

LMG 1 – Machine Learning Tools to Solve First-Principles Challenges

Krystof Brezina¹, Burak Gurlek¹, Paolo Lazzaroni¹, Zekun Lou¹, Alexandre Reily Rocha^{1, 2}, Shubham Sharma¹, Elia Stocco¹ and Mariana Rossi¹

1) Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany 2) Instituto de Física Teórica – IFT, Universidade Estadual Paulista (UNESP)

Advancing Interatomic Potentials

Reliable Vibrational Properties

Led by Shubham Sharma & Burak Gurlek Accurately modeling vibrational dynamics of molecular crystals is challenging due to the importance of anharmonic contributions. Reliability of vibrational properties derived from machine-learning potentials (MLIPs) is only quantifiable through error propagation. We propose an active-learning strategy to train a general committee model for polyacene crystals [1]. We show error propagation to harmonic and anharmonic vibrational density of states (VDOS), quantifying peak-position uncertainty.

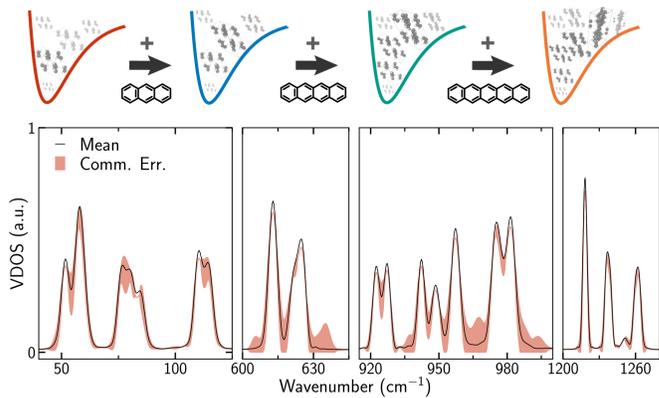


Figure 1: Top: Sketch showing the active-learning scheme used to create general MLIP for naphthalene, anthracene, tetracene and pentacene crystals. Bottom: Anharmonic VDOS of naphthalene crystal at 80 K, calculated from molecular dynamics using the committee model. The red shaded region represents the propagated uncertainty on different frequency ranges of the VDOS. Around 600 cm⁻¹ and in the 900–1000 cm⁻¹ range (carbon rings deformation modes), the uncertainty indicates variability in predicted peak positions.

Polarization-Oriented Vibrational Raman

Led by Paolo Lazzaroni Polarization-Oriented (PO) Raman spectroscopy is a powerful tool to probe the evolution of phonon symmetry with temperature. Relying on the MLIPs for polyacenes and different strategies to compute and predict polarizability tensors, we simulate PO-Raman spectra of large unit cells at different temperatures. The temperature-dependence of the intensity at different angles is sensitive to vibrational coupling. With $\alpha(\theta) := \mathbf{e}_i \cdot \boldsymbol{\alpha} \cdot \mathbf{e}_s$,

$$I(\omega, T, \theta) \propto \mathcal{F} \left\{ \left\langle \alpha(\theta, 0) \alpha(\theta, t) \right\rangle_T \right\}.$$

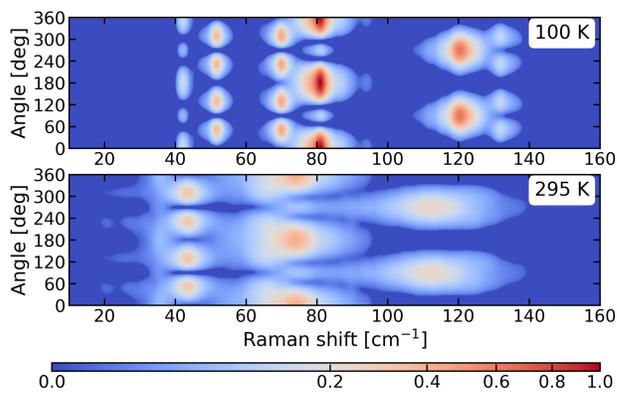


Figure 2: Temperature dependence of the PO-Raman spectrum of anthracene in the intermolecular motion range of 10 to 160 cm⁻¹, obtained with a combination of machine-learning potentials for the dynamics and Raman tensor weighted Γ -VDOS for the polarizabilities. The intensity is normalized to the maximum intensity at 100 K.

