

## Supplementary Materials

### **Data-efficient adaptation of DPA-4 force fields to DFT+*U* energetics: a case study in NiO**

**Fengyu Xie<sup>1</sup>, Peiheng Jiang<sup>2,\*</sup>, Zhicheng Zhong<sup>1,3</sup>**

<sup>1</sup>College of Artificial Intelligence and Data Science, Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, Jiangsu, China.

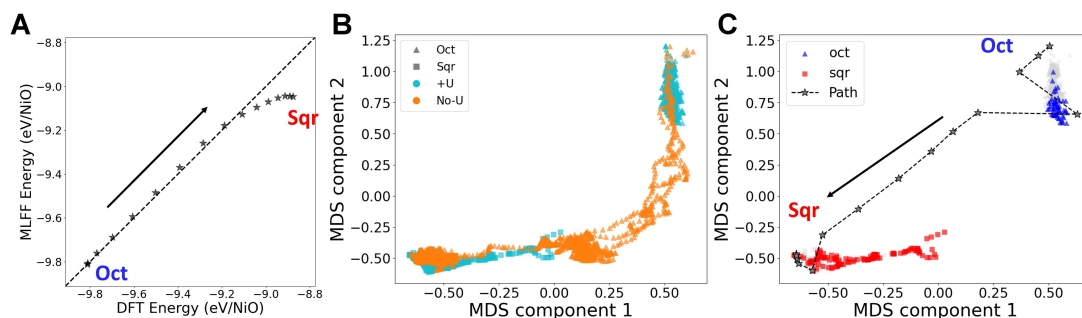
<sup>2</sup>School of Physics, MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China.

<sup>3</sup>Suzhou Lab, Suzhou 215123, Jiangsu, China.

**\*Correspondence to:** Dr. Peiheng Jiang, School of Physics, MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China. E-mail: [jiangph@xjtu.edu.cn](mailto:jiangph@xjtu.edu.cn)

## Supplementary Section 1. Analysis of the near-*Sqr* discrepancy

### 1.1 Endpoint error localization



**Supplementary Figure 1.** (A) Parity plot of configuration energies along the fixed *Oct*–*Sqr* interpolation path. The horizontal axis shows static energies obtained using the FM-initialized DFT+*U* protocol, and the vertical axis shows predictions of DPA4-Neo-OMat fine-tuned on the corresponding target dataset. The *Oct* and *Sqr* endpoints and path direction are indicated. (B) Two-dimensional MDS visualization of CrystalNN structural fingerprints for configurations from the FM-initialized DFT+*U* dataset and the non-spin-polarized PBE dataset. (C) MDS visualization of the target dataset, with the  $N_{\text{train}}=170$  configurations used for fine-tuning highlighted in color. The configurations in the validation and test sets are shown in grey. The fixed *Oct*–*Sqr* interpolation path is overlaid as connected stars.

Supplementary Figure 1A compares the energies predicted by the DPA4-Neo-OMat model fine-tuned on the FM-initialized DFT+*U* dataset with static DFT+*U* energies along the *Oct*–*Sqr* interpolation path. The MLFF remains close to the reference at the *Oct* endpoint and over most of the path, whereas the deviation increases sharply only near the *Sqr* endpoint.

### 1.2 Structure coverage

To assess whether the *Sqr* endpoint is simply a structural outlier, we examined the structural relationship between the sampled configurations in the AIMD datasets and the fixed *Oct*–*Sqr* interpolation path using CrystalNN-based fingerprints implemented in Matminer<sup>[1]</sup>. Multidimensional scaling (MDS) was used to approximately preserve pairwise distances between fingerprint vectors in a two-dimensional representation. We used matminer 0.10.1 with all its default settings to build local fingerprints for each site in a structure and aggregated the site fingerprints into structural fingerprints

by averaging over sites. MDS analysis was conducted using two components, four independent initializations, a Euclidean dissimilarity metric, and normalized stress. As shown in Supplementary Figure 1B and C, both the full target dataset and the selected  $N_{\text{train}} = 170$  subset populate the structural regions surrounding the *Oct* and *Sqr* basins. Indeed, parts of the intermediate and near-*Oct* interpolation path appear less densely represented than the vicinity of the *Sqr* basin, although the largest MLFF–DFT deviation occurs near *Sqr*. We emphasize that a two-dimensional MDS projection cannot establish complete coverage of the original high-dimensional descriptor space. However, it provides no evidence that the *Sqr* endpoint is merely an isolated structural outlier and motivates examining an additional source of electronic-structure-dependent error.

### 1.3 Electronic/magnetic multistability

We repeated static DFT+ $U$  calculations for exactly the same *Sqr* endpoint geometry while varying only the initial Ni magnetic moments. The Hubbard parameter was fixed at  $U = 6.2$  eV, the initial O moment was  $0.01 \mu_B$ , and the electronic convergence threshold was  $10^{-6}$  eV. The results are summarized in Supplementary Table 1.

**Supplementary Table 1.** Static PBE+ $U$  calculations for the same *Sqr* endpoint geometry using different initial Ni magnetic moments. The unit cell contains two NiO formula units. The initialization marked by an asterisk corresponds to that used to obtain the static reference curve in the main text.

| Method/Initial Ni moments ( $\mu_B$ ) | Final Ni moments ( $\mu_B$ ) | Final energy (eV/NiO f.u.) |
|---------------------------------------|------------------------------|----------------------------|
| PBE+ $U$ /(3.0, 3.0)*                 | (1.756, 1.756)               | -8.876                     |
| PBE+ $U$ /(1.0, 1.0)                  | (0.000, 0.000)               | -9.164                     |
| PBE+ $U$ /(0.0, 3.0)                  | (0.000, 1.350)               | -8.964                     |
| Fine-tuned MLFF                       | N/A                          | -9.047                     |

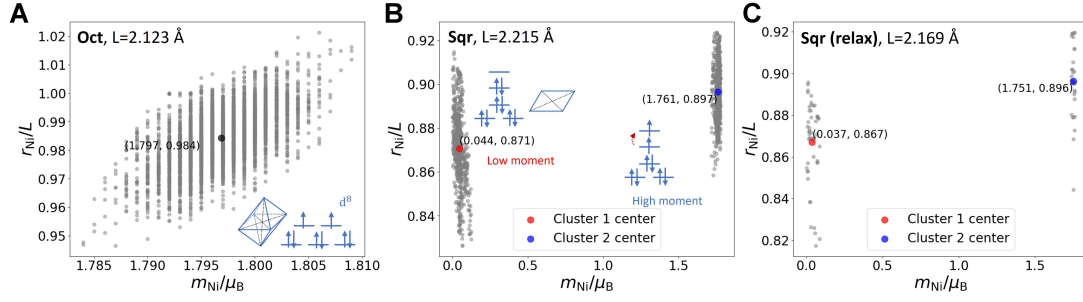
To test whether the multistability identified above could arise from insufficient SCF convergence, we repeated representative calculations with a tighter electronic convergence threshold of  $10^{-8}$  eV. As summarized in Supplementary Table 2, the three distinct self-consistent solutions persist and retain the same energetic ordering. The

multistability therefore cannot be attributed simply to the  $10^{-6}$  eV SCF convergence threshold used in the original dataset.

