

Fidelity of Machine-Learned Potentials: Quantitative Assessment for Protonated Oxalate

Chen Qu^{1,*}, Paul L. Houston², Qi Yu³, Apurba Nandi⁴, Joel M. Bowman^{4,†},
Valerii Andreichev⁵, Silvan Käser⁵, and Markus Meuwly^{5,‡}

¹Independent Researcher, Toronto, Ontario M9B0E3, Canada

²Department of Chemistry and Chemical Biology, [Cornell University](#), Ithaca, New York 14853, USA

³Department of Chemistry, [Fudan University](#), Shanghai 200438, People's Republic of China

⁴Department of Chemistry and Cherry L. Emerson Center for Scientific Computation,
[Emory University](#), Atlanta, Georgia 30322, USA

⁵Department of Chemistry, [University of Basel](#), Basel, Switzerland



(Received 20 April 2026; accepted 2 September 2026; published 21 October 2026)

There has been a veritable explosion of methods and software to perform machine-learned regression on datasets of electronic energies and forces to develop high-dimensional machine-learned potential energy surfaces (ML-PESs). A major, but not deeply studied aspect is how well different ML-PESs represent the same dataset on which they were trained, beyond the standard statistical precision measures. Here, this is examined using several “stress tests,” for two widely used machine-learned potential approaches. One is based on permutationally invariant polynomial linear least-square regression and the other is the message-passing neural network PhysNet approach. The energies from the two PESs are directly compared as are the IR spectra determined from vibrational self-consistent field/vibrational configuration interaction calculations using fitted dipole moment surfaces. In addition, tunneling splittings for the hydrogen transfer between two equivalent structures are reported using three different methods: ring polymer instanton theory, diffusion Monte Carlo simulations, and the Q_{im} -path method. These calculations require the evaluation of on the order of one billion energies that are widely dispersed in the 15-dimensional configurational space. The two PESs yield results for these quantities in excellent agreement with each other.

DOI: [10.1103/jfjx-djk2](#)

I. INTRODUCTION

Reliable, quantitatively accurate potential energy surfaces (PESs) are key for investigating, characterizing, and understanding chemical and biological systems in the gas and condensed phases based on molecular simulations [1–4]. Advances in machine learning (ML) methodologies have led to the development of machine-learned potential energy surfaces (ML-PESs), which are now widely used to simulate such systems and have become part of the standard toolbox for studying physical and chemical processes. Recent applications include the simulation of a water box with ≈ 12 M atoms and a solid-state system of Cu with 100 M atoms [5] using the DeepMD model [6]. Similarly, ML-PESs have been used in atomistic simulations of biological systems [7,8]. One of the largest such simulations was that of the human immunodeficiency virus capsid with 44 M atoms using the Allegro model

[9,10]. On the other hand, such impressive system sizes are still small compared to what can be handled using empirical energy functions (billions of atoms for several million time steps) [11–16]. Depending on the level of theory used for the reference data, the use of ML-PESs promises to reach new levels of accuracy and realism compared with all-atom simulations based on empirical and even physics-enhanced energy functions [17].

Within the broader physical chemistry community, ML-PESs together with established molecular dynamics (MD) suites and methods have become a new standard approach [18–24]. Likewise, performance and stability challenges for ML-PESs have been established that explicitly target biomolecular systems [25,26]. Such studies are required to pave the way toward confident and competent use of these new technologies by the broader science community as had been the case with empirical energy functions and electronic structure methods.

The profound and lasting impression ML-PESs make on the field of computer-based approaches applied to the dynamics and reactivity of chemical, biological, and materials systems in the gas and condensed phase is also reflected in the large number of computer codes that have become available over the past decade. A recent review article presents a sweeping overview of the status of the field up to 2025 and lists more than 90 software packages [27]. Two among this list are used here to address a major, but not deeply studied,

*Contact author: szquchen@gmail.com

†Contact author: jmbowma@emory.edu

‡Contact author: m.meuwly@unibas.ch

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

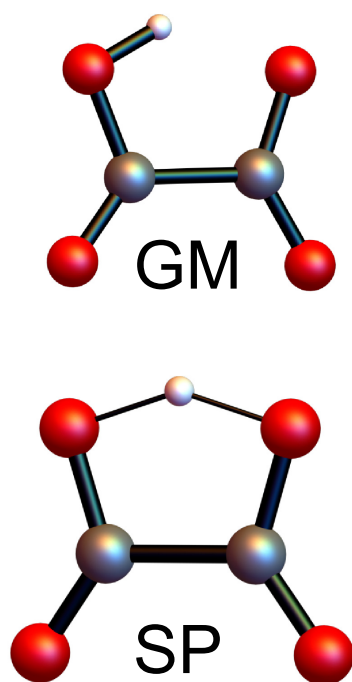


FIG. 1. Structures of protonated oxalate anion, OxH, at the global minimum (GM) and proton-transfer saddle point (SP).

aspect of ML-PESs. Namely, how well different approaches represent the same dataset on which they are trained going well beyond statistical precision metrics [25,28].

The present work reports results on several “stress tests” beyond fitting precision for a permutationally invariant polynomial PES, based on standard linear regression and a PhysNet neural network (NN)-based potential [29]. The PIP and related methods using PIPs have been reviewed in a recent Perspective, where it was noted that more than 200 ML-PESs have been reported using PIPs [30]. PhysNet belongs to the family of message passing neural networks and was one of the early architectures developed in that class of NN-based approaches [29]. Specifically, the features in PhysNet are continuously refined and together with a model for the total energy PhysNet also provides geometry-dependent partial charges. The implementation and broader scope of PhysNet have also been described in the literature [21,31,32]. Importantly, the present work assesses the quality of the ML-PESs with a particular focus on comparisons with experiment for a recently characterized system for which high-level (HL) reference calculations are feasible. The spirit of the present work readily extends and generalizes to other systems for which rigorous and validated experimental results are available. However, no survey extending to questions related to chemical space is attempted here.

The present work examines the performance of the PIP and PhysNet ML-PESs for a number of vibrational properties of the protonated oxalate anion ($[\text{HO}_2\text{CCO}_2]^-$), hereafter referred to as OxH (see Fig. 1), with a major focus on normal modes, the infrared (IR) spectrum, and the ground-state tunneling splitting. The splitting was determined using ring polymer instanton (RPI) theory, unbiased, fixed-node diffusion Monte Carlo (DMC) simulations, and the one-

dimensional (1D) Q_{im} path. The remainder of this article is organized as follows. Brief descriptions of the theoretical methods are given, followed by results, discussion, a summary, and conclusions.

II. METHODS

In the following, the datasets and construction of the ML-PESs are described. Methods for the spectroscopic studies [33–36], diffusion Monte Carlo calculations [37–40], and computing tunneling splittings [41–46] are provided in the Supplemental Material [47].

A. Reference data for the PESs

For the present work, two reference datasets were used. Set2025-LL (LL for “low level”; MP2/aug-cc-pVTZ) comprised 22100 OxH structures for which energies, forces, and dipole moments were determined at the MP2/aug-cc-pVTZ level of theory. To improve the performance of the LL model, transfer learning (TL) to an HL dataset determined at the Coupled Cluster with Single, Double, and perturbative Triple excitations [CCSD(T)/aug-cc-pVTZ] level of theory was carried out [43,48–51]. This high-level dataset, Set2025-HL, comprised a total of 2688 structures selected semirandomly from Set2025-LL: 1067 structures were sampled using MD simulations at 1000 K, and 960 and 661 structures were obtained from normal mode sampling around the minimum and transition state (TS), respectively, for which energies, forces, and dipole moments were determined at the CCSD(T)/aug-cc-pVTZ level of theory using molpro [52]. Set2025-HL is the dataset from which PES2025 was trained, *vide infra* [53]. A detailed description of the structural sampling and dataset generation procedures can be found in previous work [53,54].

