Supplementary material for 'Machine Learning Small Polaron Dynamics'

Viktor C. Birschitzky,[1] Luca Leoni,[2] Michele Reticcioli,[1] and Cesare Franchini[1,2]

[1] *University of Vienna, Faculty of Physics and Center for Computational Materials Science, Vienna, Austria*
[2] *Department of Physics and Astronomy "Augusto Righi", Alma Mater Studiorum - Università di Bologna, Bologna, 40127 Italy*

I. DATA GENERATION

A. Density Functional Theory Methods

All computations were performed using the Vienna Ab initio Simulation package (VASP)[1–4] and using Perdew–Burke–Ernzerhof functionals[5] within the rotationally invariant DFT+U formalism[6]. The bulk systems taken into consideration are rocksalt MgO+h^+, rutile TiO_2+e^-, and F-doped rutile TiO_2 with a doping concentration of 1/192. In particular, to allow the simulation of polarons, we used a $2 \times 2 \times 2$ supercell for MgO (resulting in a supercell with $8.449 \times 8.449 \times 8.449 Å^3$ and 64 atoms) and a $2 \times 2 \times 6$ supercell for the other two TiO_2-systems (resulting in a supercell with $12.973 \times 12.973 \times 17.724 Å^3$ and 288 atoms). We used a Γ-centered $2 \times 2 \times 2$ k-points mesh for the former and Γ only mesh for the latter materials. Spin-polarized calculations were performed for every material and the energy cutoff was set to 300 eV for TiO_2-systems and 500 eV for MgO, respectively. We used standard projector augmented wave pseudopotentials[4] for MgO, whereas in TiO_2-compounds we used standard pseudopotentials for Ti (treating d- and s-orbitals as valence) and soft O and F pseudopotentials. Effective U values for the different materials were set to 10 eV for the p-orbitals of O atoms in MgOwithin the suggested range of U-values[7] and 3.9 eV for the d-orbitals of Ti atoms in TiO_2 materials as previously determined by constrained RPA[8]. Polarons were induced by either adding one excess electron for TiO_2, removing an electron in MgO, or by replacing 1 of the 192 O with an F-dopant in $TiO_{1.99}F_{0.01}$.

B. Initial Dataset

The training dataset for every material was constructed from frames of Molecular Dynamics (MD) runs in the NVT ensemble using Nosé-Hoover thermostat[9] with a 1 fs timestep and the computational setup described in the previous section. The systems were first thermalized in a shorter run performing a temperature ramping through a velocity rescaling that increased the temperature by 0.5 K at every step until the desired temperature. MgOdata were collected at 600 K, $TiO_2$and $TiO_{1.99}F_{0.01}$data at 500 K. The temperature was chosen to provide a sufficient number of polaron hopping events within the initial dataset generation. In particular, the chosen temperatures resulted in ∼20 hopping events for every material after the thermalization.

The 10 ps FPMD dataset consisting of 10^4 structures was first divided by reserving the last 20% of the frames as test set, while the training and validation set were constructed from the inital 80%. To sufficiently represent the possible hopping pathways in the training dataset, we extracted all jumps frames from the remaining 80%, where we selected 20 frames before and after maximal magnetization changed from one site to another. Randomly selected structures and two randomly selected hopping events were used in the validation set. The remaining hopping structures in combination with randomly selected structures were used for training. For train and validation datasets, a 1 to 2 proportion of hopping to random structures was used.

C. Active learning data generation

We extract hopping events from long initial MLMD runs performed with a model trained on the initial training dataset to enhance polaron transition sampling. This process might be unnecessary for simple materials with few symmetrically inequivalent transition pathways (i.e., MgO), but becomes necessary in more complex materials such as TiO_2. Polaron hopping events are extracted from the MLMD simulations by identifying timesteps in which argmax($\mathbf{m}$) changed. Similarly to the previous section, we extracted configurations of 40 timesteps centered around the hopping event in MLMD and fed them back into DFT. For every series of extracted configurations $\{\mathbf{r}_i\}_{i=1}^{40}$, we performed self consistent computations on all $\mathbf{r}_i$ using the charge density and wavefunction of the previous timestep $\mathbf{r}_{i-1}$, reproducing the operating principle of FPMD. Additionally, we diversified our sampling by extracting hopping events

RMSE	E in meV/Atom			$\mathbf{F}$ in meV/Å			$\mathbf{n}$		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
TiO_2+e^-	0.074	0.069	0.056	16.421	17.154	16.481	0.002	0.002	0.002
MgO+h^+	0.116	0.112	0.129	20.640	26.895	25.730	0.002	0.003	0.003
$TiO_{1.99}F_{0.01}$	0.114	0.139	0.133	17.651	19.591	15.871	0.002	0.002	0.001
	$\mathbf{F}_{\mathrm{pol}}$ in meV/Å			$n^{\uparrow}_{\mathrm{pol}} + n^{\downarrow}_{\mathrm{pol}}$			m_{pol}		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
TiO_2+e^-	38.330	46.186	48.038	0.010	0.009	0.011	0.032	0.028	0.034
MgO+h^+	29.199	45.035	44.592	0.003	0.005	0.004	0.012	0.020	0.015
$TiO_{1.99}F_{0.01}$	46.206	65.600	40.274	0.011	0.015	0.009	0.025	0.040	0.022

