



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

Fengyu Xie¹ , Peiheng Jiang^{2,*} , Zhicheng Zhong^{1,3}

Keywords:

Machine-learned force field, transfer learning, fine-tuning, foundation model, electronic-structure adaptation

Citation: Xie, F.; Jiang, P.; Zhong, Z. Data-efficient adaptation of DPA-4 force fields to DFT+*U* energetics: a case study in NiO. *J. Mater. Inf.* 2026, 6, 49. <https://dx.doi.org/10.20517/jmi.2026.71>

Received: 12 Aug 2026

First Decision: 8 Sep 2026

Revised: 3 Oct 2026

Accepted: 13 Oct 2026

Published: 26 Oct 2026

Academic Editor:

Sergei Manzhos

Copy Editor:

Pei-Yun Wang

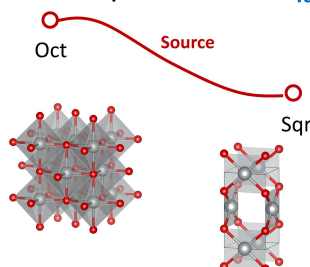
Production Editor:

Pei-Yun Wang

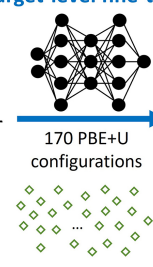
Fine-tuning transfers MLFFs across electronic-method PESs

Source MLFF: PBE without *U*

Incorrect source ordering
Sqr below *Oct*

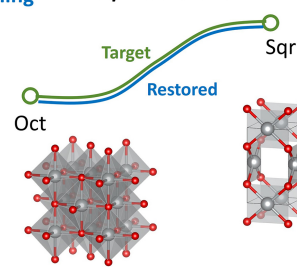


Target-level fine-tuning



Recovered MLFF: PBE + *U*

Target ordering restored
Sqr above *Oct*



Broad and consistent pretraining → Efficient fine-tuning → Correct target energetics

Abstract

Foundation machine-learned force fields (MLFFs) are often pretrained on broad materials datasets whose electronic-structure conventions may not reproduce the phase energetics required for a specific correlated material. Using NiO as a case study, we examine whether distinct source-protocol energetics can be adapted efficiently through target-protocol fine-tuning. Along a common fixed structural interpolation, non-spin-polarized Perdew–Burke–Ernzerhof (PBE) and ferromagnetically initialized PBE+*U* predict opposite energetic orderings of the octahedral *Oct* and square-planar *Sqr* phases. When fine-tuned directly, the pretrained DPA-4 models adapt rapidly to the NiO PBE+*U* surface, reaching energy, force, and virial root-mean-square errors (RMSEs) of approximately 0.4–0.5 meV/atom, 30–35 meV/Å, and 5–6 meV/atom, respectively, with 170 PBE+*U* training configurations. Crucially, models previously fine-tuned to the opposing no-*U* surface recover the qualitative target-protocol energetic ordering along the fixed structural path with nearly the same target-label efficiency as models fine-tuned directly from their respective pretrained initializations. Our results show that distinct source-protocol energetics can be reversed through target-protocol fine-tuning, and suggest a practical



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

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

³Suzhou 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

multi-fidelity strategy in which pretraining prioritizes broad, consistent, and affordable data, while compact target-protocol datasets impose energetics through application-specific fine-tuning.

INTRODUCTION

In transition-metal compounds, conventional semilocal density-functional theory (DFT) may over-delocalize the *d* electrons and inadequately describe the associated magnetic moments. Corrective approaches such as DFT+*U* are therefore frequently employed to improve the description of localization, magnetism, and relative phase energetics^[1–3]. In some cases, these corrections are not merely quantitative. The relative stability of competing crystal structures can be sensitive to the treatment of localized electronic states, and the inclusion or choice of an on-site interaction can even reverse their energetic ordering^[4–6]. A machine-learned force field (MLFF) trained on uncorrected DFT calculations will generally inherit the corresponding reference-level errors and may consequently predict incorrect phase stability.