Set2026-HL, which is based on Set2025-HL, was specifically introduced to validate the computed H -tunneling splittings. This set consisted of Set2025-HL, which was augmented in a targeted fashion as follows. (1) 200 points were extracted along the minimum energy path (MEP), which includes 100 points sampled in each direction from the TS toward the global minimum, and (2) 200 points were generated along the instanton path (IP) generated using PES2025, which was available from previous work [53]. For all new structures, energies, forces, and dipole moments were computed at the CCSD(T)/aug-cc-pVTZ level of theory using molpro [52]. This yields Set2026-HL that is the combination of Set2025-HL and the 400 structures generated as described in (1) and (2) above. Since Set2026-HL is entirely based on Set2025-HL together with improved coverage around the MEP and IP, normal modes and infrared spectra are expected to be very close.

B. PIPs and Δ -machine learning

PIPs provide a direct representation of a PES that respects permutational symmetry. In this representation, the potential is given by the equation

$$V(\mathbf{y}) = \sum_{i=1}^{n_p} c_i p_i(\mathbf{y}), \quad (1)$$

where c_i are linear coefficients, p_i are PIPs, n_p is the total number of polynomials for a given maximum polynomial order, and $\mathbf{y}$ are the variables of the potential and the PIPs. In this work, $\mathbf{y}$ are Morse variables, given by $y_{ij} = e^{-r_{ij}/\alpha}$, where r_{ij} are the $N(N-1)/2$ internuclear distances among N atoms.

The general theory underlying this representation was given by Braams and Bowman [3]. Two approaches were outlined to obtain PIPs. One is given in terms of so-called primary and secondary invariants and the second is based on “monomial symmetrization.” The second approach can be illustrated for the simple case of a triatomic. In that case, the potential is given by

$$V = \sum_{m=0}^M D_m \mathcal{S}[y_{12}^a y_{13}^b y_{23}^c], \quad (2)$$

where “ $\mathcal{S}$ ” is the operator that symmetrizes monomials. This yields the symmetrized basis function for the representation of V , which for convenience we write in compact notation of Eq. (1). An efficient algorithm, denoted by “monomial symmetrization approach,” has been coded to obtain higher-order polynomials from lower-order ones [55,56].

The coefficients, c_i , are obtained using a linear least-squares procedure; i.e., they are optimized to minimize the L2 loss, i.e., sum of the squared differences between the PIP representation of the potential and the electronic energies. If gradients are also available, these are included in the L2 loss, where the analytical gradient of V is used.

In general, the number of coefficients is much less than the dataset size and so this is an overdetermined linear least-squares problem that was solved by using `dgelss` from the MKL library [57] to obtain the linear coefficients. The `dgelss` routine is chosen because it uses singular value decomposition and therefore is the most robust (numerically stable) algorithm for such a problem. Details of the software to generate the PIPs as well as further processing of them, e.g., fast reverse differentiation using Mathematica scripts, have been given previously [56,58].

Δ -machine learning PIPs. A Δ -ML approach can be used to efficiently elevate a low-level PIP ML-PES, e.g., from extensive density functional theory or MP2 calculations to the CCSD(T) level [59–62]. In brief, the Δ -ML approach is given by the equation

$$V_{\text{LL} \rightarrow \text{CC}} = V_{\text{LL}} + \Delta V_{\text{CC-LL}}, \quad (3)$$

where $V_{\text{LL} \rightarrow \text{CC}}$ is the corrected PES, V_{LL} is a PES fit to low-level reference electronic structure data, and $\Delta V_{\text{CC-LL}}$ is the correction PES based on high-level data, e.g., coupled cluster energies. Since the difference between LL and HL energies, $\Delta V_{\text{CC-LL}}$, is not as strongly varying as V_{LL} with respect to the nuclear configurations, only 2688 high-level electronic energies were required to fit the correction PES. It is also possible to simply refit the original low-level dataset by incorporating (I) the difference energies directly into a new dataset to yield Set2025-I [63,64]. The two approaches, as expected, lead to nearly identical final results [64]. The latter approach is used here.

C. PIP fits to dipole moment data

Separate fits to dipole moment data are done, using a physically motivated expression for this vector quantity, namely,

$$\boldsymbol{\mu} = \mathbf{X} \mathbf{q}, \quad (4)$$

where $\mathbf{X}$ is the $3 \times N$ matrix of Cartesian coordinates and $\mathbf{q}$ is an $N \times 1$ column vector representing the “effective charges” on the N atoms, which are scalar quantities that can be expanded using a basis of PIPs to reflect relevant permutational invariance and with the linear coefficients determined by a linear least-squares fit. More details can be found in Refs. [3,65].

D. PhysNet and transfer learning

PhysNet is a message-passing neural network that learns to represent an atom in its chemical environments directly from nuclear charges and Cartesian coordinates [29]. The atomic feature vectors are used to predict atomic energy contributions and partial charges for any molecular configuration. Forces are obtained through reverse-mode automatic differentiation [66], while the partial charges are used to predict the electric dipole moment.

Following previous work [53], PhysNet’s learnable parameters are adjusted to minimize the loss function

$$\begin{aligned} \mathcal{L} = & w_E |E - E^{\text{ref}}| + \frac{w_F}{3N} \sum_{i=1}^N \sum_{\alpha=1}^3 \left| -\frac{\partial E}{\partial r_{i,\alpha}} - F_{i,\alpha}^{\text{ref}} \right| \\ & + w_Q \left| \sum_{i=1}^N q_i - Q^{\text{ref}} \right| + \frac{w_p}{3} \sum_{\alpha=1}^3 \left| \sum_{i=1}^N q_i r_{i,\alpha} - p_\alpha^{\text{ref}} \right| + \mathcal{L}_{\text{nh}} \end{aligned} \quad (5)$$

using the Adam optimizer [67,68]. The hyperparameters [21,29] w_i $i \in \{E, F, Q, p\}$ differentially weigh the contributions to the loss function and were $w_E = 1$ [1/energy], $w_F \sim 52.92$ [length/energy], $w_Q \sim 14.39$ [1/charge], and $w_p \sim 27.21$ [1/charge/length], respectively, and the term $\mathcal{L}_{\text{nh}}$ is a “nonhierarchical penalty” that regularizes the loss function [29]. For training, a random 80%/10%/10% split of the data as training/validation/test sets was used.

The present work employs two PhysNet-based PESs: PES2025 (available from previous work [53]) and PES2026. For PES2025, first an LL model was trained on Set2025-LL. To improve the quality of the LL model for molecular spectroscopy and tunneling splitting calculations, TL was used to refine and elevate the PES from the MP2/aug-cc-pVTZ to the gold-standard CCSD(T)/aug-cc-pVTZ level of theory using Set2025-HL [43,48–51]. For transfer learning, the learning rate was reduced to 10^{-4} and the hyperparameters were $w_E = w_F = w_Q = w_p = 1$. This transfer-learned model is referred to as PES2025 and additional details are provided in the original publication [53]. To validate and potentially further improve PES2025 for calculating H -tunneling splittings, PES2026 was trained based on Set2026-HL using transfer learning with Set2025-LL as the low-level model.

In addition to the energies, PhysNet also learns an underlying atom-centered charge model with geometry-dependent charges q_i that recovers the molecular dipole moment for each configuration. Hence, the final PhysNet model provides the total molecular energy as the sum of atomic energies

TABLE I. Statistical performance for PES2025, PES2026, and PIP PES across the MP2/aug-cc-pVTZ (Set2025-LL, Set2025-I) and CCSD(T)/aug-cc-pVTZ (Set2025-HL, Set2026-HL) datasets. Results on MAE and RMSE for energies, forces, and molecular dipole moment are reported as indicated together with the datasets the models were trained on.