Supplementary Table I: Root mean squared errors of the models for energy E, forces $\mathbf{F}$ and occupations $\mathbf{n}$ and root mean squared errors of the models at the polaronic site for forces $\mathbf{F}_{\mathrm{pol}}$, the total occupation $n^{\uparrow}_{\mathrm{pol}} + n^{\downarrow}_{\mathrm{pol}}$, and the magnetization m_{pol}.

with distinct transition pathways (*e.g.*, regarding the crystallographic direction or the relation to other symmetry-breaking features such as dopants) and utilizing hoppings from various temperatures. Randomly sampled structures from this hopping database were then fed back to the training, validation, and test set to retrain the model. In cases of multiple symmetry-breaking features (such as in $TiO_{1.99}F_{0.01}$), extrapolation issues become more severe because many polaronic configurations are not sampled during the initial FPMD run. To address this, we augment the training database with snapshots of transitions to and from each site, as well as random snapshots of the polaron residing at a site for extended periods, ensuring coverage of all distinct polaronic configurations.

II. MACHINE LEARNING METHODS

A. Training

The model was optimized to fit energies E, forces $\mathbf{F}$ and the occupation of each site. While fitting energies and forces is a standard procedure to fit potential energy surface models, fitting the occupation requires additional loss terms. In particular, we fit the invariant quantity $\mathrm{Tr}[\mathbf{n}_i^\sigma]$, whose difference in spin channels $\sigma =\uparrow,\downarrow$ equal the magnetization of the i-th atom. For simplicity of notation, we shorten $\mathrm{Tr}[\mathbf{n}_i^\sigma]$ as n_i^σ, and the difference $n_i^\uparrow - n_i^\downarrow$ as the magnetization m_i. The total magnetization of a configuration $\sum_i m_i$ is labeled as M. We found that fitting and predicting $n_i^\uparrow$ and $n_i^\downarrow$ directly resulted in the most stable model. Remaining quantities such as the atomic magnetization m_i, the total magnetization M, or the total atomic occupation $n_i^\uparrow + n_i^\downarrow$ can be derived from these model outputs and are fitted by soft constraints implemented in our loss function. By assuming a batch of K structures, each containing N atoms, we can write the loss function $\mathcal{L}$ used to fit the model as follows:

$$\mathcal{L} = \frac{1}{K}\sum_{k=1}^{K}\Bigg\{\lambda_E\left(E_k^{\mathrm{ML}} - E_k^{\mathrm{DFT}}\right)^2 + \frac{\lambda_F}{3N}\sum_{i=1}^{N}\sum_{j=1}^{3}(F_{kij}^{\mathrm{ML}} - F_{kij}^{\mathrm{DFT}})^2 + \lambda_M\left(M_k^{\mathrm{ML}} - M_k^{\mathrm{DFT}}\right)^2 + \frac{\lambda_T}{N}\sum_{i=1}^{N}\left[\sum_{\sigma=\uparrow,\downarrow}(n_{ki}^{\sigma,\mathrm{ML}} - n_{ki}^{\sigma,\mathrm{DFT}})^2 + (m_{ki}^{\mathrm{ML}} - m_{ki}^{\mathrm{DFT}})^2 + \left(\sum_{\sigma=\uparrow,\downarrow} n_{ki}^{\sigma,\mathrm{ML}} - n_{ki}^{\sigma,\mathrm{DFT}}\right)^2\right]\Bigg\}.$$

The weight parameters λ are set based on whether they apply to global or atomic quantities. Following the same approach as the NequIP paper[10], we set the weight parameters of atomic quantities (i.e., λ_F for forces and λ_T for occupations) to N^2, while for global quantities (i.e., λ_E for energies and λ_M for the total magnetization) the weight is set to 1. Similarly to the suggestions of the paper we used a batch size of 1, which we found to produce the lowest fitting errors. The model was implemented in JAX[11] and optimization was performed using the optax-library's implementation of the Adam optimizer[12] with a learning rate of 5×10^{-4}. A decrease on plateau procedure was used during training – halving the learning rate if fitting errors did not improve in 5 consecutive epochs. Using this procedure the model usually converged after ∼7 hours of training on an Nvidia Ampere 100 graphic card, requiring on the order of $10^3 - 10^4$ epochs.

Supplementary Table I presents the achieved RMSE for key quantities within our test systems, which are within the expected range for the NequiP architecture[10].