**Supplementary Table 2.** Static PBE+ $U$  calculations for the same *Sqr* endpoint geometry using different initial Ni magnetic moments, performed using a substantially tighter SCF convergence threshold of  $10^{-8}$  eV. The unit cell contains two NiO formula units. The initialization marked by an asterisk corresponds to that used to obtain the static reference curve in the main text.

| Method/Initial Ni moments ( $\mu_B$ ) | Final Ni moments ( $\mu_B$ ) | Final energy (eV/NiO f.u.) |
|---------------------------------------|------------------------------|----------------------------|
| PBE+ $U$ /(3.0, 3.0)*                 | (1.324, 1.335)               | -8.806                     |
| PBE+ $U$ /(1.0, 1.0)                  | (0.000, 0.000)               | -9.164                     |
| PBE+ $U$ /(0.0, 3.0)                  | (0.008, 1.350)               | -8.965                     |

These calculations provide direct evidence that the identical atomic geometry supports multiple self-consistent DFT+ $U$  solutions with substantially different local moments and energies. The energy span among these solutions is approximately 0.288 eV/f.u., exceeding the 0.171 eV/f.u. residual difference under discussion. Thus, the *Sqr* reference energy is strongly dependent on the converged electronic/magnetic branch, even at fixed atomic species, cell, and coordinates. This distinction is directly relevant to the MLFF representation. The present MLFF maps atomic species and coordinates onto a single energy and contains no explicit electronic-state or magnetic-state variable. **Consequently, disagreement with one selected static solution in a multistable region cannot be interpreted in exactly the same way as an ordinary interpolation error on a unique, single-valued potential-energy surface.**



**Supplementary Figure 2.** Distribution of local Ni magnetic moments ( $m_{\text{Ni}}$ ) and normalized local average Ni–O distance ( $r_{\text{Ni}}/L$ ), where  $r_{\text{Ni}}$  is the average Ni–O distance within the selected coordination shell and the normalization factor  $L$  is the cubic root of average volume per atom, i.e.,  $L = \sqrt[3]{V/n_{\text{atoms}}}$ . (A) *Oct* configurations sampled from 300 K FM-initialized DFT+*U* AIMD. (B) *Sqr* configurations sampled under the same conditions as in (A). (C) Final configurations from supercell *Sqr* relaxations initialized with randomly symmetry-broken Ni moments. Red and blue markers indicate the centers of the two magnetic populations. Schematics illustrate possible  $d^8$  configurations consistent with the observed low- and high-moment populations.

We next examined whether this behavior is limited to deliberately varied magnetic initialization in a small static cell. Supplementary Figure 2 analyzes the local Ni moments together with normalized local average Ni–O distance in the FM-initialized DFT+*U* AIMD data. The *Oct* configurations form essentially one high-moment population centered near  $1.8 \mu_{\text{B}}$ . In contrast, the *Sqr* configurations exhibit two populations, a low-moment population near zero and a high-moment population near  $1.76 \mu_{\text{B}}$ . Their corresponding local Ni–O distances overlap substantially. A similar overlapping bimodal pattern persists in zero-temperature supercell relaxations initialized with randomly symmetry-broken Ni moments [Supplementary Figure 2C]. Supplementary Table 1 therefore provides direct evidence of multistability at identical geometry, while Supplementary Figure 2 shows that related magnetic bimodality persists across closely overlapping *Sqr* local environments in both AIMD and relaxed supercells.

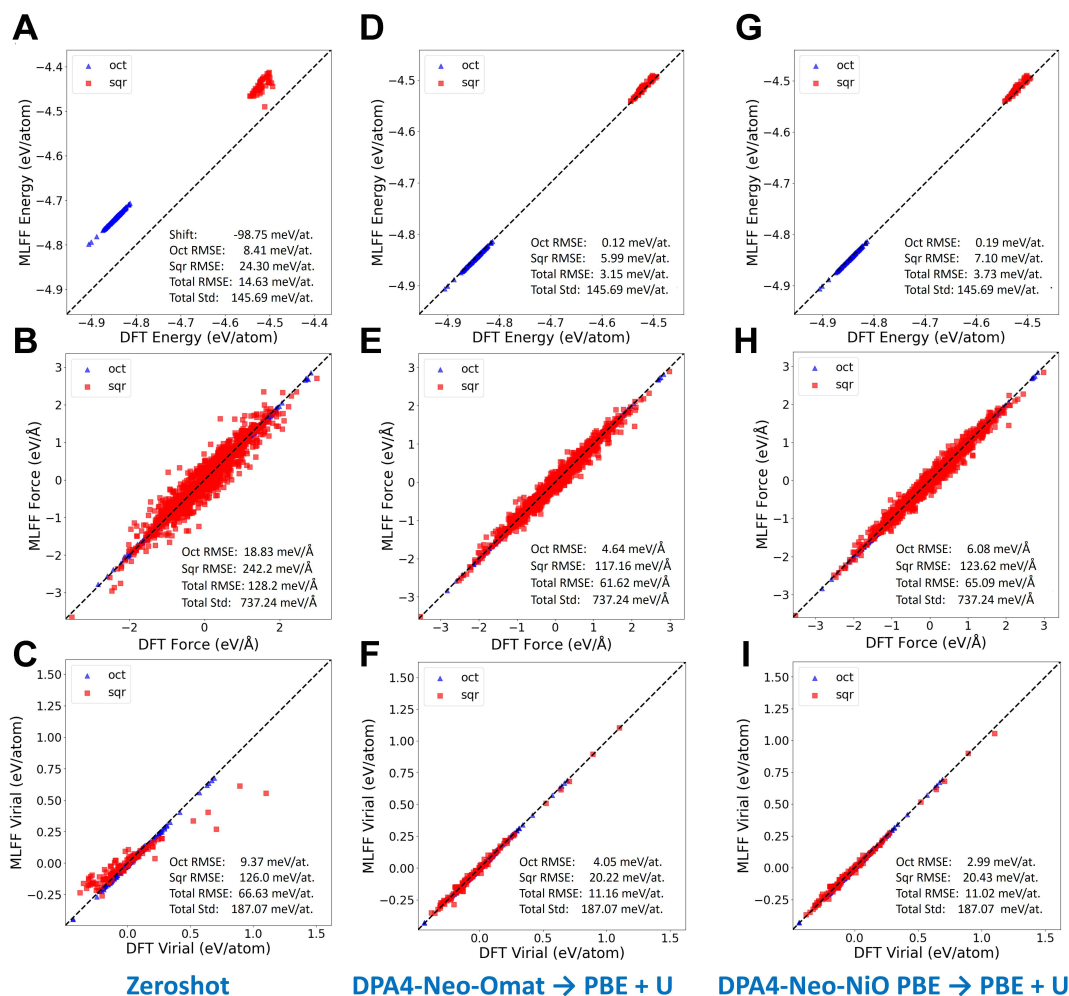
The two populations qualitatively suggest low-spin-like and high-spin-like solutions in the square-planar ligand field. We present this only as a physical interpretation, because a definitive microscopic assignment would require orbital-resolved