Nevertheless, using DFT+*U* for large-scale MLFF pretraining may introduce a separate difficulty. Commonly used materials databases, such as Materials Project (MP)^[7,8], Alexandria^[9,10] and OMat24^[11] may combine calculations performed under composition-dependent DFT and DFT+*U* conventions^[3,12]. Although these conventions were often designed to improve agreement between calculated and experimental thermodynamics for transition-metal compounds, they effectively mix labels generated by different Hamiltonians into a single training set. Consequently, an MLFF must reconcile potentially incompatible energetic descriptions across chemical environments, which may lead to nonphysical interpolation. Recent work^[13] has investigated the effect of such dataset inconsistency in depth and argued that consistently generated data without *U* could be more suitable for pretraining foundation MLFFs than datasets containing heterogeneous *U* conventions.

This proposal naturally suggests a separation of roles between pretraining and system-specific fine-tuning. During pretraining, an MLFF learns broadly transferable chemical interactions from an affordable and consistently parameterized reference DFT dataset. Quantitative energetics for a particular correlated material can subsequently be imposed through fine-tuning with a compact dataset at a more appropriate theory level, such as DFT+*U* or hybrid-functional DFT. For this strategy to be practically useful, an MLFF pretrained without *U* must remain adaptable even when its pretraining encodes qualitatively incorrect phase energetics. Our previous work demonstrated that foundation MLFFs can be fine-tuned for specific materials with remarkable data efficiency^[14]. However, it remains unclear whether prior fine-tuning to a distinct potential-energy surface impedes subsequent adaptation, and whether a small target dataset can efficiently reverse a strongly learned energetic preference.

Here, we investigate these questions using NiO, an archetypal strongly correlated transition-metal oxide, as a case study. We consider a rocksalt-like structure with octahedrally coordinated Ni, denoted *Oct*, and a PdO-like tetragonal structure with square-planar Ni coordination, denoted *Sqr* [Figure 1]. Along a common structural interpolation between these phases, non-spin-polarized Perdew–Burke–Ernzerhof (PBE) and ferromagnetically initialized (FM-initialized) PBE+*U* predict opposite relative energetics. This pronounced reversal provides a stringent test of whether an MLFF carrying a distinct source-protocol energetic preference can be efficiently adapted to a target-protocol surface.

We fine-tune several pretrained DPA models^[15,16] using a NiO dataset generated with FM-initialized PBE+*U* under consistent computational settings. With 170 training configurations, the fine-tuned models achieve low test errors for PBE+*U* energies, forces, and virials, while recovering the qualitative target-protocol energetic ordering along the fixed *Oct*–*Sqr* path. For the DPA-4 models, fine-tuning with 170 PBE+*U*

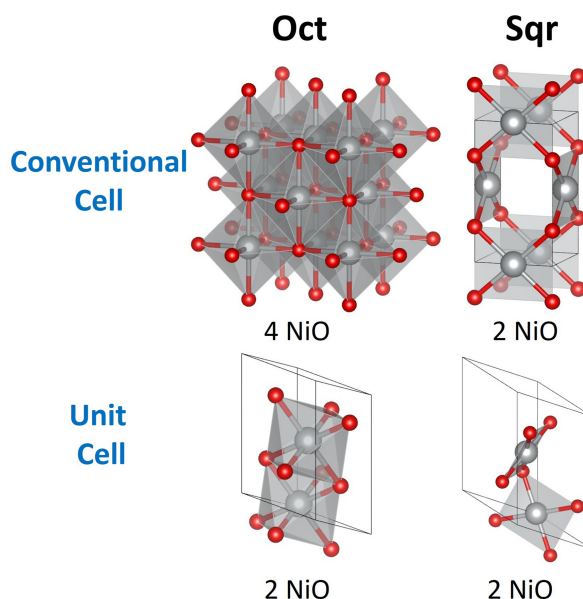


Figure 1. Crystal structures of NiO with octahedral (*Oct*) and square-planar (*Sqr*) Ni coordination. The upper row shows the conventional rocksalt-like cubic cell of *Oct*-NiO containing four formula units (left) and the conventional PdO-like tetragonal cell of *Sqr*-NiO containing two formula units (right). The lower row shows the corresponding triclinic cells of *Oct*-NiO (left) and *Sqr*-NiO (right), each containing two formula units. The two-formula-unit triclinic cells were used for the structural interpolation and as the parent cells for constructing the $2 \times 2 \times 2$ AIMD supercells. AIMD: *Ab initio* molecular dynamics.