Quantification of H-bond Strengths

Led by Alexandre Reily Rocha Taking advantage of MLIPs trained on several different exchange-correlation density functionals, rigorous convergence of thermodynamic-integration paths can be achieved. We probe the strength of hydrogen bonds in crystals at different temperatures and pressures. We take Ice Ih as an example and a path with three steps. The hydrogen bond between a water molecule and the ice matrix is switched on, in between the switching on/off of a harmonic potential restraint on the molecule:

$$V = \lambda_1 V_{ice} + (1 - \lambda_1) (V_{ice+defect} + V_{H_2O}) + \lambda_2 U_{harm}$$

Nuclear quantum effects require additional thermodynamic integration steps based on path integral molecular dynamics.

Reliable Vibrational Prop. (cont.)

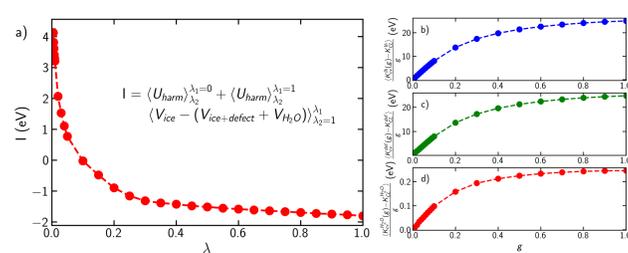


Figure 3: Panel a): Integrand for the classical thermodynamic integration step. Further panels: Integrands of steps to include nuclear quantum effects on b) the full system; c) ice with a molecular vacancy; d) a single water molecule. With a MLIP trained on the optB88 functional, $\Delta F_{\text{classical}} = 277 \pm 7$ meV and $\Delta F_{\text{QM}} = 293 \pm 10$ meV.

Reliable Chemical Reactions

Led by Krystof Brezina Obtaining MLIPs for chemical reactions is challenging because of the need for an expensive sampling of thermally inaccessible structures around the transition state. To reach this goal, we combine

- Transition tube sampling [2] to cheaply generate such structures and
- Active learning via query by committee [3] to select the most relevant ones.

This methodology allows us to quantify the free-energy profiles and thermal rates of water splitting on Pd(111) using path-integral umbrella sampling (Fig. 4), shedding light on the crucial role of nuclear quantum effects and the hydrogen-bond network in the neighborhood of the reaction site.

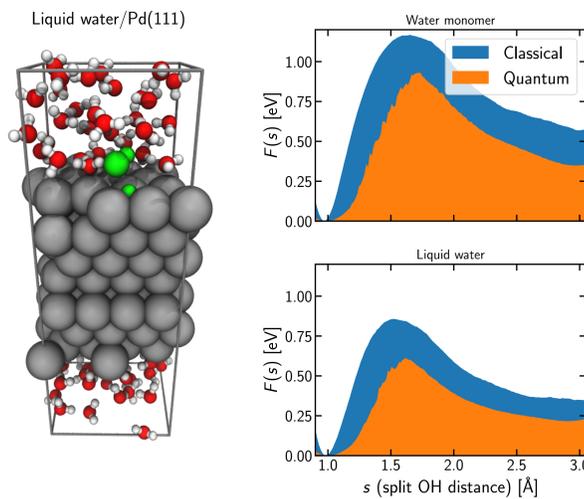


Figure 4: Surface-mediated water splitting on Pd(111). Left: A snapshot of the simulated system consisting of a Pd(111) slab in contact with liquid water under periodic boundary conditions. The split water molecule is highlighted in green. Top right: Free energy as a function of the distance of the split O-H bond obtained from classical and path-integral umbrella sampling simulations of a single water molecule on a Pd surface. Bottom right: The same as above, just for the full liquid slab.

New Electronic-Structure Models

Multi-Valued Dipolar Models

Led by Elia Stocco Simulations with external static or time-dependent electric fields require the knowledge of response tensors. For this purpose, we developed a MACE [4] model to predict the multi-valued dipole $\boldsymbol{\mu}$ in periodic systems, able to capture the correct behavior even far from equilibrium conditions [5]:

$$\boldsymbol{\mu}(\mathbf{R}) = \sum_l^{\text{atoms}} \underbrace{\boldsymbol{\mu}^l(\mathbf{R})}_{\text{machine learned}} + e \sum_l^{\text{atoms}} \underbrace{\mathcal{N}^l \mathbf{R}^l}_{\text{oxidation numbers}}$$

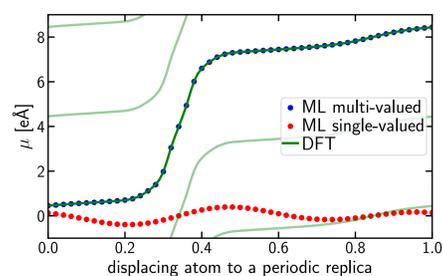


Figure 5: Predicted dipoles from single- and multi-valued models along a trajectory where an atom is displaced to a periodic replica. The DFT reference (green) is shown for different branches. The multi-valued model matches DFT, while the single-valued model fails for structures far from equilibrium.