occupations and dedicated electronic-structure analysis beyond the scope of the present MLFF study. The conclusion drawn here is narrower: **the *Sqr* region exhibits pronounced electronic/magnetic branch sensitivity over identical or closely related atomic environments.**

In summary, Supplementary Table 1 and Supplementary Figure 2 suggest that *Sqr* NiO may lie in a particularly delicate regime in which crystal-field splitting, exchange–correlation interactions, local lattice distortion, and magnetic state are strongly coupled. Small changes in geometry or electronic initialization can lead to qualitatively different self-consistent states and shift the DFT+*U* energy substantially. Most importantly, identical *Sqr* coordinates support multiple solutions spanning 0.288 eV/f.u., exceeding the 0.171 eV/f.u. MLFF–DFT discrepancy. **We therefore identify electronic-state multistability and branch sensitivity as the most plausible origin of the discrepancy near the *Sqr* endpoint.**

**Resolving which competing electronic state is physically relevant or energetically preferred is a separate electronic-structure problem and it is not required to establish the central conclusion of this work.** Both direct and sequential adaptation routes use exactly the same target configurations, labels, data partition, and fine-tuning protocol. Despite prior adaptation to a source dataset with the opposite *Oct*–*Sqr* ordering, the sequentially adapted model reaches essentially the same target errors and energetic ordering as the directly adapted model. **Thus, the controlled comparison provides a clear answer to the question posed in this work: prior adaptation to the opposing source electronic-structure protocol does not prevent efficient subsequent adaptation to the target protocol.**

## Supplementary Section 2. Validation of MLFFs on independent AIMD trajectories



**Supplementary Figure 3.** Parity plots of independent-trajectory validation of DPA4-Neo models on DFT+*U* NiO at 600 K. Energy (top row), force (middle row), and virial (bottom row) predictions are compared against DFT+*U* reference values for independently initialized AIMD trajectories of *Oct* and *Sqr* NiO that were not used during fine-tuning. (A–C) Zero-shot predictions of the pretrained DPA4-Neo-OMat model. (D–F) Predictions after direct fine-tuning of DPA4-Neo-OMat on the DFT+*U* training data. (G–I) Predictions after sequential adaptation, in which DPA4-Neo-OMat was first fine-tuned on non-spin-polarized PBE NiO data and subsequently fine-tuned on the DFT+*U* data. Blue triangles and red squares denote *Oct* and *Sqr* configurations, respectively. Dashed lines indicate perfect agreement with the DFT+*U* references. RMSE values for the individual phases and the combined dataset are reported in each panel. The zero-shot energy error is reported as the test-set energy-offset-aligned zero-shot RMSE, obtained by applying a uniform per-atom energy shift

determined from the mean DFT–MLFF energy difference over the corresponding test set. The standard deviation of the corresponding DFT+ $U$  reference quantity is also provided for comparison.

To assess whether sequential fine-tuning also restores the MLFF to the target protocol on configurations that are not directly correlated with the original DFT+ $U$  AIMD training data, we performed validation on independently generated AIMD trajectories. Specifically, we generated two independently initialized DFT+ $U$  AIMD trajectories, one initialized from *Oct*-NiO and the other from *Sqr*-NiO, at 600 K and 1 bar using the same computational settings as the original trajectories described in the main text. Both simulations used a time step of 4.0 fs. The *Oct* trajectory was run for 192 steps, all of which were electronically converged, whereas the *Sqr* trajectory was run for 80 steps, yielding 73 electronically converged frames. These new trajectories were used only for evaluation, while all fine-tuned models were trained exclusively on the original dataset. Supplementary Figure 3 compares the zero-shot DPA4-Neo-OMat model, the model directly fine-tuned on DFT+ $U$ , and the model sequentially adapted from non-spin-polarized PBE to PBE+ $U$ .

