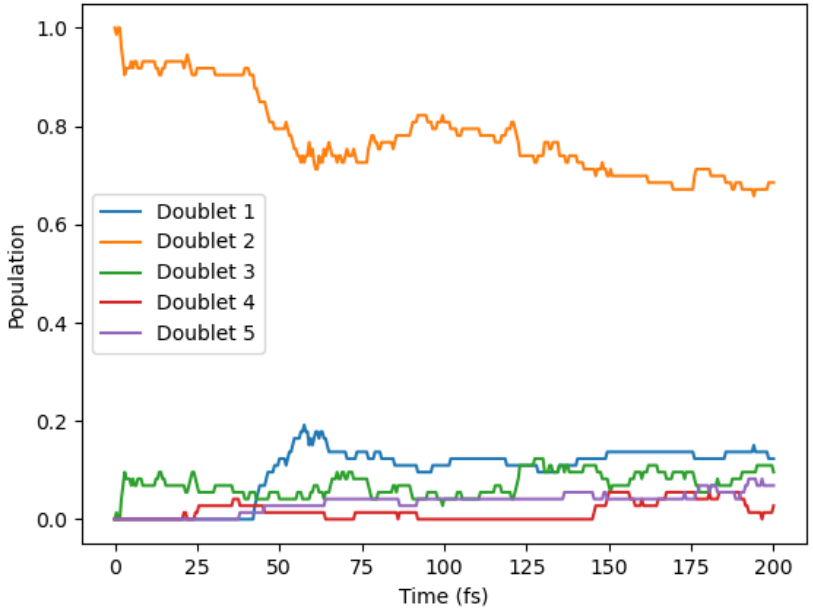
(b) 8 doublets starting from different states

Photoproduct	ML (N = 73, 200 fs)		QM (N = 73, ~353 fs)	
	N°	%	N°	%
HO+SO	59	84.1	55	79.7
HOS+O	2	2.9	6	8.7
$H+SO_2$	1	2.9	2	2.9
No photolysis	7	10.1	6	8.7

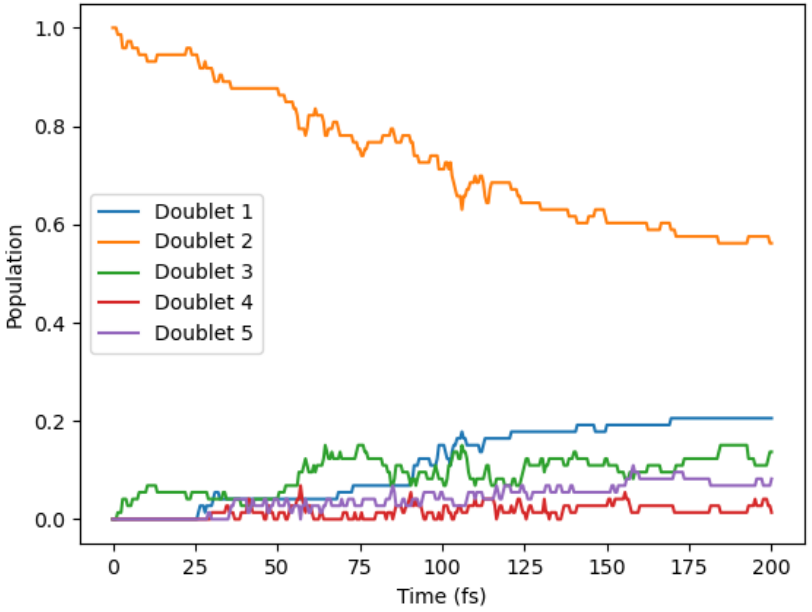
PD Results – State Population

(a) 5 doublets starting from first excited state for 200 fs

QM + FSSH (N = 73)



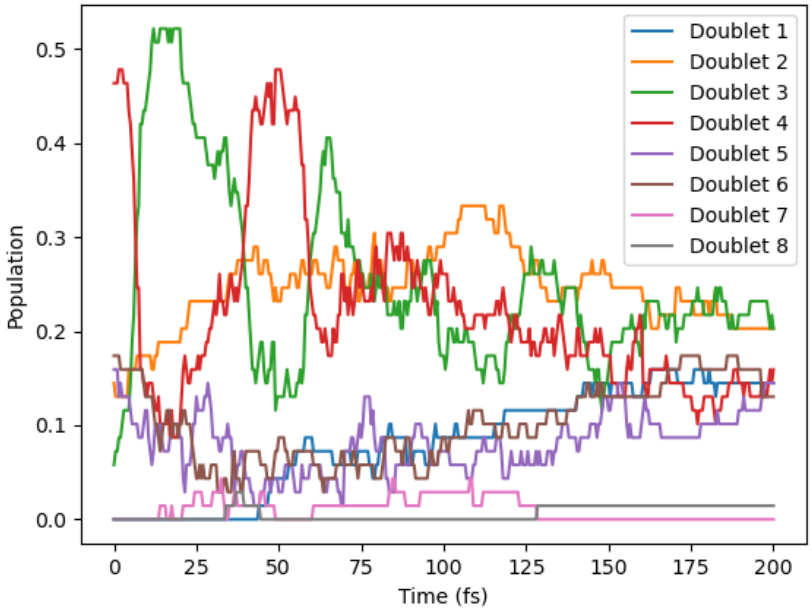
QM/ML + ZNSH (N = 73)



PD Results – State Population

(b) 8 doublets starting from different states for 200 fs

QM + FSSH (N = 73)



QM/ML + ZNSH (N = 73)