Multi-Valued Dipolar Models (cont.)

Simulations with Applied Electric-Fields

By defining $H = H_0 + \boldsymbol{\mu} \cdot \mathbf{E}$, we can combine ML models to run classical and path-integral molecular dynamics at varying electric field intensities [5].

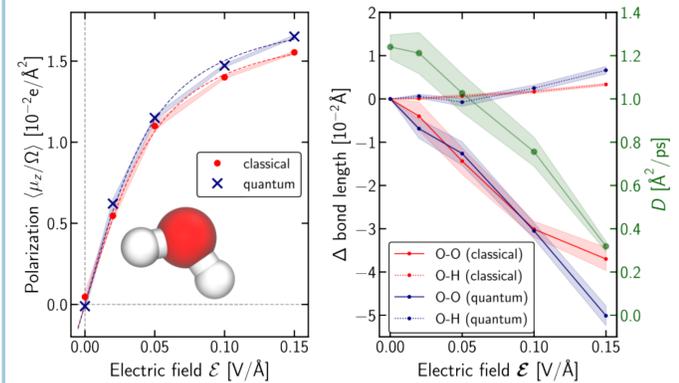


Figure 6: Comparison between classical and quantum results in liquid water at varying electric field intensities. Nuclear quantum effects induce a more pronounced polarization at high electric fields (left panel). The intramolecular O-H bond elongates, and the intermolecular O-O distance reduces at increasing fields (right panel). Quantum simulations show an enhancement of these effects at higher field intensities. The molecular self-diffusion (green curve) drops substantially, showing the onset of electrofreezing.

Models for the Electronic Density

Led by Zekun Lou By “density-fitting” the electronic density n_e of isolated and periodic systems, one can machine-learn the expansion coefficients c_λ with symmetry-adapted Gaussian process regression [6, 7], based on equivariant similarity kernels \mathbf{K}_λ :

$$c_\lambda = \mathbf{w}_\lambda^T \mathbf{K}_\lambda.$$

Numerical improvements of this model enable extrapolation from density learning on small bilayer 2D materials to the prediction of derived properties of small-angle, large twisted bilayer systems.

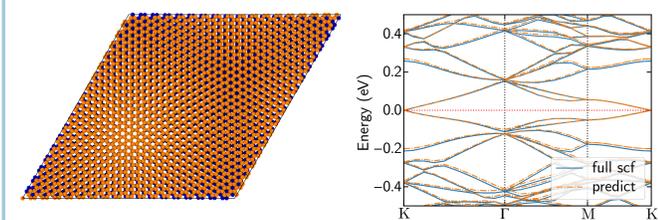


Figure 7: Left: Twisted bilayer graphene at $\theta = 1.89^\circ$, 3676 atoms. Right: band structure derived from electronic density prediction, matching full SCF results.

Led by Mariana Rossi Following the same density-fitting ansatz for the electronic density response to homogeneous electric fields, we proposed a fully equivariant kernel to predict the vector field of the density response [8].

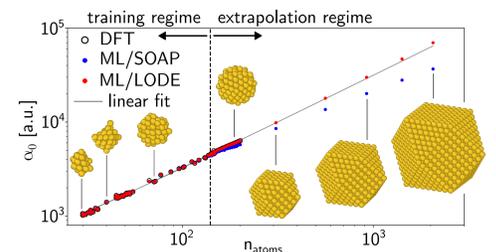


Figure 8: Isotropic polarizability α_0 of gold nanoparticles of increasing size, derived from density-response predictions, trained on density-functional perturbation theory. Long-range (LODE) descriptors are essential to capture the scaling law of α_0 with size.

Outlook

- ML for reactions at electrified interfaces
- ML for non-equilibrium quantum dynamics
- General and autodifferentiable models for the density

References

- [1] B. Gurlek *et al.*, arXiv:2504.11224 (2025).
- [2] K. Brezina *et al.*, *J. Chem. Theory Comput.* **19**, 6589 (2023).
- [3] C. Schran *et al.*, *J. Chem. Phys.* **153**, 104105 (2020).
- [4] I. Batatia *et al.*, *Adv. Neural Inf. Process. Syst.* (2022).
- [5] E. Stocco *et al.*, arXiv:2502.02413 (2025).
- [6] A. M. Lewis *et al.*, *J. Chem. Theory Comput.* **17**, 7203 (2021).
- [7] A. Grisafi *et al.*, *J. Chem. Theory Comput.* **19**, 4451 (2023).
- [8] M. Rossi *et al.*, *J. Phys. Chem. Lett.* **16**, 2326 (2025).