training configurations reduces the energy root-mean-square error (RMSE) from approximately 15–20 meV/atom to below 0.4–0.5 meV/atom and the force RMSE from approximately 125–175 meV/Å to below 30–35 meV/Å across three data splits. The corresponding virial RMSEs decrease from approximately 50–80 meV/atom to around 5–6 meV/atom. Most importantly, for representative split 0 in the separate two-stage experiment, DPA4-Neo-OMat reaches energy, force, and virial RMSEs of 0.59 meV/atom, 31.96 meV/Å, and 6.13 meV/atom, respectively, with the same 170 target configurations. DPA-4 models previously fine-tuned on non-spin-polarized PBE data, and therefore encoded the opposite *Oct*–*Sqr* energetic preference, can be transferred to the FM-initialized PBE+*U* surface with nearly the same data efficiency as models fine-tuned directly from the corresponding pretrained initialization. Supporting analyses of the near-*Sqr* discrepancy, independent-trajectory validation, electronic-structure controls, and detailed Vienna *ab initio* Simulation Package (VASP) settings are provided in [Supplementary Sections 1–4](#), respectively.

Our results demonstrate that distinct source-protocol phase energetics do not necessarily make a pretrained MLFF unsuitable for subsequent target-protocol adaptation. More broadly, they suggest a practical multi-fidelity strategy for developing and using foundation MLFFs. Pretraining can prioritize broad configurational coverage, consistency of the DFT labels, and computational affordability, whereas compact target-protocol datasets can provide the electronic and energetic fidelity required for a specific application. This suggestion does not imply that arbitrary source-protocol errors are harmless. Rather, the present NiO example shows that target-protocol energetic ordering need not be reproduced during pretraining for the resulting MLFF representation to remain useful. Foundation MLFFs should therefore be regarded as transferable priors that require target-specific validation and, where necessary, appropriate target-protocol fine-tuning.

MATERIALS AND METHODS

DFT calculations

All DFT calculations, including structural relaxations and single-point calculations, were carried out using VASP with the projector-augmented-wave (PAW) method^[17–20]. A plane-wave cutoff of 520 eV was used. Input files were generated using the PBE pseudopotentials through pymatgen^[21] MPRelaxSet

(POT_GGA_PAW_PBE, corresponding to the 06-05-2010 PBE PAW dataset in VASP documentation). Specifically, the Ni_pv potential (PAW_PBE Ni_pv 06Sep2000) was used with $3p^6 3d^8 4s^2$ electrons treated as valence, together with the standard O potential (PAW_PBE O 08Apr2002) with $2s^2 2p^4$ electrons treated as valence. For the interpolated unit-cell configurations, self-consistent (SCF) static calculations were first carried out on a Γ -centered $8 \times 8 \times 5$ k-point grid to obtain the energies, magnetic moments, and charge densities. Subsequent non-self-consistent (NSCF) calculations reused the converged charge densities and employed a denser Γ -centered $16 \times 16 \times 10$ k-point grid to obtain the projected density of states (PDOS) in [Supplementary Figure 6](#). For all *ab initio* molecular dynamics (AIMD) calculations, Γ -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. Gaussian smearing with ISMEAR = 0 and SIGMA = 0.05 eV was employed. Electronic convergence (EDIFF) was set to 1×10^{-6} eV, and structural relaxations were converged to a force threshold of 0.02 eV/Å. For all DFT+*U* calculations, we consistently employed the rotationally invariant form introduced by Dudarev *et al.*^[22] with $U_{\text{eff}} = U - J = 6.2$ eV applied to the Ni 3d orbitals, following the convention implemented in pymatgen. For all spin-polarized calculations, including the interpolated static calculations and AIMD simulations, the initial magnetic moment of O was set to $+0.01 \mu_B$, while the magnitude of the initial Ni magnetic moment was set to $3.0 \mu_B$. The signs of the Ni moments were assigned according to the FM or antiferromagnetic (AFM) initialization protocol. Throughout this work, we use “FM-initialized” and “AFM-initialized” to refer to the imposed electronic/magnetic initialization rather than to guarantee uniform convergence to ferromagnetic or antiferromagnetic states, as the SCF magnetic solutions can be more complex, particularly in the *Sqr* region.