Model	Technique	Dataset	MAE(E) (cm ⁻¹)	RMSE(E) (cm ⁻¹)	RMSE(F) (cm ⁻¹ / a_0)	RMSE(D) a.u. (D)
PES2025	Base	Set2025-LL	3.1	16.4	84.9	2.1×10^{-3} (5.4×10^{-3})
	TL	Set2025-HL	1.7	4.1	12.0	1.0×10^{-3} (2.6×10^{-3})
PES2026	TL	Set2026-HL	3.7	7.8	26.3	2.1×10^{-3} (5.3×10^{-3})
PIP PES	Δ -learn	Set2025-I	7.6	23.1	47.8	1.0×10^{-3} (2.6×10^{-3})
	Δ -learn	Set2025-HL	12.2	31.1	65.1	1.1×10^{-3} (2.7×10^{-3})

and the total dipole moment $\vec{\mu}(\mathbf{R}) = \sum_{i=1}^N q_i(\mathbf{R})\vec{r}_i$ for each geometry, which yield the full-dimensional potential energy and dipole moment surfaces (DMSs). Machine-learned fluctuating charge models, which capture the nuances of electronic redistribution during nuclear motion, have recently also been derived from kernel-based approaches and from an equivariant neural network [69,70].

The fluctuating charges in PhysNet are directly trained on the molecular dipole moment $\vec{\mu}$ and are smooth functions of the geometry. Because these charges are forced to be a smooth function of the coordinates to minimize the dipole loss, the resulting dipole moment surface is inherently continuous and differentiable. This is vital for obtaining high-quality relative intensities in IR signatures, as the model accurately captures the dipole derivatives $\frac{\partial \mu}{\partial Q}$ across the PES. This methodology aligns with previous work on formaldehyde [71], which demonstrated that allowing electrostatics to fluctuate naturally with geometry provides a robust foundation for reproducing complex spectroscopic features. High-quality dipole moment surfaces also allow for the comparison of relative intensities of infrared responses. However, for characterizing the energies at which particular vibrational modes appear, more qualitative models based on, for example, point charges are sufficient.

III. RESULTS

A. Fits and validation of the machine-learned PESs

The present work employs three different PESs. The PhysNet-based PES2025 was previously trained on Set2025-LL and Set2025-HL and used for spectroscopic applications (VPT2 calculations and MD simulations). To extend its validity to tunneling splitting calculations, it was supplemented by additional reference calculations at the same level of theory along the IP and MEP to give Set2026-HL (see Sec. II). Compared with PES2025 [mean absolute error (MAE) of $E = 3.1$ cm⁻¹; root mean squared error (RMSE) of $E = 16.4$ cm⁻¹ on Set2025-LL; MAE(E) = 1.7 cm⁻¹; RMSE(E) = 4.2 cm⁻¹ on Set2025-HL], PES2026 on Set2026-HL for an independent test set features RMSE(E) = 7.8 cm⁻¹, 26.3 cm⁻¹/ a_0 for gradients (compared with 12.0 cm⁻¹/ a_0 for Set2025-HL), and 2.1×10^{-3} a.u. (5.3×10^{-3} D) for dipoles (see Table I).

Complementary to these PhysNet-based PESs, a PIP-represented PES [3,30] was generated using the PES2025-LL dataset. The PIPs used full “421 symmetry,” indicating that all four oxygens permute with one another and the two carbons permute with one another. The polynomial basis has 9923 polynomials/coefficients. As described in Sec. II, this fit was

then elevated to the CCSD(T) level using the Δ -ML method and the Set2025-HL dataset. The rms errors of the final PES on Set2025-I are 23.1 cm⁻¹ for energies and 47.8 cm⁻¹/ a_0 for gradients, and on Set2025-HL they are 31.1 cm⁻¹ for energies and 65.1 cm⁻¹/ a_0 for gradients. The DMS using PIPs featured a training error of 1.0×10^{-3} a.u. (2.6×10^{-3} D).

The representations PES2025 and PIP-PES are compared in Fig. 2(a). For this, 1000 configurations were randomly sampled from an MD trajectory generated using PES2025 at a temperature of 300 K. This dataset covers a comparatively narrow energy range of ~ 10 kcal/mol. The two PESs feature an MAE(E) and RMSE(E) of 0.010 and 0.014 kcal/mol. Similarly, PES2025 and PES2026 are juxtaposed in Fig. 2(b) for which MAE(E) and RMSE(E) are even closer: 0.002 and 0.003 kcal/mol, respectively. These comparisons demonstrate that all three PESs are consistent with one another. On the other hand, the subtle differences are relevant when comparing observables obtained from using them in the subsequent simulations.

PES2026 features a barrier height for H-transfer of 3.349 kcal/mol compared with 3.351 kcal/mol for PES2025, which matches the CCSD(T) reference value. Normal mode frequencies for the global minimum and the TS were also determined using PES2026 and their MAE(ω) match those from PES2025 to within better than 0.1 cm⁻¹. Compared with the frequencies from CCSD(T) calculations for the minimum energy and TS structure, the MAE(ω) are 0.41 and 0.55 cm⁻¹. All these comparisons underline the high quality of both PES2025 and PES2026. The quality of the normal mode frequencies is also in line with previous work using the same methods [72].

For the PIP-PES, the barrier height for H-transfer is 1171.7 cm⁻¹, in excellent agreement with the CCSD(T) value of 1172.2 cm⁻¹ (3.351 kcal/mol). The harmonic frequencies of these two stationary points are listed in Table II. The MAE between the reference values and those from the PIP-PES are 0.47 and 0.85 cm⁻¹, respectively. This compares with 0.34 cm⁻¹ [CCSD(T)/aug-cc-pVTZ] and 0.09 (MP2/aug-cc-pVTZ) for the global minimum of PES2025 [53].

B. Infrared spectroscopy

The computed IR spectra from PES2025 and the PIP-PES, computed using their respective dipole moment surfaces, together with the most recent experiment, are shown in Fig. 3 [53]. It can be seen that the spectra (blue for the PIP-PES and red for PES2025) from vibrational self-consistent field/vibrational configuration interaction (VSCF/VCI) calculations using both PESs successfully capture the