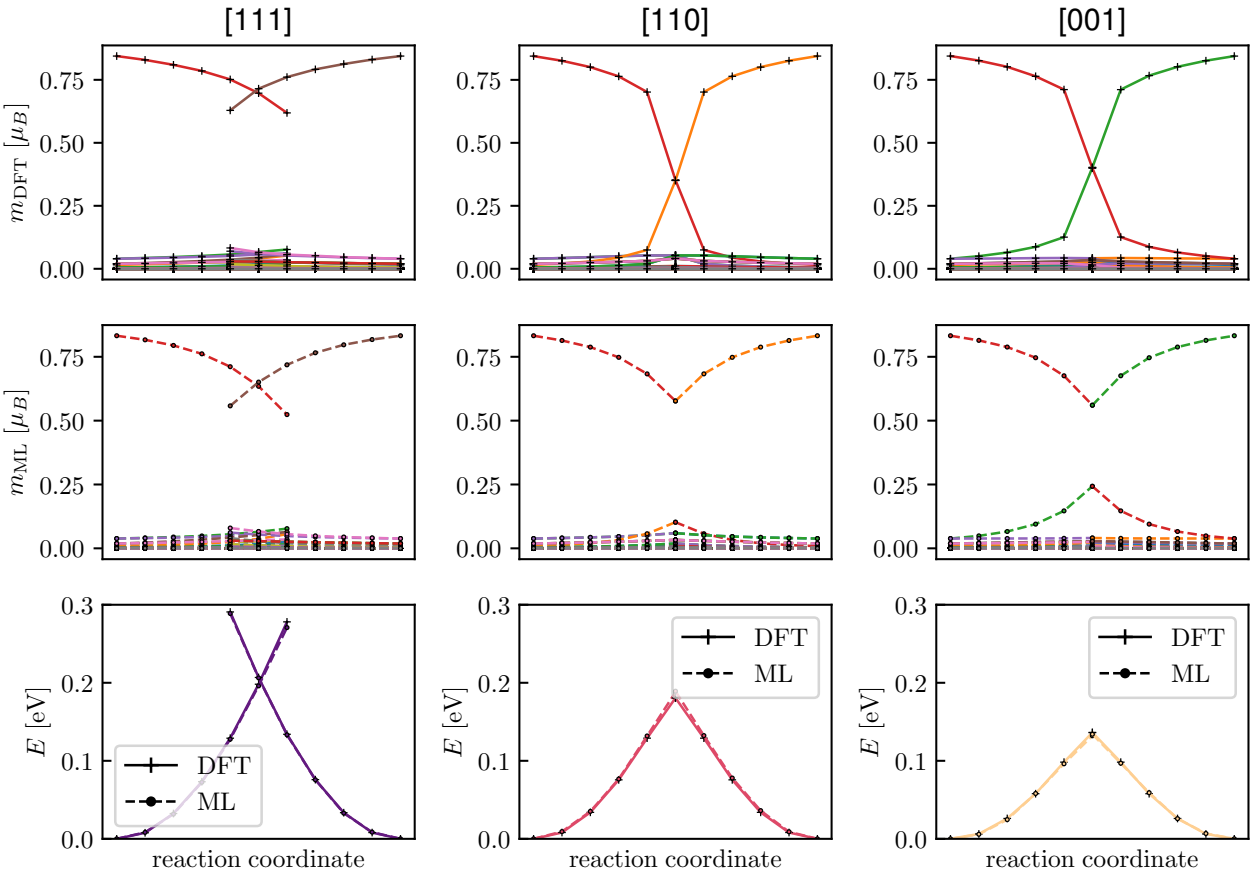
B. Machine-learned molecular dynamics

The machine-learned molecular dynamics (MLMD) was performed using the JAX-MD framework[13]. To be consistent with the training dataset every simulation was performed in the NVT ensemble using the Nosé-Hoover thermostat and with a timestep to 1 fs. Simulations were performed at various temperatures ranging from 100 K to 600 K in steps of 100 K. At 100 K polaron hopping was usually too infrequent to perform statistical analysis. Stable simulations for the full 10 ns where achieved for $TiO_{1.99}F_{0.01}$ and $MgO+h^+$ at every temperature, while MLMD simulations in TiO_2+e^- became unstable for 500 K and 600 K, still allowing to collect statistics for 4 ns and 1.3 ns, respectively.