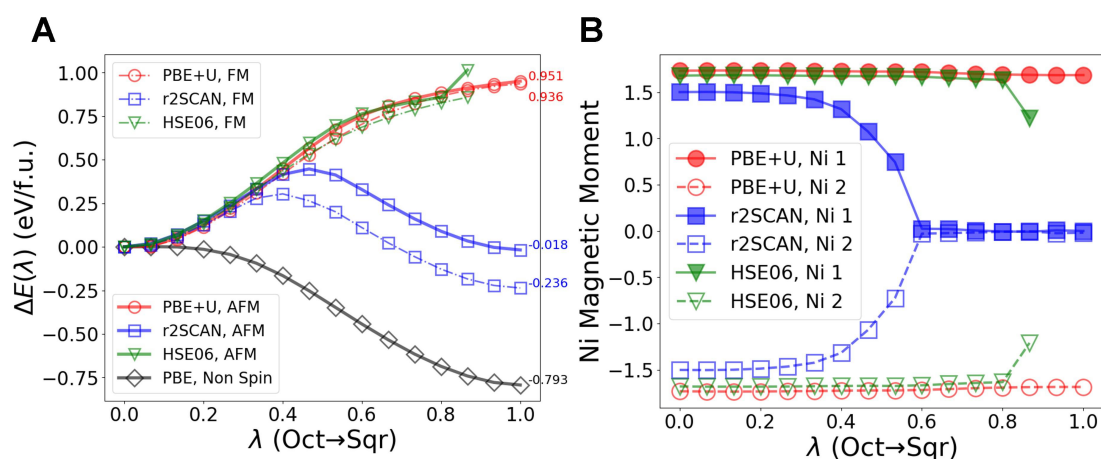
On these independent trajectories, direct and sequential adaptation reduce the total energy RMSE from 14.63 meV/atom to 3.15 and 3.73 meV/atom, respectively; the force RMSE from 128.2 meV/Å to 61.62 and 65.09 meV/Å; and the virial RMSE from 66.63 meV/atom to 11.16 and 11.02 meV/atom. Although these errors are higher than those obtained from the original trajectory-based splits, both adaptation routes still provide substantial improvements over the zero-shot model and preserve the qualitative target-protocol *Oct*–*Sqr* energetic ordering. These results show that the observed adaptation is not merely a consequence of correlations between neighboring AIMD frames, and that target-protocol fine-tuning remains effective for configurations sampled from independently initialized trajectories outside the original training trajectories.

The remaining independent-test error can be predominantly attributed to the *Sqr* configurations. As discussed in Section 1.3, *Sqr* NiO exhibits pronounced coupling between local geometry and electronic/magnetic multistability, such that identical or closely related geometries may converge to distinct DFT+ $U$  branches. These labels can appear intrinsically less single-valued and therefore harder to fit to a coordinate-only MLFF than those of the *Oct* configurations.

## Supplementary Section 3. Electronic-structure energetics and magnetic evolution along the *Oct*–*Sqr* interpolation

### 3.1 Comparison among functionals and magnetic initializations

To provide additional electronic-structure context for the source and target labeling protocols used in the main text, we compared the relative energies of NiO configurations along the same fixed structural interpolation between the octahedrally coordinated (*Oct*) and square-planar (*Sqr*) structures using several electronic-structure treatments. The *Oct* and *Sqr* endpoint structures were first relaxed using non-spin-polarized PBE, and intermediate configurations were generated by linearly interpolating both the atomic coordinates and lattice matrices. The resulting configurations were then evaluated by single-point calculations using PBE+*U*, r2SCAN, and HSE06 under ferromagnetic (FM) and antiferromagnetic (AFM) initializations. Given the electronic/magnetic complexity, we used “FM-initialized” and “AFM-initialized” to distinguish only the imposed magnetic initialization protocol rather than implying that all Ni sites necessarily retain the correspondingly ordered high-moment magnetic state after self-consistency. Because the same set of atomic geometries is used, the comparison reflects differences among the electronic-structure treatments rather than method-dependent structural relaxation. The interpolation should therefore be regarded as a fixed configurational path rather than as a minimum-energy path or a determination of equilibrium phase stability.



**Supplementary Figure 4.** Electronic-structure energetics and site-resolved magnetic moments along the fixed *Oct*–*Sqr* interpolation path in NiO. (A) Energy per NiO formula unit relative to the *Oct* configuration as a function of the interpolation coordinate  $\lambda$ , where  $\lambda=0$  and 1 correspond to the PBE-relaxed *Oct* and *Sqr* endpoints,

respectively. PBE+ $U$ , r2SCAN, and HSE06 results are shown for FM- and AFM-initialized calculations, while non-spin-polarized PBE is shown as the reference source-level profile. All energies are evaluated on the same fixed set of configurations. The final two HSE06 points are omitted because consistent magnetic solutions could not be converged. (B) Site-resolved local magnetic moments of the two Ni sites along the interpolation for the AFM-initialized calculations. Marker colors and shapes follow panel (A); filled and open symbols denote Ni 1 and Ni 2, respectively.

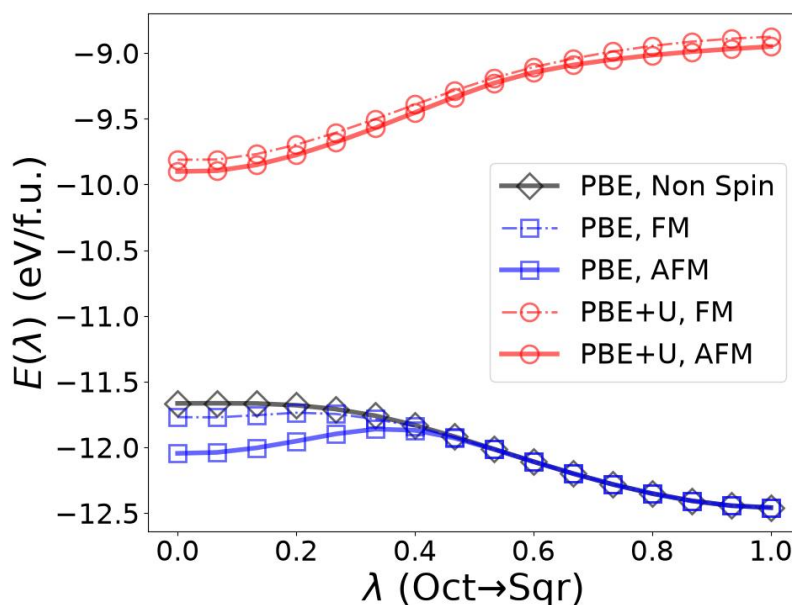
Supplementary Figure 4A shows the corresponding relative-energy profiles. The energetic preference between *Oct* and *Sqr* varies substantially with the electronic-structure treatment. At the two fixed endpoint geometries, non-spin-polarized PBE places *Sqr* 0.793 eV/f.u. below *Oct*. PBE+ $U$  reverses this ordering strongly, placing *Sqr* 0.936 and 0.951 eV/f.u. above *Oct* for the FM- and AFM-initialized calculations, respectively. The FM- and AFM-initialized PBE+ $U$  curves remain close to each other over most of the interpolation and give essentially the same qualitative energetic preference. On this basis, the consistently generated FM-initialized PBE+ $U$  data were used as the target reference for the MLFF adaptation experiments. HSE06 shows a qualitatively similar energetic evolution to PBE+ $U$  over the portion of the interpolation for which consistent magnetic solutions could be obtained, with the energy increasing substantially toward the *Sqr* side. The final two HSE06 configurations were excluded because consistent magnetic solutions could not be converged. In contrast, r2SCAN exhibits substantially different behavior. Under both FM and AFM initializations, the relative energy first increases away from *Oct* but subsequently decreases as the *Sqr* geometry is approached. At the *Sqr* endpoint, r2SCAN places *Sqr* 0.018 eV/f.u. below *Oct* under FM initialization and 0.236 eV/f.u. below *Oct* under AFM initialization, in qualitative disagreement with the PBE+ $U$  profiles and with the available HSE06 results.