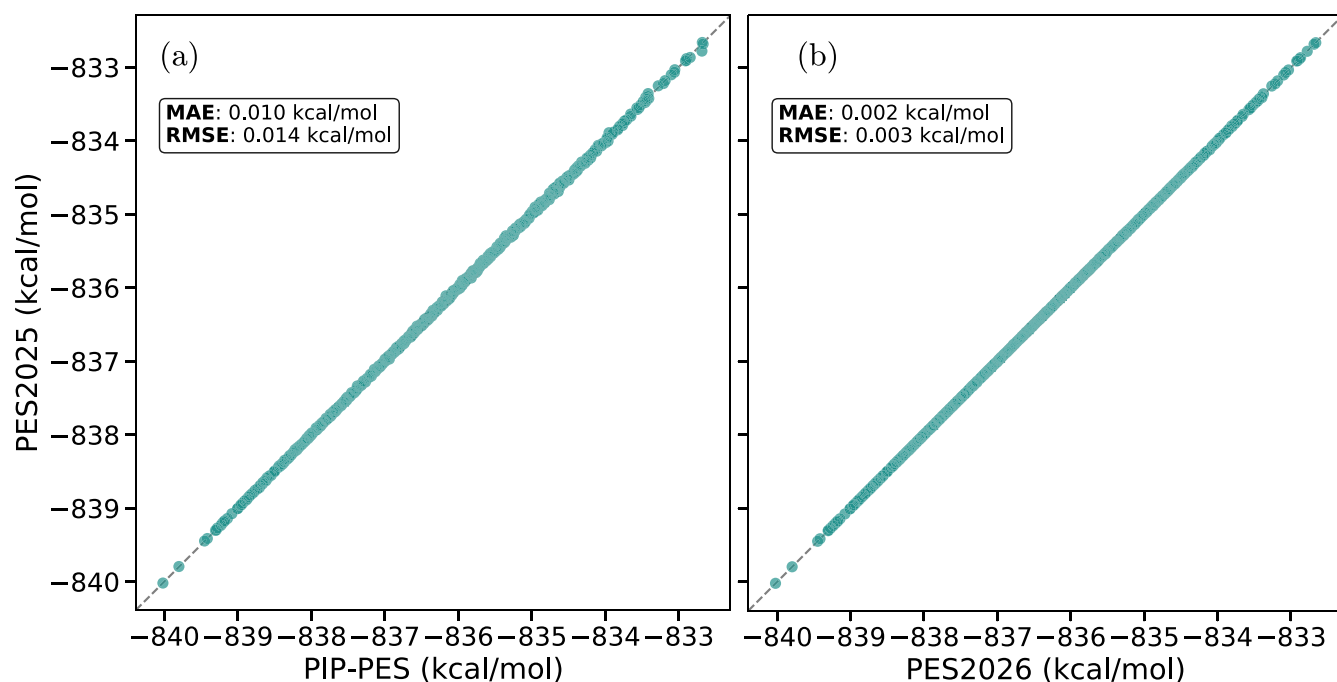


FIG. 2. (a) Comparison between PES2025 (based on PhysNet) and the PIP-based PES. (b) Comparison between PES2025 and PES2026. The statistical performance for PES2025, PES2026, and PES-PIP on the CCSD(T)/aug-cc-pVTZ reference data is summarized in Table I.

location and breadth of the highly dispersed band in the expected OH-stretch region ranging from ~ 2500 to $\sim 3200\text{ cm}^{-1}$. A detailed analysis of this band and its implications for quantum chaos using the PIP-PES has been recently discussed [36]. The focus in the present work is on the fidelity of the potential and dipole moment surfaces that are constructed with two different approaches (neural network versus polynomial regression) using the same reference data. A direct comparison of the VSCF/VCI energies from both

ML-PESs is given in Fig. S1 of the Supplemental Material [47].

The two computed spectra in Fig. 3 (blue and red traces) are visually identical, even for the fine structures in the highly dispersed band between 2000 and 3500 cm^{-1} . The quantitative difference between the two computed spectra is the green trace: peak positions for MULTIMODE (MM) calculations using the two PESs differ by a few cm^{-1} and arise primarily for the framework modes ($1200\text{--}1800\text{ cm}^{-1}$). It is also

TABLE II. Comparison of harmonic frequencies (cm^{-1}) from CCSD(T) and the PIP-PES for the two structures shown in Fig. 1. For PES2025, the frequencies have been reported previously with MAEs of 0.34 cm^{-1} for the minimum and 0.57 cm^{-1} for the transition state for H-transfer [53]. Given the very small overall $\text{MAE}(\omega) = 0.1\text{ cm}^{-1}$ between PES2025 and PES2026, the results for PES2026 are not explicitly reported.

Min mode	CCSD(T)	PES	Diff	TS Mode	CCSD(T)	PES	Diff
1	104.15	104.44	0.29	1	1143.01 <i>i</i>	1141.21 <i>i</i>	1.80
2	300.53	300.48	0.05	2	142.08	142.68	0.60
3	434.85	435.22	0.37	3	333.55	334.10	0.55
4	488.55	488.20	0.35	4	492.74	493.82	1.08
5	565.63	565.45	0.18	5	602.51	602.91	0.40
6	700.53	700.64	0.11	6	715.09	714.51	0.58
7	829.31	828.73	0.58	7	750.74	749.58	1.16
8	849.26	849.41	0.15	8	852.74	851.81	0.93
9	968.64	967.52	1.12	9	865.80	865.14	0.66
10	1140.19	1141.01	0.82	10	1296.66	1298.50	1.84
11	1335.09	1335.15	0.06	11	1299.88	1299.93	0.05
12	1455.66	1455.68	0.02	12	1322.91	1323.09	0.18
13	1726.89	1727.56	0.67	13	1748.71	1748.66	0.05
14	1816.52	1816.89	0.37	14	1790.83	1791.51	0.68
15	3203.91	3205.81	1.90	15	2101.26	2099.06	2.20
MAE			0.47				0.85

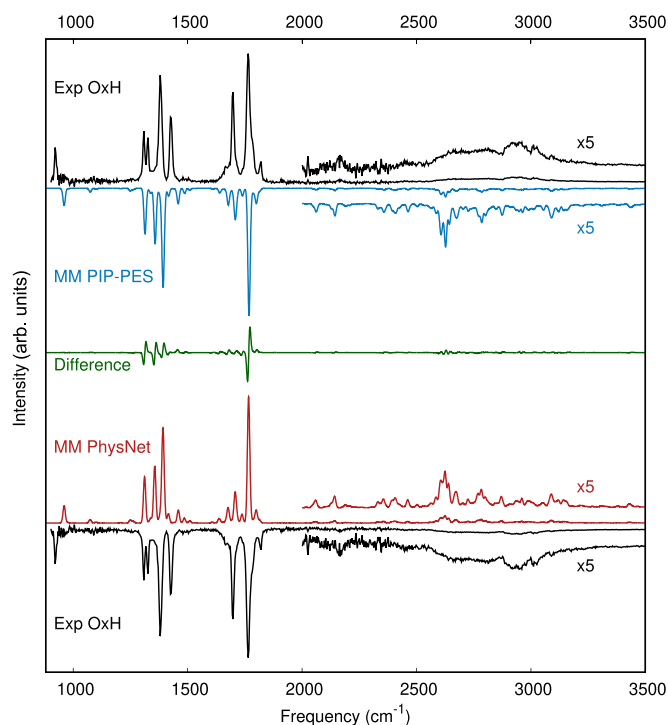


FIG. 3. IR spectrum for OxH calculated at the respective global minimum using MM and a 4MR with two different potential and dipole moment surfaces. The spectra using the PIP-PES and PES2025 are reported in blue and red, respectively, whereas the difference between the two computed spectra is shown in green. The IR spectrum for PES2026 was not calculated explicitly because (1) PES2025 and PES2026 agree to within better than 1 cm^{-1} [see Fig. 2(b)] and (2) the normal modes from the two PESs agree to within 0.1 cm^{-1} . The experimentally measured spectrum is displayed in black [53,73]. Results with a 3MR are reported in Fig. S2 of the Supplemental Material [47], and those for PES2026 using 4MR are in Fig. S3 of the Supplemental Material [47].

noted that the relative intensities of the modes are realistic and in reasonable agreement with the measurements. It is recalled that PhysNet uses fluctuating charges from which the fluctuating molecular dipole moment is computed, whereas for the PIP-PES the dedicated DMS PIP representation is used.

The spectra from MULTIMODE calculations are compared over the full range that was reported from the measurements [53,73], starting at the red side of the spectrum. The weak band below 1000 cm^{-1} is not correctly described by the MM calculations, whereas the quartet below 1500 cm^{-1} rather closely follows the experiments. The spacing for the first doublet is somewhat too wide and the intensity of the highest-frequency peak is too low. For the doublet (experiment) between 1500 and 2000 cm^{-1} , the MM calculations reveal more detail for the low-frequency peak, whereas the high-frequency peak is correctly described. Overall, the agreement between the two computed spectra is excellent and they closely match the experiments. For the diffuse band extending from 2500 to 3200 cm^{-1} (see black traces in Fig. 3), the MM calculations using the two different PESs (PES2025 and PIP-PES) yield nearly identical results (green trace in Fig. 3).