In order to compare *Oct*–*Sqr* energetics under different correlation treatments, we employed PBE^[23,24], PBE+*U*, r2SCAN^[25], and HSE06^[26]. No additional Hubbard correction was applied to r2SCAN and HSE06. In these computations, we used triclinic cells containing two NiO formula units [[Figure 1](#), lower panels]. The *Oct* and *Sqr* endpoint structures were first relaxed using non-spin-polarized PBE without *U*. Intermediate configurations were then generated by linearly interpolating both the atomic coordinates and lattice matrices. The resulting fixed interpolation pathway was evaluated using single-point calculations at each electronic-structure level.

To obtain PBE and FM-initialized PBE+*U* datasets for fine-tuning MLFFs, we performed AIMD in the isothermal–isobaric (NPT) ensemble at $T = 300, 600, 900$ K and $P = 1$ bar. Both the cell volume and the cell shape were allowed to vary. The AIMD trajectories were propagated using the Langevin thermostat and barostat, with lattice friction coefficients of 10 ps^{-1} and a fictitious lattice mass of 1,000 atomic mass units. At each temperature, we performed two simulations, initialized from the *Oct* and the *Sqr* structures, respectively. Each simulation lasted 250 steps with a time step of 4.0 fs. From the resulting PBE trajectories of the *Oct* and the *Sqr* phases, we extracted 750 electronically converged frames for each phase. From the FM-initialized PBE+*U* trajectories, we extracted 750 and 250 electronically converged frames for the *Oct* and the *Sqr* phases, respectively. For *Sqr*-NiO, only 250 frames in total across the three PBE+*U* trajectories reached electronic self-consistency; the remaining unconverged frames were excluded, as detailed in [Supplementary Section 4.4](#). These extracted frames subsequently formed the datasets for fine-tuning and validating MLFFs. For AIMD, the two-formula-unit cells were replicated $2 \times 2 \times 2$, yielding supercells containing 16 NiO formula units to sample a broader range of local configurations.

Representative INCAR files used for the FM-initialized PBE+*U* fixed-path static calculations and 300 K AIMD simulations, along with detailed representative VASP settings, are provided in [Supplementary Section 4](#).

Dataset preparation

For the datasets generated from AIMD calculations, we first held out 10% of all configurations at random for testing. Training configurations were then selected from the remaining configurations by maximizing the information entropy implemented in QUESTS^[27]. The information entropy of a set of n structures (H) is defined as^[28]

$$H(\{\mathbf{X}\}) = -\frac{1}{n} \sum_{i=1}^n \log \left[\frac{1}{n} \sum_{j=1}^n K_h(\mathbf{X}_i, \mathbf{X}_j) \right], \quad (1)$$

where $\mathbf{X}_i$ denotes the vectorized local atomic features of structure i , K_h is a Gaussian kernel, and h is the kernel bandwidth, set to $h = 0.015$ in this work. The local atomic features were constructed using 32 channels and a radial cutoff of 5.0 Å. The differential entropy gain upon adding a candidate structure $\mathbf{Y}$ into the existing dataset $\{\mathbf{X}\}$ is given by

$$\delta H(\mathbf{Y}|\{\mathbf{X}\}) = -\log \left[\sum_{i=1}^n K_h(\mathbf{Y}, \mathbf{X}_i) \right]. \quad (2)$$

All DFT-labeled structures were ranked by δH , and the top N_{train} configurations were selected to maximize the information diversity of the training set for a given training-set size.

Using this approach, we selected $N_{\text{train}} = 660$ configurations from the PBE dataset to fit the pretrained MLFFs accurately to the non-spin-polarized PBE surface. For the PBE+ U dataset, we varied N_{train} from 50 to 290 in increments of 40 to investigate how rapidly test errors decrease with increasing training-set size. Three independent train/validation-test splits were first generated for the PBE+ U dataset. Within each split, the test set was held fixed across all training-set sizes, while progressively larger training subsets were selected from the corresponding train/validation pool by information-entropy maximization. The configurations not selected for training within the train/validation pool served as the validation set and were used only for on-the-fly monitoring of the validation loss during training with deepmd-kit. They did not contribute to gradient updates and were not used for final model evaluation, checkpoint selection, or hyperparameter optimization; all reported performance metrics were calculated exclusively on the held-out test set. The same matched data partitions were used for direct fine-tuning, sequential adaptation, and training-from-scratch comparisons.