The magnetic-moment evolution provides further context. As shown in Supplementary Figure 4B, the PBE+ $U$  and HSE06 calculations retain relatively large local Ni moments throughout the interpolation. By comparison, the local Ni moments obtained with r2SCAN decrease sharply in the intermediate region and approach zero toward the *Sqr* side. The change in the r2SCAN energetic profile occurs concurrently with a transition to a substantially different magnetic solution. This behavior suggests

strong coupling between the predicted energetic landscape and the electronic/magnetic state sampled by a given electronic-structure treatment. Thus, although r2SCAN-based pretraining has recently been proposed for improving the thermodynamic and zero-shot accuracy of foundation MLFFs, system-specific validation and target-protocol adaptation remain important, particularly in strongly correlated systems.

### 3.2 Spin-polarized PBE control calculations

To distinguish the effects of spin polarization from those of the Hubbard  $U$  correction, we performed additional spin-polarized PBE calculations without  $U$  along the same fixed  $Oct-Sqr$  structural path.



**Supplementary Figure 5.** Absolute energies along the fixed structural interpolation path from octahedral ( $Oct$ ,  $\lambda=0$ ) to square-planar ( $Sqr$ ,  $\lambda=1$ ) NiO, evaluated using non-spin-polarized PBE, spin-polarized PBE with FM and AFM initializations, and PBE+ $U$  with FM and AFM initializations. For all spin-polarized calculations shown here, the magnitude of the initial Ni magnetic moment was set to  $3.0 \mu_B$ , corresponding to the (3.0, 3.0) initialization listed in the accompanying Supplementary Tables 2 and 3 for the FM case, with opposite signs assigned for the AFM initialization. The spin-polarized PBE curves converge toward the same low-energy nonmagnetic branch near the  $Sqr$  endpoint and closely follow the non-spin-polarized PBE result, whereas PBE+ $U$  produces a qualitatively different and substantially higher-energy branch along the same path.

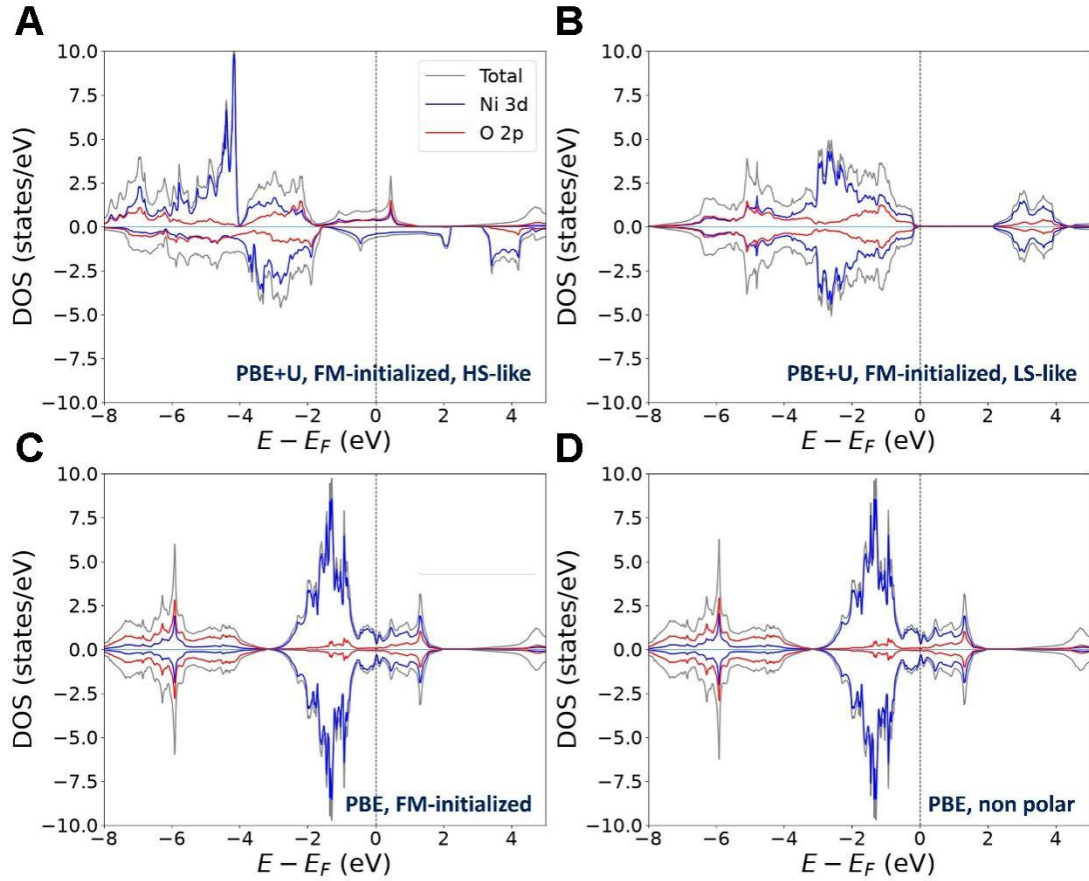
The new results show that spin polarization alone does not reproduce the qualitative energetic reversal observed between non-spin-polarized PBE and FM-initialized PBE+ $U$  (Supplementary Figure 5). In addition, at the *Sqr* endpoint (Supplementary Table 3), a range of initial magnetic configurations, including FM, AFM, asymmetric, and low-moment initializations, all converge under PBE to essentially the same nonmagnetic solution with the same final energy of -12.459 eV per NiO formula unit. Thus, the distinct electronic multistability observed for *Sqr*-NiO under PBE+ $U$  cannot be attributed simply to the inclusion of spin polarization.