The computations using MM confirm that the intensity of the spectroscopic feature(s) associated with H-transfer are spread out across several hundred wave numbers. On the other hand, the computed lineshape does not find the plateau centered around 2750 cm^{-1} .

C. Tunneling splittings

Next, tunneling splittings from the methods described above are presented here. Most determinations are for the ground vibrational state, but some excited state splittings are also reported using the Q_{im} method.

Instanton calculations. Tunneling splittings had been previously determined using PES2025, which was transfer learned from the MP2 level of theory to CCSD(T). For this PES, the training dataset was not specifically tailored to tunneling splitting calculations as had been done previously for malonaldehyde or tropolone [43,74]. In order to further improve PES2025 for tunneling splittings, reference energies at the CCSD(T)/aug-cc-pVTZ level of theory were determined along the minimum energy and instanton paths. Subsequently, the TL procedure was repeated that yielded PES2026 (see Fig. 2). The tunneling splitting calculated using perturbatively corrected (pcRPI) RPI theory using PES2025 was 35.0 cm^{-1} . This compares with a value of 34.9 cm^{-1} when employing PES2026. The difference of 0.1 cm^{-1} indicates that both surfaces provide comparable performance and that the prediction at the CCSD(T)/aug-cc-pVTZ level of theory is $\sim 35.0\text{ cm}^{-1}$.

DMC calculations. Diffusion Monte Carlo calculations were performed with both the PIP-PES and the PES2025 surfaces to determine the zero point and first excited levels of OxH. The latter calculation was accomplished by preventing the walkers from crossing the nodal surface. The calculation with the PIP-PES was performed in terms of normal modes, while that using the PES2025 surface was performed in Cartesian coordinates. Uncertainties were calculated from the standard deviation of five simulations. For the zero-point energy, the PIP-PES gave $7820.1 \pm 1.6\text{ cm}^{-1}$, and the PES2025 surface gave $7819.0 \pm 2.0\text{ cm}^{-1}$. For the first excited levels, the results were 7853.1 ± 0.9 and $7856.9 \pm 1.0\text{ cm}^{-1}$, respectively, resulting in tunneling splittings of 33.0 ± 1.8 and $37.9 \pm 2.3\text{ cm}^{-1}$. Given that PES2026 and PES2025 agree to within better than 1 cm^{-1} on average [Fig. 2(b)] and the differences on the 1 to few cm^{-1} scale between PES2025 and the PIP-PES, DMC simulations were not carried out using PES2026.

Q_{im} calculations in 1D and 2D. For additional scientific interest and deeper understanding, tunneling splittings were also determined from reduced-dimensionality Q_{im} calculations. This was done only for the PIP-PES; however, connections between the PESs will be drawn at the end of this subsection. Figure 4 reports $V(Q_{im})$ together with the energy levels and probability densities derived from the wavefunctions. The barrier is 1172 cm^{-1} . Using the normal coordinate system of the saddle point, the normal coordinate of the minimum is $\vec{Q}_{min} = [34.2, 0.0, 20.2, 0.0, 0.6, 31.0, 10.3, 0.0, -14.2, 0.0, -4.5, 3.2, 3.4, 0.1, -8.4]$ (or $-\vec{Q}_{min}$ for the other minimum), where the 15 numbers represent distances in multiples of the unit vectors associated with the normal modes of the saddle

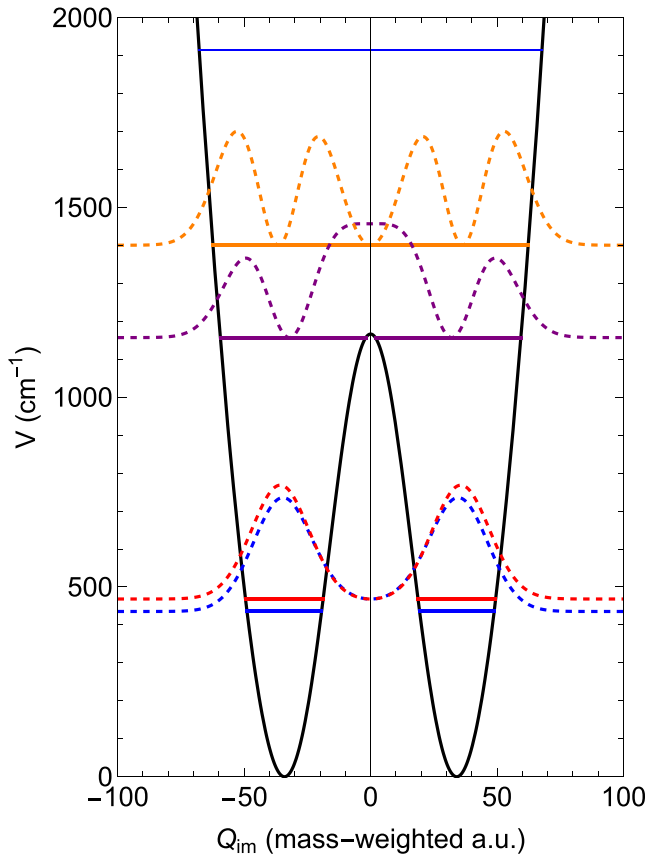


FIG. 4. $V(Q_{im})$ potential for OxH (black) with DVR energy levels (solid horizontal lines) and $|\Psi^2|$ distributions (dashed). The coloring of the energy levels corresponds to the coloring of the distributions. The ground-state tunneling splitting between the lowest doublet (blue/red solid lines) is clearly visible.

point, starting with that of the imaginary frequency and then in order of the frequencies from lowest to highest (see Table II).

Energy levels were also calculated for a barrier height of 1250 cm^{-1} to agree with the complete basis set (CBS) limit extrapolated from CCSD(T)-F12a/aug-cc-pVXZ ($X = D$ and T) calculations (see Table S1 and Fig. S4 of the Supplemental Material [47]). Using the geometries of the Q_{im} path for the 1172 cm^{-1} barrier, points along the path were calculated at the CCSD(T)-F12a CBS level and then fit by a polynomial function that was used in the DVR calculation. Use of DVR for this curve gave a tunneling splitting of 28 cm^{-1} .

A two-dimensional (2D) version of the PES was also examined using Q_{im} and Q_{15} , the highest-frequency normal mode at the saddle point, as the two dimensions, and then performing an optimization of the remaining $3N - 8$ normal coordinates at each value of Q_{im} between $-100.0 a_0$ and $100.0 a_0$; and at each value of Q_{15} between $-72.0 a_0$ and $40 a_0$; Q is reported in mass weighted atomic units, whereby the mass is in units of m_e . The potential surface is shown in colored contour form in Fig. 5. The grid spacing for the diagram is $0.5 a_0$ in each dimension. The black solid line shows the Q_{im} path on the PIP-PES, the green dashed line shows the minimum energy path on PES2025, and the red line shows the instanton path on PES2025. Importantly, despite using two different PES