Here, N_{train} denotes the number of configurations used for adaptation and therefore measures target-label efficiency within the fixed candidate pool generated in this study. Thus, N_{train} measures training-sample efficiency within the fixed 1,000-configuration PBE+ U candidate dataset and should not be interpreted as either the total number of DFT+ U calculations or the computational cost required to generate this candidate pool. In practical workflows, the candidate configurations need not be generated by full DFT-AIMD trajectories and may instead be obtained using lower-cost sampling such as pretrained MLFF MD followed by selective target-protocol labeling.

Foundation MLFFs and fine-tuning

We compared four foundation MLFFs: the *MPtraj* branch and the *OMat24* branch of DPA-3.1-3M^[15,29], denoted as DPA3-MP and DPA3-OMat, respectively, and two DPA4^[16] models, denoted as DPA4-Air-MP and DPA4-Neo-OMat. The neighbor cutoff radii were set to 6 Å for all these models. Models were fine-tuned and evaluated using deepmd-kit^[30]. For the training-from-scratch baseline, we used the DPA4-Neo architecture as a representative controlled baseline because it is the principal DPA-4 architecture examined in this work. The model was randomly initialized and trained using the same training-set sizes, matched data splits, learning-rate schedule, batch settings, loss definition, and other training settings used for DPA4-Neo fine-tuning. For direct fine-tuning, second-stage adaptation, and training from scratch, one training run with

a fixed random seed was performed for each combination of data split and N_{train} ; no additional repetitions with different optimization seeds were performed.

Models were trained for approximately 120 epochs, corresponding to $\sim 120 \times N_{\text{train}}/N_{\text{batch}}$ steps. Automatic batch sizing was used with a target of 128 atoms per batch. The gradient norm was clipped to a maximum of 1.0. For DPA-3, the learning rate decayed exponentially from 0.001 to 3×10^{-5} over the full training schedule at an interval of N_{train} steps. The DPA-3 loss function combined energy, force, and virial mean squared error (MSE) terms. Their weights were varied from 0.2:100:0.02 at the beginning of training to 20:60:1 at the end. For DPA-4, the learning rate followed a warmup-stable-decay (WSD) schedule, starting at 1.4×10^{-4} , warming up to 7.0×10^{-4} in N_{train} steps, remaining stable until 65% of the total training steps had been completed, after which it decayed sinusoidally to 1×10^{-6} . The DPA-4 loss combined energy, force, and virial mean absolute error (MAE) terms with fixed weights of 1:1:0.25, respectively.

For the test-set energy-offset-aligned zero-shot RMSE, a uniform per-atom energy shift was applied to each pretrained model before calculating the RMSE. The shift was defined as the difference between the mean DFT energy per atom and the mean MLFF-predicted energy per atom over the corresponding test set, and the same constant was then added to all MLFF energy predictions in that test set. This removes arbitrary differences in the absolute energy references of pretrained MLFFs while preserving all relative energy differences among configurations. This aligned quantity is used only as a descriptive zero-shot baseline and is not used for model selection, hyperparameter optimization or quantitative claims regarding the recovery of relative energetics.

RESULTS AND DISCUSSION

Introduction to the MLFF models

The MLFFs considered in this work are local, energy-conserving neural-network potentials that approximate the potential energy as a function of atomic species and geometry. For a configuration with atomic species Z_i and coordinates $\mathbf{r}_i$, the total energy is expressed as a sum of atomwise contributions determined by the local environment N_i within a finite cutoff radius,

$$E(\{Z_i, \mathbf{r}_i\}) = \sum_i \varepsilon_i(N_i). \quad (3)$$

Forces and virials are subsequently obtained from derivatives of the predicted total energy with respect to atomic coordinates and cell degrees of freedom. Consequently, these models represent a single-valued mapping from atomic geometry to energy and do not contain explicit electronic or magnetic state variables. This feature is particularly relevant to the present study, where different electronic-structure protocols may assign distinct energetic landscapes to closely related atomic configurations.

Here, we compare two generations of the Deep Potential architecture. DPA-3 is a multilayer graph neural network (GNN) constructed on a line graph series (LiGS), in which atoms, bonds, and angular relationships are represented through coupled graph levels and updated by message passing^[15]. The DPA-3.1-3M architecture used here employs an order-2 LiGS with 16 interaction layers and contains approximately 3.27 million trainable parameters. We evaluate two released DPA-3.1-3M checkpoints, denoted DPA3-MP and DPA3-OMat, corresponding to models pretrained on MP and OMat24 data, respectively.