**Supplementary Table 3.** Static PBE calculations without  $U$  for the same *Sqr* endpoint geometry using different initial Ni magnetic moments, performed with a SCF convergence threshold of  $10^{-8}$  eV. The unit cell contains two NiO formula units.

| Method/Initial Ni moments ( $\mu_B$ ) | Final Ni moments ( $\mu_B$ ) | Final energy (eV/NiO f.u.) |
|---------------------------------------|------------------------------|----------------------------|
| PBE/(3.0, 3.0)                        | (0.000, 0.000)               | -12.459                    |
| PBE/(1.0, 1.0)                        | (0.000, -0.001)              | -12.459                    |
| PBE/(0.0, 3.0)                        | (0.000, 0.000)               | -12.459                    |
| PBE/(-3.0, 3.0)                       | (0.000, 0.000)               | -12.459                    |

### 3.3 Representative PDOS of *Sqr*-NiO

Supplementary Figure 6 provides a basic electronic-structure characterization of *Sqr*-NiO under different DFT treatments and magnetic-moment initializations. The PBE *Sqr*-NiO calculations converge to metallic and essentially nonmagnetic solutions. In contrast, PBE+ $U$  calculations for *Sqr*-NiO at  $U = 6.2$  eV can converge to distinct self-consistent states depending on the magnetic initialization, including a metallic HS-like state obtained from initial Ni magnetic moments of (3.0, 3.0)  $\mu_B$ , and a semiconducting LS-like state obtained from (1.0, 1.0)  $\mu_B$ , with a band gap of approximately 2 eV.



**Supplementary Figure 6.** Projected density of states (PDOS) of the *Sqr*-NiO structure under different electronic-structure treatments and self-consistent solutions. (A) PBE+*U*, FM-initialized, HS-like solution obtained from initial Ni moments of (3.0, 3.0)  $\mu_B$ ; (B) PBE+*U*, FM-initialized, LS-like solution obtained from (1.0, 1.0)  $\mu_B$  initialization; (C) spin-polarized PBE with a (3.0, 3.0)  $\mu_B$  FM initialization; and (D) non-spin-polarized PBE. The grey, blue, and red curves denote the total DOS and the projected Ni 3*d* and O 2*p* contributions, respectively. The Fermi level is set to  $E - E_F = 0$ . Positive and negative DOS values in the spin-polarized panels denote the two spin channels.

## **Supplementary Section 4. VASP settings**

### **4.1 Static FM-initialized PBE+*U* calculations on the fixed interpolation path**

ALGO = All  
TIME = 0.1  
EDIFF = 1e-06  
ENCUT = 520.0  
IBRION = -1  
ISMEAR = 0  
ISPIN = 2  
LASPH = True  
LDAU = True  
LDAUJ = 0 0  
LDAUL = 2 0  
LDAUU = 6.2 0  
LDAUPRINT = 1  
LDAUTYPE = 2  
LMAXMIX = 4  
LORBIT = 11  
LREAL = False  
LWAVE = False  
MAGMOM = 2\*3.0 2\*0.01  
NELM = 300  
NSW = 0  
PREC = Accurate  
SIGMA = 0.05  
AMIX = 0.1  
BMIX = 0.0001  
AMIX\_MAG = 0.2  
BMIX\_MAG = 0.0001  
NPAR = 4

### **4.2 FM-initialized PBE+*U* AIMD calculations at 300 K**

ALGO = All  
AMIX = 0.1  
AMIX\_MAG = 0.2

BMIX = 0.0001  
BMIX\_MAG = 0.0001  
EDIFF = 1e-06  
ENCUT = 520.0  
IBRION = 0  
ISIF = 3  
ISMEAR = 0  
ISPIN = 2  
LANGEVIN\_GAMMA = 10.0 10.0 10.0  
LANGEVIN\_GAMMA\_L = 10.0  
LASPH = True  
LCHARG = False  
LDAU = True  
LDAUJ = 0 0  
LDAUL = 2 0  
LDAUPRINT = 1  
LDAUTYPE = 2  
LDAUU = 6.2 0  
LMAXMIX = 4  
LORBIT = 11  
LREAL = False  
LWAVE = False  
MAGMOM = 16\*3.0 16\*0.01  
MDALGO = 3  
NELM = 300  
NPAR = 4  
NSW = 250  
PMASS = 1000.0  
POTIM = 4.0  
PREC = Accurate  
PSTRESS = 0.001  
SIGMA = 0.05  
TEBEG = 300.0  
TIME = 0.1

### 4.3 Pseudopotential and k-point settings

Throughout this study, we used pymatgen’s default MPRelaxSet pseudopotential settings for the POT\_GGA\_PAW\_PBE set (corresponding to the 06-05-2010 PBE PAW dataset in VASP documentation) together with its automatic k-point generation scheme. Specifically, all calculations employed the Ni\_pv potential (PAW\_PBE Ni\_pv 06Sep2000), with the  $3p^6 3d^8 4s^2$  electrons treated as valence, and the regular O potential (PAW\_PBE O 08Apr2002), with the  $2s^2 2p^4$  electrons treated as valence. For the interpolated unit-cell configurations, self-consistent static calculations were first carried out on a  $\Gamma$ -centered  $8 \times 8 \times 5$  k-point grid to obtain the energies, magnetic moments, and charge densities. Subsequent non-self-consistent calculations reused the converged charge densities and employed a denser  $\Gamma$ -centered  $16 \times 16 \times 10$  k-point grid to obtain the PDOS in Supplementary Figure 6. For all AIMD calculations,  $\Gamma$ -centered k-point grids of  $4 \times 4 \times 2$  and  $3 \times 5 \times 2$  were used for the *Oct* and the *Sqr* phases, respectively.

### 4.4 AIMD data filtering

For the original PBE+U AIMD dataset, three *Sqr*-NiO trajectories were generated at 300, 600, and 900 K, with 250 MD steps attempted for each trajectory. Because electronic self-consistency was substantially more difficult to achieve for *Sqr*-NiO, only 250 frames in total across the three trajectories satisfied the electronic-convergence criterion and were retained in the final dataset. Frames that did not reach electronic self-consistency were excluded from subsequent MLFF training, validation, and testing.

## Supplementary Section 5. Detailed RMSE analysis for directly fine-tuned and sequentially adapted MLFFs

### 5.1 Split-wise test-set RMSEs at $N_{\text{train}} = 170$ in main Figure 2 and Figure 3

The following tables report the exact test-set RMSEs at  $N_{\text{train}} = 170$  for each predefined data split. Energy and virial RMSEs are reported in meV/atom, and force RMSEs in meV/Å. The final row for each model gives the mean  $\pm$  sample standard deviation across the three splits.