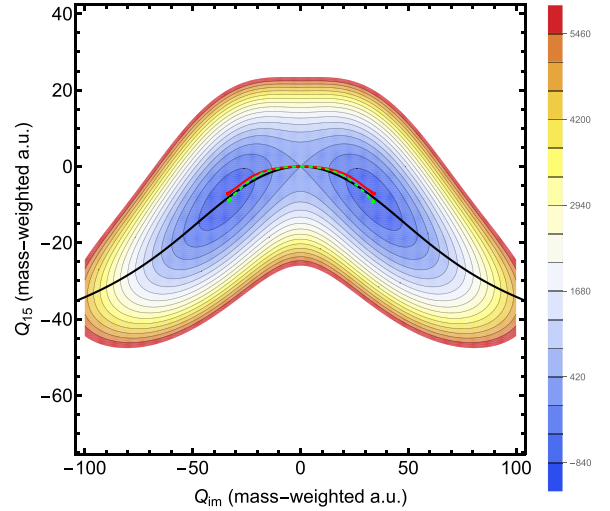


FIG. 5. Two-dimensional (colored contours) for OxH as a function of Q_{im} and Q_{15} with all other normal modes optimized at each point using PIP-PES. The instanton, minimum energy, and Q_{im} path are the red, green dashed, and black paths. The instanton and minimum energy paths were determined on PES2025 and the Q_{im} path on the PIP-PES. One division along either coordinate corresponds to the unit vector along the relevant normal mode.

representations (PES2025, PIP-PES) and the different methods used for computing them (RPI, MEP, 2D- Q_{im}), the three paths are essentially superimposed in the regions between the transition state and the global minimum. The associated tunneling splitting is also 33 cm^{-1} , nearly identical to that from the 1D Q_{im} analysis and in close agreement with the DMC simulations using the PIP-PES as well as with the RPI calculations using PES2025. Because the RPI and Q_{im} paths pass through the saddle point, corner cutting in OxH is small, unlike in tropolone, where pronounced corner cutting was observed [61]. This contributes to the present finding that Δ_H from the 1D- Q_{im} approach, which is simpler and more approximate than DMC and RPI simulations, compares rather favorably with these methods and representations of the PESs. A paper comparing the Q_{im} method for a variety of H-transfer molecules has recently appeared [75].

Tunneling summary. A summary of the tunneling splittings for OxH is provided in Table III. Given the similarities between the various PESs, it is not surprising that all splitting from using PESs with a barrier of 1172 cm^{-1} give compara-

TABLE III. Comparison of the tunneling splittings Δ_H for H-transfer in OxH using the different methods described in the text.

PES	Barrier (cm^{-1})	Tunneling splitting (cm^{-1})	Method	Reference
PES2025	1172	35	RPI	[53]
PES2026	1171	35	RPI	This work
PIP-PES	1172	33	DMC	This work
PES2025	1172	38	DMC	This work
PIP-PES	1172	33	1D Q_{im}	This work
PIP-PES	1172	33	2D Q_{im}	This work
PIP-PES	1250	28	1D Q_{im}	This work

ble splittings. What is more encouraging is that the different methods—RPI, DMC, and Q_{im} —also give consistent results. For a barrier of 1172 cm^{-1} , all tunneling splittings are quite consistently between 33 and 36 cm^{-1} . A comparison of the instanton and Q_{im} curves on the 2D PIP surface shows that they are almost identical. This explains why the RPI and Q_{im} calculations agree to within $\sim 2\text{ cm}^{-1}$, i.e., within $\sim 5\%$. The corrected barrier height of 1250 cm^{-1} —estimated from the CCSD(T)/CBS extrapolation and used for a corrected 1D Q_{im} curve—leads to a smaller splitting $\Delta_H = 28\text{ cm}^{-1}$. Given the high level of the electronic structure calculation for that higher barrier, the splitting of 28 cm^{-1} may indicate that the actual splitting is in the $28\text{--}33\text{ cm}^{-1}$ range. This is a testable prediction from highest-level treatments for experimental verification.

Summary of validations. For all three PESs considered in the present work, their statistical performance on the test set and their capability to correctly describe the normal mode frequencies at the CCSD(T)/aug-cc-pVTZ level were established. The overlap for the IR spectra from the MM simulations for PES2025 and PIP-PES with each other and with measurements is remarkable. Using PES2025 for the tunneling splitting calculations based on RPI and DMC simulations, Δ_H differs by 3 cm^{-1} (see Table III). This is within the expected errors of the RPI and DMC methods, respectively. For tunneling simulations based on PES2025 and PES2026 from RPI simulations, $\Delta\Delta_H$ is negligible. This is consistent with Fig. 2(b). Hence, PES2026 establishes the validity of PES2025 for computing tunneling splittings. Overall, since the MAE between PES2025 and PES2026 is negligible (0.002 kcal/mol), $\Delta\Delta_H \sim 0$ from RPI simulations, and such simulations probe regions around the minimum and the TS, no MM calculations are needed to further validate PES2026. Nevertheless, Fig. S3 of the Supplemental Material [47] compares the IR spectrum with that computed from PES2025 and finds the two to be rather close, as expected. DMC simulations with PES2025 and PIP-PES yield a difference of 2 cm^{-1} for the tunneling splittings. Again, this is within the expected error of the nuclear dynamics method. The consistent agreement across this set of observables demonstrates the comparable performance of the PESs and suggests that exhaustively computing every observable with every method considered is largely redundant for mutual PES verification.

IV. SUMMARY AND OUTLOOK

The performance of two machine-learned potentials, Δ -ML PIP and TL-PhysNet, trained on the same datasets of energies and forces, was examined for OxH. Dipole moment fits were also done using two corresponding approaches.

With both representations of the PESs, the IR spectra and ground-state tunneling splittings were calculated by using state-of-the-art quantum methods, with nearly identical results and good agreement with experiments for the vibrational spectra. This suggests that, at the given level of theory [CCSD(T)], for the given datasets, and with the methods to represent these ML-PESs (PhysNet and PIP), a saturation in the accuracy of the ML-PESs has been reached. Each of these calculations—the VSCF/VCI calculations and the DMC simulations—sample of order 10^9 configurations. Hence, this represents a major and quantitative test of the agreement between the predictions of the two ML-PESs.

From the full body of simulations carried out in the present work, it emerges that both representations—PhysNet and PIP—are of comparable quality without a clear preference of one over the other. This confirms earlier findings [76] for the $\text{O} + \text{O}_2$ reaction and indicates that the field of small-molecule reaction dynamics has reached a level of maturity that enables predictive simulations of measurable quantities, as demonstrated in the present work for oxalate ($[\text{HO}_2\text{CCO}_2]^-$), a considerably larger system.

Although not the main point of the paper, it is worth noting that the width of the highly diffuse and weak OH-stretch band from using both PESs is consistent with experiment. This also indicates the high accuracy of the two ML-PESs. However, the fine details are not in perfect agreement with the measurements. The ground-state double-well tunneling splitting, predicted to be $\sim 30\text{ cm}^{-1}$, has not been measured nor explicitly treated in any of the calculations of the IR spectrum, including the ones shown here. Hence, the influence of the double-well splitting on the computed IR spectrum is as yet unknown and remains challenging. Hopefully, this will be uncovered both in future experiments and calculations. The present PESs are of sufficient quality to be used in such future studies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Swiss National Science Foundation through Grants No. 200020_219779 (to M.M.) and No. 200021_215088 (to M.M.), and the University of Basel (to M.M.). This article was also based upon work within COST Action COSY CA21101, supported by COST (European Cooperation in Science and Technology) (to M.M.).

DATA AVAILABILITY

The codes and data for the present study are available in Ref. [77].

-
- [1] M. Meuwly, Reactive molecular dynamics: From small molecules to proteins, *WIREs Comput. Mol. Sci.* **9**, e1386 (2019).
- [2] S. Manzhos and T. J. Carrington, Neural network potential energy surfaces for small molecules and reactions, *Chem. Rev.* **121**, 10187 (2021).

- [3] B. J. Braams and J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- [4] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, *Chem. Rev.* **121**, 10142 (2021).