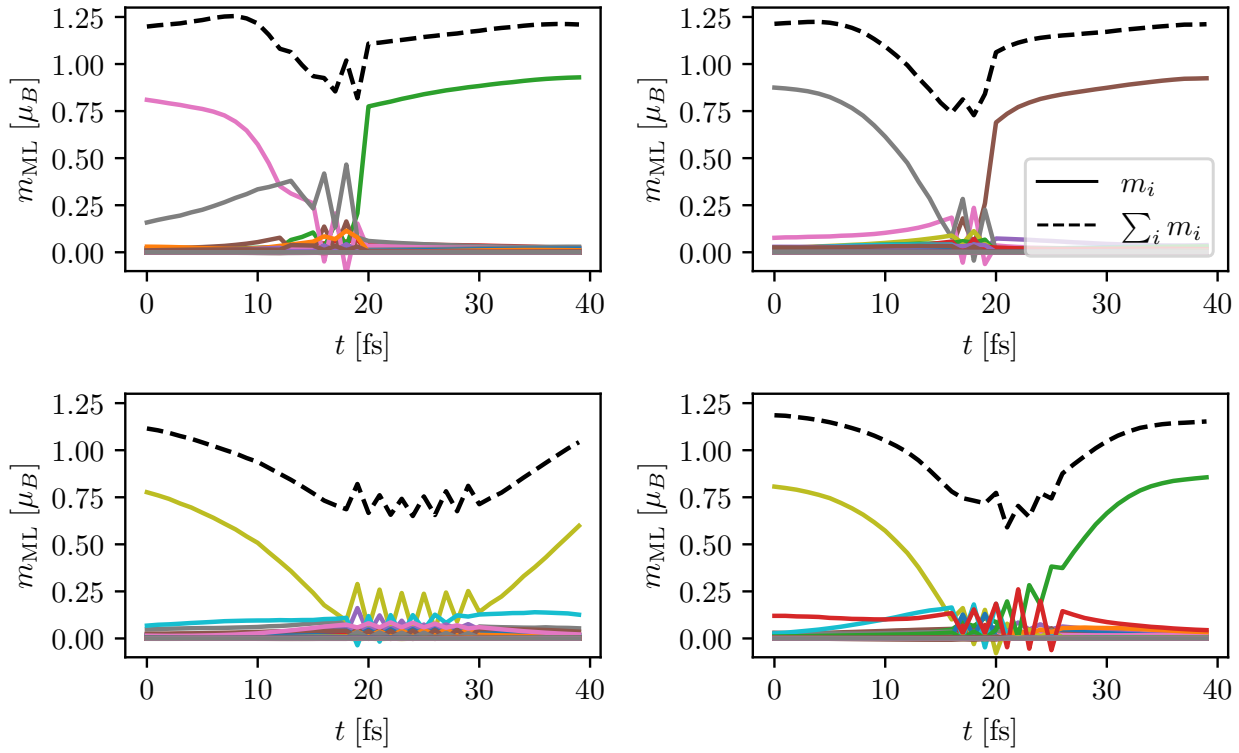
C. Hopping at the ML-level

Small polaron dynamics within FPMD is driven by thermally activated hops, where transitions occur in few time steps (tens of fs) of the dynamical simulation. Most of the time, the polaron remains localized at a specific site. During these intervals, our assumption of an integer charge state holds well, as the polaron can be assigned to a single site. However, during the transition steps of the rare hopping events, this approximation may break down. In these cases, the polaron hopping occurs through the coupling of the initial and final polaronic states, forming an adiabatic transition state where the charge is briefly delocalized across two sites. Additionally, due to thermal distortions, the polaronic band can approach the conduction band, resulting in partial occupations of the conduction band. During electronic minimization, this can lead to the localization of the polaron at a different site.

In both of these instances, our approximation complicates the explicit treatment of these events using machine learning, as the model is trained under the assumption that the polaron is localized on a single site. Nevertheless, we have found that this assumption produces the most reliable models in terms of long-term stability. The artifacts



Supplementary Figure 1: Comparison of DFT- and ML-predicted magnetizations m along different hopping pathways within TiO_2+e^-.



Supplementary Figure 2: Oscillations in ML-predicted magnetization m in TiO_2+e^-associated to partial delocalization of the polaron. The total magnetization $\sum_i m_i$ is shown in dashed black and indicates partial delocalization of the polaron.

resulting from this issue are evident in the figures below.