**Supplementary Table 4.** Split-wise test-set RMSEs at  $N_{\text{train}} = 170$  for the direct PBE+U fine-tuning experiments in Figure 2.

| Model            | Data split    | Energy RMSE (meV/atom) | Force RMSE (meV/Å) | Virial RMSE (meV/atom) |
|------------------|---------------|------------------------|--------------------|------------------------|
| DPA3-MP          | Split 0       | 1.827                  | 42.57              | 9.13                   |
| DPA3-MP          | Split 1       | 3.644                  | 51.21              | 10.06                  |
| DPA3-MP          | Split 2       | 2.068                  | 41.74              | 9.00                   |
| DPA3-MP          | Mean $\pm$ SD | $2.513 \pm 0.987$      | $45.17 \pm 5.24$   | $9.39 \pm 0.58$        |
| DPA3-OMat        | Split 0       | 2.101                  | 40.36              | 8.35                   |
| DPA3-OMat        | Split 1       | 2.239                  | 49.13              | 8.92                   |
| DPA3-OMat        | Split 2       | 0.887                  | 39.22              | 7.84                   |
| DPA3-OMat        | Mean $\pm$ SD | $1.742 \pm 0.744$      | $42.90 \pm 5.42$   | $8.37 \pm 0.54$        |
| DPA4-Air-MP      | Split 0       | 0.449                  | 30.41              | 5.43                   |
| DPA4-Air-MP      | Split 1       | 0.400                  | 40.04              | 5.08                   |
| DPA4-Air-MP      | Split 2       | 0.432                  | 28.01              | 4.76                   |
| DPA4-Air-MP      | Mean $\pm$ SD | $0.427 \pm 0.025$      | $32.82 \pm 6.37$   | $5.09 \pm 0.34$        |
| DPA4-Neo-OMat    | Split 0       | 0.467                  | 29.54              | 5.90                   |
| DPA4-Neo-OMat    | Split 1       | 0.423                  | 39.05              | 6.25                   |
| DPA4-Neo-OMat    | Split 2       | 0.456                  | 28.66              | 5.48                   |
| DPA4-Neo-OMat    | Mean $\pm$ SD | $0.448 \pm 0.023$      | $32.42 \pm 5.76$   | $5.88 \pm 0.39$        |
| DPA4-Neo scratch | Split 0       | 0.655                  | 41.97              | 11.31                  |

| Model            | Data split    | Energy RMSE (meV/atom) | Force RMSE (meV/Å) | Virial RMSE (meV/atom) |
|------------------|---------------|------------------------|--------------------|------------------------|
| DPA4-Neo scratch | Split 1       | 0.647                  | 48.82              | 10.88                  |
| DPA4-Neo scratch | Split 2       | 0.696                  | 38.16              | 10.82                  |
| DPA4-Neo scratch | Mean $\pm$ SD | $0.666 \pm 0.027$      | $42.98 \pm 5.40$   | $11.00 \pm 0.27$       |

**Supplementary Table 5.** Split-wise test-set RMSEs at  $N_{\text{train}} = 170$  for the sequential-adaptation experiments in Figure 3.

| Model         | Data split    | Energy RMSE (meV/atom) | Force RMSE (meV/Å) | Virial RMSE (meV/atom) |
|---------------|---------------|------------------------|--------------------|------------------------|
| DPA3-OMat     | Split 0       | 1.642                  | 30.54              | 9.27                   |
| DPA3-OMat     | Split 1       | 1.138                  | 47.20              | 9.44                   |
| DPA3-OMat     | Split 2       | 0.999                  | 37.01              | 9.54                   |
| DPA3-OMat     | Mean $\pm$ SD | $1.259 \pm 0.338$      | $38.25 \pm 8.40$   | $9.42 \pm 0.14$        |
| DPA4-Air-MP   | Split 0       | 0.461                  | 31.81              | 6.19                   |
| DPA4-Air-MP   | Split 1       | 0.369                  | 41.02              | 6.29                   |
| DPA4-Air-MP   | Split 2       | 0.563                  | 30.44              | 5.39                   |
| DPA4-Air-MP   | Mean $\pm$ SD | $0.464 \pm 0.097$      | $34.42 \pm 5.76$   | $5.95 \pm 0.49$        |
| DPA4-Neo-OMat | Split 0       | 0.585                  | 31.96              | 6.13                   |
| DPA4-Neo-OMat | Split 1       | 0.432                  | 40.99              | 5.93                   |
| DPA4-Neo-OMat | Split 2       | 0.576                  | 28.92              | 4.61                   |
| DPA4-Neo-OMat | Mean $\pm$ SD | $0.531 \pm 0.086$      | $33.96 \pm 6.28$   | $5.56 \pm 0.83$        |

## 5.2 Phase composition and phase/temperature-resolved test-set RMSEs at $N_{\text{train}} = 170$

To make the phase dependence of the prediction errors explicit, we report the phase composition of the  $N_{\text{train}} = 170$  training subsets and the corresponding held-out test sets in Supplementary Table 6. The selected training subsets are *Sqr*-rich, with approximately 65% *Sqr* configurations in all three splits. Therefore, the larger *Sqr*

errors do not arise from a smaller representation of *Sqr* configurations in the training subsets. Supplementary Table 7 further decomposes the sequentially adapted DPA4-Neo-OMat test errors by phase and sampling temperature. *Sqr* exhibits consistently larger energy, force, and virial RMSEs than *Oct* across 300, 600, and 900 K, demonstrating that the aggregate RMSE obscures a systematic phase dependence.