- [5] D. Lu, H. Wang, M. Chen, L. Lin, R. Car, E. Weinan, W. Jia, and L. Zhang, 86 PFLOPS deep potential molecular dynamics simulation of 100 million atoms with *ab initio* accuracy, *Comput. Phys. Commun.* **259**, 107624 (2021).
- [6] L. Zhang, J. Han, H. Wang, R. Car, and W. E., Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, *Phys. Rev. Lett.* **120**, 143001 (2018).
- [7] O. T. Unke *et al.*, Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments, *Sci. Adv.* **10**, eadn4397 (2024).
- [8] V. Zaverkin, M. Ferraz, F. Alesiani, and M. Niepert, Performance of universal machine-learned potentials with explicit long-range interactions in biomolecular simulations, *J. Chem. Theory Comput.* **22**, 2074 (2026).
- [9] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, Learning local equivariant representations for large-scale atomistic dynamics, *Nat. Commun.* **14**, 579 (2023).
- [10] B. Kozinsky, A. Musaelian, A. Johansson, and S. Batzner, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC'23* (ACM Press, New York, 2023).
- [11] K. Kadau, T. C. Germann, and P. S. Lomdahl, Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/L, *Int. J. Mod. Phys. C* **17**, 1755 (2006).
- [12] T. C. Germann and K. Kadau, Trillion-atom molecular dynamics becomes a reality, *Int. J. Mod. Phys. C* **19**, 1315 (2008).
- [13] Y. Shibuta, S. Sakane, E. Miyoshi, S. Okita, T. Takaki, and M. Ohno, Heterogeneity in homogeneous nucleation from billion-atom molecular dynamics simulation of solidification of pure metal, *Nat. Commun.* **8**, 10 (2017).
- [14] L. Casalino, C. Seitz, J. Lederhofer, Y. Tsybovsky, I. A. Wilson, M. Kanekiyo, and R. E. Amaro, Breathing and tilting: Mesoscale simulations illuminate influenza glycoprotein vulnerabilities, *ACS Cent. Sci.* **8**, 1646 (2022).
- [15] K. Santos *et al.*, Breaking the molecular dynamics timescale barrier using a wafer-scale system, in *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, New York, 2024), pp. 1–13.
- [16] D. Ugarte La Torre, J. Jung, and Y. Sugita, Scaling molecular dynamics for large-scale simulation of biological systems on AMD CPU/GPU supercomputers: Lessons from LUMI: Optimizing GENESIS for maximizing the computational efficiency of CPU and GPU kernels on the LUMI supercomputer, in *Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region* (ACM ICPS, New York, 2025), pp. 1–12.
- [17] K. Lek Chaton and M. Meuwly, Enhancing empirical energy functions using physics- and machine learning-based extensions: Structure, dynamics and spectroscopy of modified benzenes, *J. Comput. Chem.* **46**, e70162 (2025).
- [18] S.-L. J. Lahey and C. N. Rowley, Simulating protein–ligand binding with neural network potentials, *Chem. Sci.* **11**, 2362 (2020).
- [19] M. Gastegger, K. T. Schütt, and K.-R. Müller, Machine learning of solvent effects on molecular spectra and reactions, *Chem. Sci.* **12**, 11473 (2021).
- [20] T. J. Inizan, T. Plé, O. Adjoua, P. Ren, H. Gökcen, O. Isayev, L. Lagardère, and J.-P. Piquemal, Scalable hybrid deep neural networks/polarizable potentials biomolecular simulations including long-range effects, *Chem. Sci.* **14**, 5438 (2023).
- [21] K. Song, S. Käser, K. Töpfer, L. I. Vazquez-Salazar, and M. Meuwly, PhysNet meets CHARMM: A framework for routine machine learning/molecular mechanics simulations, *J. Chem. Phys.* **159**, 024125 (2023).
- [22] K. Zinovjev, Electrostatic embedding of machine learning potentials, *J. Chem. Theory Comput.* **19**, 1888 (2023).
- [23] K. Zinovjev, L. Hedges, R. Montagud Andreu, C. Woods, I. Tuñón, and M. W. van der Kamp, *emle-engine*: A flexible electrostatic machine learning embedding package for multiscale molecular dynamics simulations, *J. Chem. Theory Comput.* **20**, 4514 (2024).
- [24] J. Kalayan, I. Ramzan, C. D. Williams, R. A. Bryce, and N. A. Burton, A neural network potential based on pairwise resolved atomic forces and energies, *J. Comput. Chem.* **45**, 1143 (2024).
- [25] I. Poltavsky *et al.*, Crash testing machine learning force fields for molecules, materials, and interfaces: Model analysis in the TEA challenge 2023, *Chem. Sci.* **16**, 3720 (2025).
- [26] I. Poltavsky *et al.*, Crash testing machine learning force fields for molecules, materials, and interfaces: Molecular dynamics in the TEA challenge 2023, *Chem. Sci.* **16**, 3738 (2025).
- [27] J. Xia, Y. Zhang, and B. Jiang, The evolution of machine learning potentials for molecules, reactions and materials, *Chem. Soc. Rev.* **54**, 4790 (2025).
- [28] W. G. Stark, C. van der Oord, I. Batatia, Y. Zhang, B. Jiang, G. Csányi, and R. J. Maurer, Benchmarking of machine learning interatomic potentials for reactive hydrogen dynamics at metal surfaces, *Mach. Learn.: Sci. Technol.* **5**, 030501 (2024).
- [29] O. T. Unke and M. Meuwly, PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges, *J. Chem. Theory Comput.* **15**, 3678 (2019).
- [30] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu, A perspective marking 20 years of using permutationally invariant polynomials for molecular potentials, *J. Chem. Phys.* **162**, 180901 (2025).
- [31] S. Käser, L. I. Vazquez-Salazar, M. Meuwly, and K. Töpfer, Neural network potentials for chemistry: Concepts, applications and prospects, *Digital Discovery* **2**, 28 (2023).
- [32] W. Hwang *et al.*, CHARMM at 45: Enhancements in accessibility, functionality, and speed, *J. Phys. Chem. B* **128**, 9976 (2024).
- [33] J. M. Bowman, S. Carter, and X. Huang, MULTIMODE: A code to calculate rovibrational energies of polyatomic molecules, *Int. Rev. Phys. Chem.* **22**, 533 (2003).
- [34] Q. Yu *et al.*, MULTIMODE, the n-mode representation of the potential and illustrations to IR spectra of glycine and two protonated water clusters, in *Vibrational Dynamics of Molecules* edited by J. M. Bowman (World Scientific, Singapore, 2022), pp. 296–339.
- [35] C. Qu, T. C. Allison, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, Computational spectroscopy using MULTIMODE and machine-learned potentials, *J. Chem. Phys.* **164**, 134116 (2026).
- [36] C. Qu, P. L. Houston, and J. M. Bowman, Does the diffuse OH-stretch band in the IR spectrum of protonated oxalate exhibit quantum chaos? *J. Chem. Phys.* **164**, 061103 (2026).
- [37] J. B. Anderson, A random-walk simulation of the Schrödinger equation: H^+_3 , *J. Chem. Phys.* **63**, 1499 (1975).