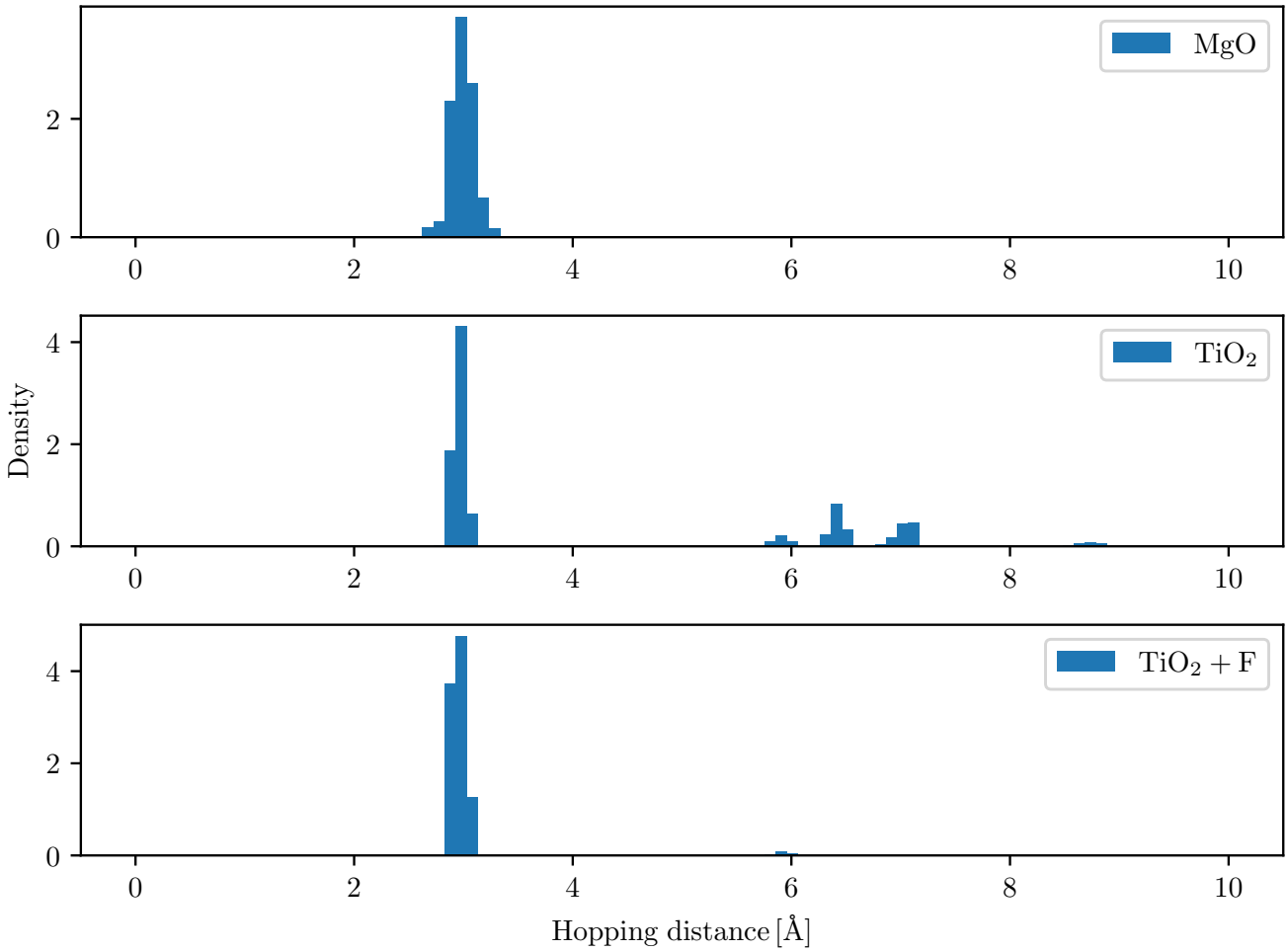
In Supplementary Figure 1 the magnetization along the reaction coordinate as predicted from DFT and ML for three distinct transition pathways ([001], [110], [111]) in rutile TiO_2+$e^-$are displayed. Reaction coordinates are determined by linear interpolation of the relaxed initial and final polaronic configurations:

$$\mathbf{r}_{\mathrm{int}} = \alpha \mathbf{r}_{\mathrm{init}} + (1-\alpha)\mathbf{r}_{\mathrm{final}} \tag{1}$$

It is apparent that for the [110] and [001] transitions, the central interpolated structure (i.e., $\alpha = 0.5$) cannot reproduce the desired shared state at the ML level, due to the assumption that the polaronic state is assigned to either site. This is reflected in the ML model, which produces two distinct sets of magnetizations at the central structure, depending on the assignment of the polaronic state. Nevertheless, we found that the model yields well-defined energies for these states, independent of the specific polaronic state. Furthermore, polaronic hopping still occurs during the MLMD, with well-reproducible magnetizations for these adiabatic transitions at the DFT level.

We also found that Leopold can reliably detect partial delocalization in the conduction band. This behavior is primarily associated with migrations along [110] and [111]. There, a drop of ML-predicted total magnetization during MLMD is an indicator for small energy differences between the polaronic and conduction band in TiO_2 at the DFT-level. In these cases the model might produce non-physical oscillations as depicted in Supplementary Figure 2. At DFT reference level a partial occupation of the conduction band might result in a favorable initialization of the electronic wavefunction to instantaneously transfer a polaron to a different site. In MLMD, we found that explicitly searching for a suitable polaron localization site did not offer any additional benefit. The simpler criterion outlined in the maintext in Figure 5 (assigning the polaron to the maximally magnetized site) was found to enhance long term stability of the simulations and reproducibly transfers polarons adiabatically at the DFT-reference level. We note that it might be beneficial to include a search for the most stable polaronic sites to mimic the behavior observed in the reference DFT simulations.

Interestingly, Leopold approximately reproduces the ratio of diabatic and adiabatic transitions in electron doped TiO_2. As previously demonstrated by Deskins through explicit calculations of the transfer rate[14], diabatic hopping along the [111] direction is expected to occur five orders of magnitude less frequently than adiabatic transport along



Supplementary Figure 3: Distance distribution of hopping distances in MgO, TiO_2, and $TiO_{1.99}F_{0.01}$as extracted from MLMD runs at 500K.

[001]. Although our model does not yield exact hopping rates, it reproduces the relative frequency of these events with reasonable accuracy. Specifically, in our simulations, hops along [111] occur three orders of magnitude less frequently than along [001], which is in qualitative agreement with prior findings given the approximations employed.

Supplementary Figure 3 collects the distribution of hopping distances as extracted from MLMD-runs at 500K for MgO+h^+, TiO_2+$e^-$and $TiO_{1.99}F_{0.01}$. Distant hops, as in TiO_2, are a feature of both FPMD and MLMD and affect the calculated electron mobility. The degree to which distant hops occur is system-dependent and appears to be related to the energy separation of the polaronic and conduction bands. While MgO shows predominantly adiabatic nearest neighbor hopping, rutile TiO_2+e^- displays a comparatively high proportion of delocalization-driven, long-distance hopping events, as partial filling of the conduction band often leads to favorable electronic initializations that drive an instantaneous transition of the polaron over longer distances (as discusseed above). In contrast, the binding interaction of the F-dopant in TiO_2 increases the separation of the polaronic and conduction band and distant hops, therefore, very rarely occur in this case. The effects of these long-distance hops are apparent in the comparatively high electron mobility in the TiO_2+e^-case.