**Supplementary Table 6.** Phase composition of the  $N_{\text{train}} = 170$  training subsets and the corresponding held-out test sets.

| Data split | Training <i>Oct</i> | Training <i>Sqr</i> | Training <i>Sqr</i> fraction | Test <i>Oct</i> | Test <i>Sqr</i> |
|------------|---------------------|---------------------|------------------------------|-----------------|-----------------|
| Split 0    | 60                  | 110                 | 64.7%                        | 77              | 23              |
| Split 1    | 58                  | 112                 | 65.9%                        | 76              | 24              |
| Split 2    | 59                  | 111                 | 65.3%                        | 79              | 21              |

**Supplementary Table 7.** Phase- and temperature-resolved test RMSEs of sequentially adapted DPA4-Neo-OMat at  $N_{\text{train}} = 170$ . Values are mean  $\pm$  sample standard deviation across the three predefined data splits.

| Phase      | Temperature | Energy RMSE (meV/atom) | Force RMSE (meV/Å) | Virial RMSE (meV/atom) |
|------------|-------------|------------------------|--------------------|------------------------|
| <i>Oct</i> | 300 K       | 0.205 $\pm$ 0.062      | 3.51 $\pm$ 0.21    | 2.12 $\pm$ 0.64        |
| <i>Oct</i> | 600 K       | 0.164 $\pm$ 0.056      | 6.29 $\pm$ 0.35    | 3.68 $\pm$ 0.82        |
| <i>Oct</i> | 900 K       | 0.125 $\pm$ 0.013      | 6.93 $\pm$ 0.67    | 2.85 $\pm$ 0.13        |
| <i>Sqr</i> | 300 K       | 1.449 $\pm$ 0.173      | 96.78 $\pm$ 33.97  | 13.72 $\pm$ 3.90       |
| <i>Sqr</i> | 600 K       | 0.906 $\pm$ 0.241      | 43.30 $\pm$ 17.70  | 8.72 $\pm$ 1.92        |
| <i>Sqr</i> | 900 K       | 0.605 $\pm$ 0.129      | 47.06 $\pm$ 8.65   | 6.23 $\pm$ 0.16        |

### 5.3 Separate a possible temperature-distribution effect from the *Sqr*-specific increase in error at $N_{\text{train}} = 170$

To separate a possible temperature-distribution effect from the *Sqr*-specific increase in error, we analyzed the original  $N_{\text{train}} = 170$  splits by sampling temperature. All three selected training subsets contain 300, 600, and 900 K configurations (Supplementary Table 8). Moreover, the original held-out 600 K errors are not systematically larger than those at 300 or 900 K (Supplementary Table 9). A direct phase-resolved

comparison at 600 K shows that Oct errors remain nearly unchanged on the independently initialized trajectory, whereas the Sqr errors increase substantially (Supplementary Table 10). Thus, the independent-test degradation is not primarily a consequence of the nominal 600 K temperature. It is instead associated with the *Sqr* configurations sampled by the independent trajectory, consistent with the electronic/magnetic branch sensitivity discussed in Section 1.3 and with an accompanying configurational-distribution shift within the *Sqr* region.

**Supplementary Table 8.** Temperature distribution of the selected PBE+*U* training configurations at  $N_{\text{train}} = 170$ .

| Data split | 300 K     | 600 K      | 900 K       | Total |
|------------|-----------|------------|-------------|-------|
| Split 0    | 10 (5.9%) | 31 (18.2%) | 129 (75.9%) | 170   |
| Split 1    | 10 (5.9%) | 36 (21.2%) | 124 (72.9%) | 170   |
| Split 2    | 13 (7.6%) | 34 (20.0%) | 123 (72.4%) | 170   |

**Supplementary Table 9.** Temperature-resolved test-set RMSEs for DPA4-Neo-OMat at  $N_{\text{train}} = 170$  on the original held-out test sets. Values are mean  $\pm$  sample standard deviation across the three predefined data splits.

| Adaptation route      | Temperature | Energy RMSE (meV/atom) | Force RMSE (meV/Å) | Virial RMSE (meV/atom) |
|-----------------------|-------------|------------------------|--------------------|------------------------|
| Direct fine-tuning    | 300K        | $0.548 \pm 0.064$      | $43.94 \pm 12.40$  | $7.37 \pm 1.10$        |
| Direct fine-tuning    | 600K        | $0.361 \pm 0.044$      | $18.46 \pm 7.83$   | $5.11 \pm 0.75$        |
| Direct fine-tuning    | 900K        | $0.389 \pm 0.066$      | $24.24 \pm 4.69$   | $4.32 \pm 0.63$        |
| Sequential adaptation | 300K        | $0.739 \pm 0.159$      | $46.77 \pm 12.60$  | $6.95 \pm 1.94$        |
| Sequential adaptation | 600K        | $0.406 \pm 0.073$      | $19.36 \pm 6.84$   | $4.99 \pm 0.07$        |
| Sequential adaptation | 900K        | $0.323 \pm 0.049$      | $24.40 \pm 2.25$   | $4.01 \pm 0.30$        |

**Supplementary Table 10.** Phase-resolved comparison between the original held-out 600 K configurations and the independently initialized 600 K trajectories for DPA4-Neo-OMat. Original-test values are mean  $\pm$  sample standard deviation across the three predefined splits; independent-trajectory values correspond to the single independent validation trajectory for each phase.

| <b>Adaptation route</b> | <b>600 K dataset</b>    | <b>Phase</b> | <b>Energy RMSE (meV/atom)</b> | <b>Force RMSE (meV/Å)</b> | <b>Virial RMSE (meV/atom)</b> |
|-------------------------|-------------------------|--------------|-------------------------------|---------------------------|-------------------------------|
| Direct fine-tuning      | Original held-out 600 K | <i>Oct</i>   | 0.108 $\pm$ 0.010             | 4.94 $\pm$ 0.32           | 4.06 $\pm$ 0.20               |
| Direct fine-tuning      | Independent 600 K       | <i>Oct</i>   | 0.12                          | 4.64                      | 4.05                          |
| Direct fine-tuning      | Original held-out 600 K | <i>Sqr</i>   | 0.825 $\pm$ 0.112             | 41.98 $\pm$ 20.05         | 8.30 $\pm$ 2.25               |
| Direct fine-tuning      | Independent 600 K       | <i>Sqr</i>   | 5.99                          | 117.16                    | 20.22                         |
| Sequential adaptation   | Original held-out 600 K | <i>Oct</i>   | 0.164 $\pm$ 0.056             | 6.29 $\pm$ 0.35           | 3.68 $\pm$ 0.82               |
| Sequential adaptation   | Independent 600 K       | <i>Oct</i>   | 0.19                          | 6.08                      | 2.99                          |
| Sequential adaptation   | Original held-out 600 K | <i>Sqr</i>   | 0.906 $\pm$ 0.241             | 43.30 $\pm$ 17.70         | 8.72 $\pm$ 1.92               |
| Sequential adaptation   | Independent 600 K       | <i>Sqr</i>   | 7.10                          | 123.62                    | 20.43                         |

## REFERENCES

- (1) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Computational Materials Science* 2018, 152, 60–69.  
<https://doi.org/10.1016/j.commatsci.2018.05.018>