- [38] I. Kosztin, B. Faber, and K. Schulten, Introduction to the diffusion Monte Carlo method, *Am. J. Phys.* **64**, 633 (1996).
- [39] J. B. Anderson, Quantum chemistry by random walk. H^2P , $H_3^+ D_{3h}^1A_1'$, $H_2^3\Sigma_u^+$, $H_4^1\Sigma_g^+$, Be^1S , *J. Chem. Phys.* **65**, 4121 (1976).
- [40] Y. Wang, B. J. Braams, J. M. Bowman, S. Carter, and D. P. Tew, Full-dimensional quantum calculations of ground-state tunneling splitting of malonaldehyde using an accurate *ab initio* potential energy surface, *J. Chem. Phys.* **128**, 224314 (2008).
- [41] J. O. Richardson and S. C. Althorpe, Ring-polymer instanton method for calculating tunneling splittings, *J. Chem. Phys.* **134**, 054109 (2011).
- [42] J. O. Richardson, Ring-polymer instanton theory, *Int. Rev. Phys. Chem.* **37**, 171 (2018).
- [43] S. Käser, J. O. Richardson, and M. Meuwly, Transfer learning for affordable and high-quality tunneling splittings from instanton calculations, *J. Chem. Theory Comput.* **18**, 6840 (2022).
- [44] J. E. Lawrence, J. Dušek, and J. O. Richardson, Perturbatively corrected ring-polymer instanton theory for accurate tunneling splittings, *J. Chem. Phys.* **159**, 014111 (2023).
- [45] E. Kamarchik, Y. Wang, and J. Bowman, Reduced-dimensional quantum approach to tunneling splittings using saddle-point normal coordinates, *J. Phys. Chem. A* **113**, 7556 (2009).
- [46] D. T. Colbert and W. H. Miller, A novel discrete variable representation for quantum mechanical reactive scattering via the S -matrix Kohn method, *J. Chem. Phys.* **96**, 1982 (1992).
- [47] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/jfjx-djk2> for methods for the spectroscopic studies [33–36], diffusion Monte Carlo calculations [37–40], and computing tunneling splittings [41–46], and reports a comparison of IR spectra for 3MR and 4MR VSCF/VCI calculations and details of the CCSD(T)/CBS energies along the Q_{im} path.
- [48] M. E. Taylor and P. Stone, Transfer learning for reinforcement learning domains: A survey, *J. Mach. Learn. Res.* **10**, 1633 (2009).
- [49] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* **22**, 1345 (2009).
- [50] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, *Nat. Commun.* **10**, 2903 (2019).
- [51] S. Käser and M. Meuwly, Transfer learned potential energy surfaces: Accurate anharmonic vibrational dynamics and dissociation energies for the formic acid monomer and dimer, *Phys. Chem. Chem. Phys.* **24**, 5269 (2022).
- [52] H.-J. Werner, P. J. Knowles, *et al.*, MOLPRO, version 2019.2, a package of *ab initio* programs (2019), <https://www.molpro.net/>.
- [53] V. Andreichev, S. Käser, E. L. Bocanegra, M. Salik, M. A. Johnson, and M. Meuwly, Dynamics of protonated oxalate from machine-learned simulations and experiment: Infrared signatures, proton transfer dynamics and tunneling splittings, *Phys. Chem. Chem. Phys.* **27**, 23288 (2025).
- [54] S. Käser, D. Koner, and M. Meuwly, Compact kernel/neural network representation for accurate, fast, and global reactive molecular potential energy surfaces, *Precis. Chem.* **4**, 571 (2026).
- [55] Z. Xie and J. M. Bowman, Permutationally invariant polynomial basis for molecular energy surface fitting via monomial symmetrization, *J. Chem. Theory Comput.* **6**, 26 (2010).
- [56] MSA Software with Gradients, 2029, <https://github.com/szquchen/MSA-2.0>, accessed 20 January 2019.
- [57] Intel Corporation, Intel® oneAPI Math Kernel Library, 2020, <https://www.intel.com/content/www/us/en/developer/tools/oneapi/onemkl.html>.
- [58] P. L. Houston, C. Qu, Q. Yu, R. Conte, A. Nandi, J. K. Li, and J. M. Bowman, PESPIP: Software to fit complex molecular and many-body potential energy surfaces with permutationally invariant polynomials, *J. Chem. Phys.* **158**, 044109 (2023).
- [59] R. Ramakrishnan, P. Dral, M. Rupp, and O. A. von Lilienfeld, Big data meets quantum chemistry approximations: The Δ -machine learning approach, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- [60] A. Nandi, C. Qu, P. L. Houston, R. Conte, and J. M. Bowman, Δ -machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory, *J. Chem. Phys.* **154**, 051102 (2021).
- [61] A. Nandi *et al.*, Ring-polymer instanton tunneling splittings of tropolone and isotopomers using a Δ -machine learned CCSD(T) potential: Theory and experiment shake hands, *J. Am. Chem. Soc.* **145**, 9655 (2023).
- [62] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu, Δ -machine learned potential energy surfaces and force fields, *J. Chem. Theory Comput.* **19**, 1 (2023).
- [63] Y. Liu and J. Li, Permutation-invariant-polynomial neural-network-based Δ -machine learning approach: A case for the HO_2 self-reaction and its dynamics study, *J. Phys. Chem. Lett.* **13**, 4729 (2022).
- [64] P. L. Houston, C. Qu, Q. Yu, P. Pandey, R. Conte, A. Nandi, J. M. Bowman, and S. G. Kukolich, Formic acid–ammonia heterodimer: A new Δ -machine learning CCSD(T)-level potential energy surface allows investigation of the double proton transfer, *J. Chem. Theory Comput.* **20**, 1821 (2024).
- [65] X. Huang, B. J. Braams, and J. M. Bowman, *Ab initio* potential energy and dipole moment surfaces for $H_5O_2^+$, *J. Chem. Phys.* **122**, 044308 (2005).
- [66] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008), 2nd ed.
- [67] D. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [68] S. J. Reddi, S. Kale, and S. Kumar, On the convergence of adam and beyond, [arXiv:1904.09237](https://arxiv.org/abs/1904.09237).
- [69] E. Boittier, K. Töpfer, M. Devereux, and M. Meuwly, Kernel-based minimal distributed charges: A conformationally dependent ESP-model for molecular simulations, *J. Chem. Theory Comput.* **20**, 8088 (2024).
- [70] E. D. Boittier and M. Meuwly, Efficient and equivariant prediction of distributed charges for accurate molecular electrostatics, *J. Chem. Theory Comput.* **22**, 4532 (2026).
- [71] S. Käser, D. Koner, A. S. Christensen, O. A. von Lilienfeld, and M. Meuwly, Machine learning models of vibrating H_2CO : Comparing reproducing kernels, FCHL, and PhysNet, *J. Phys. Chem. A* **124**, 8853 (2020).
- [72] S. Käser, E. D. Boittier, M. Upadhyay, and M. Meuwly, Transfer learning to CCSD(T): Accurate anharmonic frequencies

- from machine learning models, *J. Chem. Theory Comput.* **17**, 3687 (2021).
- [73] C. T. Wolke, A. F. DeBlase, C. M. Leavitt, A. B. McCoy, and M. A. Johnson, Diffuse vibrational signature of a single proton embedded in the oxalate scaffold, $\text{HO}_2\text{CCO}_2^-$, *J. Phys. Chem. A* **119**, 13018 (2015).
- [74] S. Käser, J. O. Richardson, and M. Meuwly, Transfer learning for predictive molecular simulations: Data-efficient potential energy surfaces at CCSD(T) accuracy, *J. Chem. Theory Comput.* **21**, 6633 (2025).
- [75] C. Qu, A. Nandi, P. L. Houston, and J. M. Bowman, Tests demonstrate the accuracy of the Q_{im} -reaction path Hamiltonian for double-well tunneling splittings for H-atom transfer, *J. Phys. Chem. A* **130**, 3027 (2026).
- [76] J. Wang, J. C. S. V. Veliz, M. Upadhyay, and M. Meuwly, High-energy reaction dynamics of O_3 , *J. Chem. Phys.* **163**, 074109 (2025).
- [77] C. Qu, P. L. Houston, Q. Yu, A. Nandi, J. M. Bowman, V. Andreichev, S. Käser, and M. Meuwly, oxalate-2, 2026, <https://github.com/MMunibas/oxalate-2>.