III. POLARON MOBILITY

Resulting long MLMD simulations give access to the dynamical properties of the polaron, which we study through the computation of its mean squared displacement MSD(t), defined as:

$$\mathrm{MSD}(t) = \left\langle \frac{1}{T-t} \int_0^{T-t} [\mathbf{r}(t+\Delta) - \mathbf{r}(\Delta)]^2 \mathrm{d}\Delta \right\rangle. \tag{2}$$

Here, $\langle \cdot \rangle$ represents an ensemble average over all polarons in the system while $\mathbf{r}(t)$ are the polarons trajectories in a simulation with length T. Since only a single polaron is present in our system, the ensemble average in the formula can be omitted, as the dynamics of a single polaron fully characterize the charge transport properties in this case. From the obtained MSD we extracted the polaron diffusion coefficient D by using the known relationship

$$\mathrm{MSD}(t) = 2nDt, \tag{3}$$

where n is the dimensionality of the considered trajectory $\mathbf{r}$ (in our case $n = 3$). Following Eq. (3) the diffusion coefficient was extracted by linearly interpolating the MSD function obtained in the linear regime. Once D has been extracted, we can transform using the Einstein relation $\mu = qD/k_\mathrm{B}T$[15] to obtain the final estimate for the polaron mobility.

In this work, we also studied the angular dependence of the mobility in rutile $TiO_{1.99}F_{0.01}$. To obtain $\mu(\theta)$ within the (110)-plane containing the dopant, we projected the trajectory $\mathbf{r}(t)$ onto $\hat{\mathbf{n}}(\theta)$ giving access to the direction-dependent trajectory $r_\theta(t)$. We compute the MSD as follows:

$$\mathrm{MSD}(t,\theta) = \left\langle \frac{1}{T-t} \int_0^{T-t} [r_\theta(t+\Delta) - r_\theta(\Delta)]^2 \mathrm{d}\Delta \right\rangle \qquad \text{with} \qquad r_\theta(t) = \mathbf{r}(t) \cdot \hat{\mathbf{n}}(\theta). \tag{4}$$

In this way we reduce the MSD computation to a 1D trajectory in a certain direction in space, rendering Eq. (3) into the form

$$\mathrm{MSD}(t,\theta) = 2D(\theta)t, \tag{5}$$

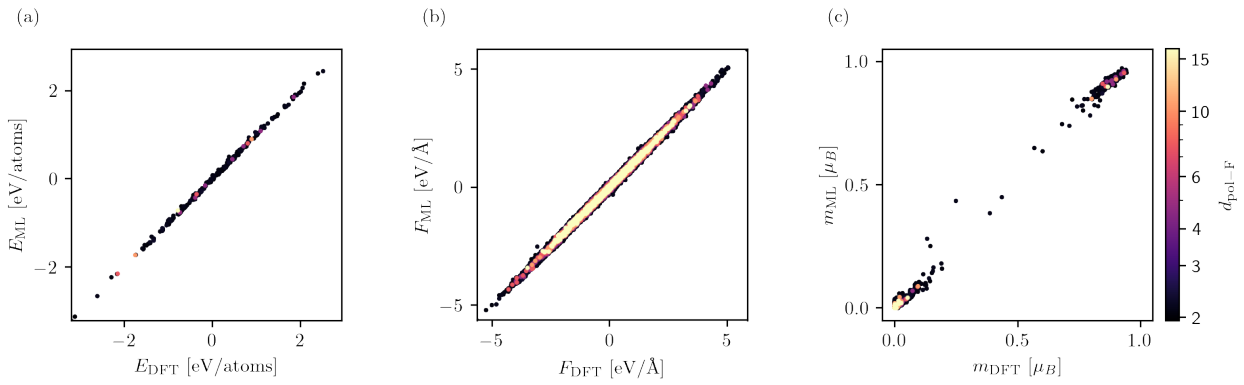
from which $D(\theta)$ can be extracted and transformed into $\mu(\theta)$ using the same approach described before.

IV. SCALABILITY AND LONG-RANGE INTERACTIONS

In our machine learning approach periodic DFT-results are reproduced, since periodic image interactions are effectively captured through the iterative message passing of NequIP. The message-passing incorporates interactions up to the cutoff radius multiplied by the number of message-passing steps (3 iterations and a cutoff of 5 Å in our case), capturing interaction of the first few periodic images. Previous studies[16–18] suggest that beyond approximately 10 Å, polaron-polaron interactions become negligible.

Clearly, long range interactions are only captured for the periodicity of the reference calculations. Since polaron-polaron interactions should become negligible after 10 Å scaling simulations is feasible for single polaron dynamics, as larger supercells result in decreasing the polaron concentration. This is illustrated in Supplementary Figure 4, which shows a comparison of DFT and ML prediction when simulating dynamics of a single polaron and F-dopant in a doubled F-doped TiO_2 cell. As the system size increases and the polaron and dopant are able to move further apart – beyond the separations achievable in the original supercell – Leopold accurately reproduces energies, forces, and magnetizations. This is emphasized by accurate predictions across various polaron-F distances occuring in the simulations, as indicated by the color scale.

A more challenging scenario for scaling polaronic or other defect-based ML simulations is the combinatorial explosion of possible configurations in larger supercells when a higher number of defects are present. A simple example is the doubling of the reference supercell in any of its spatial dimensions. Keeping polaron concentration fixed, this would double the number of polarons in the doubled cell. If MLMD is now performed on this doubled cell, one of the polarons might hop to a different site, effectively departing from the translational symmetry of the polaronic configuration of the reference data in the original supercell. Here, polaron-polaron interactions can become sizable due to the possibly reduced polaron-polaron distance. Since no reference data is available for this new polaron configuration, the model will necessarily extrapolate, if not retrained with this specific arrangement of polarons. Sampling these interactions comes with its own challenges[16–18] and are beyond the scope of this work.



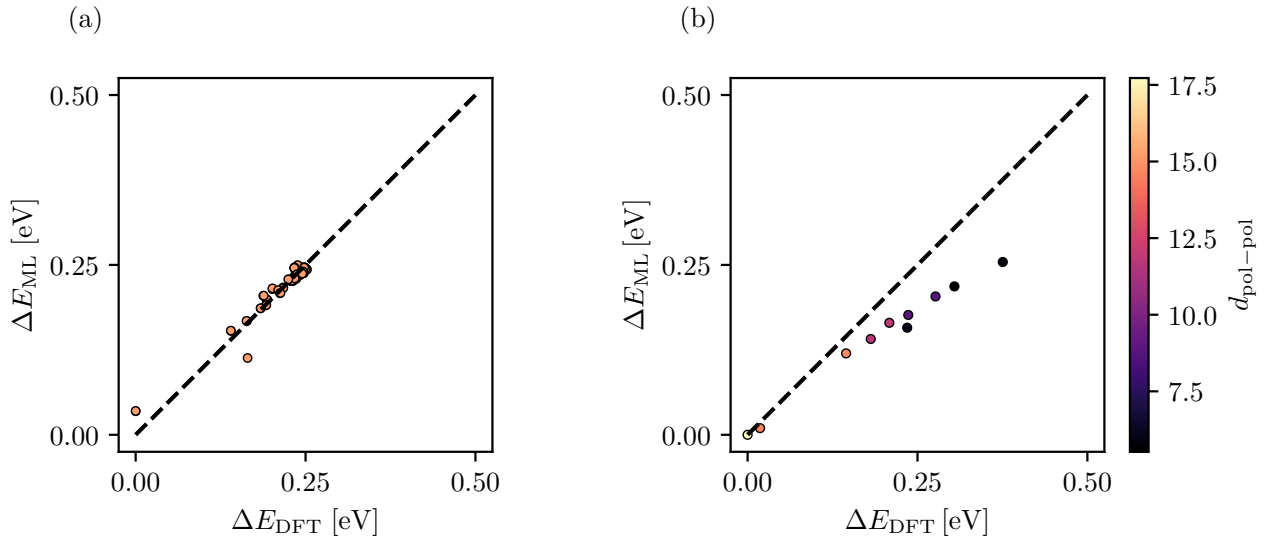
Supplementary Figure 4: Comparison of DFT- and ML-predicted (a) energies shifted by mean and rescaled by standard deviation, (b) forces, and (c) magnetizations as extracted from a scaled MLMD-simulation with a single polaron and F-dopant in a $2 \times 2 \times 12$-supercell (halving the polaron-/F-concentration). Simulations where performed at 500K for 250 ps and configurations extracted at regular intervals of 1ps and computed at DFT-level. During the simulation the polaron explored various configurations relative to the F-dopant, as indicated by the colorscale showing the polaron-F-distance $d_{\mathrm{pol-F}}$.

In the following, we demonstrate the capability to model configurational disorder, by examining F-doped TiO_2 as a case study, demonstrating that polaron-defect configurations can be accurately captured when properly sampled. By including training data, sampling each distinct polaron-F configuration, we demonstrate how Leopold can accurately reproduce the energy contribution of the occurring configurational disorder. Additionally, we explore errors when scaling simulations to larger supercells, by comparing predictions in unsampled polaron-polaron configurations. Both of these results should hold without loss of generality with respect to the defect type, demonstrating both the capability to model configurational disorder, as well as expected errors when not properly dealing with it.

Supplementary Figure 5 (a) shows a comparison of DFT- and ML-predicted polaron-F configurations in the $2\times2\times6$-supercell of $TiO_{1.99}F_{0.01}$ following a relaxation of each distinct arrangement (96 without considering symmetries) of defects (i.e. polaron and F-dopant) with Leopold and a successive DFT-calculation. The relaxation allows to investigate only the energy contribution of the configurational disorder by removing thermal noise. As expected, Leopold can reproduce the configurational disorder energy contribution, since reference data was included in the training procedure as outline above. Small deviations in some predictions most likely originate from the lack of low temperature reference data in the training of Leopold.

Supplementary Figure 5 (b) on the other hand shows a comparison of DFT- and ML-predicted polaron-polaron-configurations in a $2 \times 2 \times 12$-supercell of $TiO_{1.99}F_{0.01}$ (doubling of the cell in the [001] direction compared to the reference cell). To see the impact of extrapolating the polaron-polaron configuration, we leave the F-dopant position unchanged in both halves of the supercell (i.e. the translational symmetry of the F-dopants is identical to the refernce data). Polaron-polaron interaction is now sampled by fixing one of the polarons directly to one of the F-dopants, while the other polaron is moved along the same [001]-Ti-row (directly adjacent polarons are omitted for stability reasons). For computational efficiency, configurations are relaxed with Leopold and successively calculated at the DFT-level. To cancel possible effects from changing the cell size (e.g. k-point density), we align both DFT- and ML-energies of the configuration where both polarons are directly adjacent to the F-dopant (i.e. the combined polaron-F configuration reproducing the translational symmetry of the reference supercell) to 0 and show only energy differences to this configuration. As indicated by the colorscale, moving polarons closer together results in a systematic shift, where Leopold effectively underestimates the polaron-polaron repulsion – an expected error, since this interaction has not been sampled in the reference data. While in this case errors per atom are small (< 0.2 meV/atom), scaling simulation without explicitly sampling polaron-polaron interactions requires caution, as the error is expected to scale with the square of the number of polarons.

In conclusion, LEOPOLD is able to accurately scale the system sizes as long as the number of polarons remains fixed in the system, allowing large scale simulations for lower polaron densities respect to the training one. However, when scaling while keeping the polaron concentration fixed can instead lead to errors that remains manageable as long as polarons do not cluster together, which cannot be guaranteed in MD. Thus, scaling simulation while keeping the polaron concentration fixed requires carefully constructing reference databases to accurately account for configurational disorder. Additionally, adaption of the determination of the polaronic site in the proposed MD-routine



Supplementary Figure 5: Expected errors in capturing configuratinal disorder of polarons/defects in $TiO_{1.99}F_{0.01}$ through ML-relaxed polaron-F configurations. (a) DFT- and ML-based energies (ΔE_{DFT} and ΔE_{ML}), with the lowest DFT energy in the training dataset as baseline. Each datapoint corresponds to a polaron localized on one of the 96 Ti sites in the unscaled $2 \times 2 \times 6$ supercell, effectively capturing all polaron-F configurations. The training datasets includes reference data to each possible configuration, capturing the configurational disorder (b) DFT- and ML-based energies in a scaled $2 \times 2 \times 12$-supercell. DFT and ML-based energies of the configuration reproducing the translational symmetry of the $2 \times 2 \times 6$-supercell are aligned to 0. Each datapoint corresponds to two fixed F-dopants, and one fixed polaron, while the second polaron's position is different for each datapoint – effectively sampling the polaron-polaron interaction. The color scale encodes the polaron-polaron distance $d_{\mathrm{pol-pol}}$.

(based on the maximally magnetized site) is required, as multiple polaronic states need to be assigned.

[1] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[2] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996).
[3] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[4] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[5] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[6] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).
[7] S. Falletta and A. Pasquarello, npj Computational Materials **8**, 1 (2022), publisher: Nature Publishing Group.
[8] M. Setvin, C. Franchini, X. Hao, M. Schmid, A. Janotti, M. Kaltak, C. G. Van de Walle, G. Kresse, and U. Diebold, Physical Review Letters **113**, 086402 (2014), publisher: American Physical Society.
[9] S. Nosé, The Journal of Chemical Physics **81**, 511 (1984).
[10] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, Nature Communications **13**, 2453 (2022), publisher: Nature Publishing Group.
[11] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," (2018).
[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2017), arXiv:1412.6980 [cs.LG].
[13] S. S. Schoenholz and E. D. Cubuk, in *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Associates, Inc., 2020).
[14] N. A. Deskins and M. Dupuis, Physical Review B **75**, 195212 (2007).
[15] A. Einstein, *Investigations on the theory of the brownian movement* (Dover Publications, New York, 1956).
[16] V. C. Birschitzky, F. Ellinger, U. Diebold, M. Reticcioli, and C. Franchini, npj Computational Materials **8**, 1 (2022), number: 1 Publisher: Nature Publishing Group.
[17] V. C. Birschitzky, I. Sokolović, M. Prezzi, K. Palotás, M. Setvín, U. Diebold, M. Reticcioli, and C. Franchini, npj Computational Materials **10**, 1 (2024), publisher: Nature Publishing Group.
[18] Y. Zhang, Z.-K. Han, B. Zhu, X. Hu, M. Troppenz, S. Rigamonti, H. Li, C. Draxl, M. V. Ganduglia-Pirovano, and Y. Gao,

"Decoupling many-body interactions in CeO_2 (111) oxygen vacancy structure: Insights from machine-learning and cluster expansion," (2024), arXiv:2406.15956 [cond-mat.mtrl-sci].