DPA-4 is a later SE(3)-equivariant architecture based on edge-conditioned multi-focus attention (EMFA) SO(2)-equivariant convolution^[16]. It propagates scalar and higher-order angular features through equivariant message passing and uses the resulting scalar representation for atomwise energy prediction. In this work, we use two released DPA-4 variants: DPA4-Neo-OMat and DPA4-Air-MP. DPA4-Neo contains approximately

1.60 million trainable parameters, whereas the DPA4-Air architecture is larger and contains approximately 2.76 million parameters. The corresponding checkpoints were pretrained on data derived from OMat24 and MP, respectively. These models differ not only in pretraining history but also in architecture and model capacity, providing a useful set of pretrained priors for comparing downstream target-protocol adaptability under the same NiO fine-tuning task.

Readers are referred to the original model publications for further architectural details^[15,16].

Correlation-sensitive energetics in NiO as a transfer-learning testbed

To inspect how different electronic-structure treatments alter the NiO energy landscape and whether non-spin-polarized PBE and PBE+*U* provide a sufficiently challenging testbed for MLFF adaptability, we first relaxed the *Oct* and *Sqr* endpoint structures using non-spin-polarized PBE and generated intermediate configurations by linearly interpolating both the atomic coordinates and lattice matrices. The resulting configurations define a fixed *Oct*–*Sqr* structural path on which the same set of geometries can be evaluated using different electronic-structure treatments. Because the path endpoints were uniformly relaxed using non-spin-polarized PBE rather than separately at each electronic-structure level, this construction should be interpreted as a demonstrative comparison of reference-level energetic differences, rather than as a minimum-energy transition path or a determination of equilibrium phase stability.

Detailed PBE, PBE+*U*, r2SCAN, and HSE06 results, together with magnetic-moment analysis, are provided in [Supplementary Section 3](#). Along this fixed path, non-spin-polarized PBE and FM-initialized PBE+*U* exhibit strongly opposing *Oct*–*Sqr* energetic preferences, providing a stringent testbed for evaluating whether an MLFF previously adapted to one reference surface can subsequently be redirected to the other. Additional spin-polarized PBE control calculations show that spin polarization alone does not reproduce this energetic reversal; the PBE *Sqr* solution is metallic and essentially nonmagnetic, whereas PBE+*U* supports distinct metallic HS-like and semiconducting LS-like solutions, as detailed in [Supplementary Sections 3.2 and 3.3](#). These comparisons also suggest that, although r2SCAN-based pretraining has been recommended in several recent studies^[31–33] as a route toward improved thermodynamic accuracy and zero-shot phase predictions, using a higher-level semilocal reference such as r2SCAN may not by itself eliminate the need for system-specific validation and, where necessary, targeted fine-tuning.

Data-efficient fine-tuning of pretrained MLFFs to PBE+*U*

We next examine whether pretrained MLFFs can be efficiently adapted to the FM-initialized PBE+*U* reference surface. [Figure 2](#) shows the energy, force, and virial RMSEs on the held-out PBE+*U* test set as functions of the number of target-protocol training configurations, N_{train} . The results at $N_{\text{train}} = 0$ correspond to zero-shot predictions. We compare four pretrained models spanning three architectures: DPA3, DPA4-Air, and DPA4-Neo. For DPA3, we consider two pretrained branches, DPA3-MP and DPA3-OMat, trained on the MP^[7] and OMat24^[11] datasets, respectively. For DPA4-Air and DPA4-Neo, we use the separately released DPA4-Air-MP and DPA4-Neo-OMat checkpoints, pretrained on MP and OMat24, respectively. For each $N_{\text{train}} > 0$, we train and evaluate each model on the three predefined data splits; within each split, the test set remains fixed across all training-set sizes. For the zero-shot case, we evaluate the pretrained models on the corresponding test sets.

All four models improve substantially upon fine-tuning with a small amount of PBE+*U* data. In particular, the test-set energy-offset-aligned zero-shot RMSEs of the DPA-4 models are approximately 15–20 meV/atom and decrease to approximately 1–2 meV/atom with only 50 training configurations and approach 0.5 meV/atom after fine-tuning on 170 configurations. Their force RMSEs similarly decrease from approximately 125–175 meV/Å for zero-shot predictions to approximately 30 meV/Å with 170 training

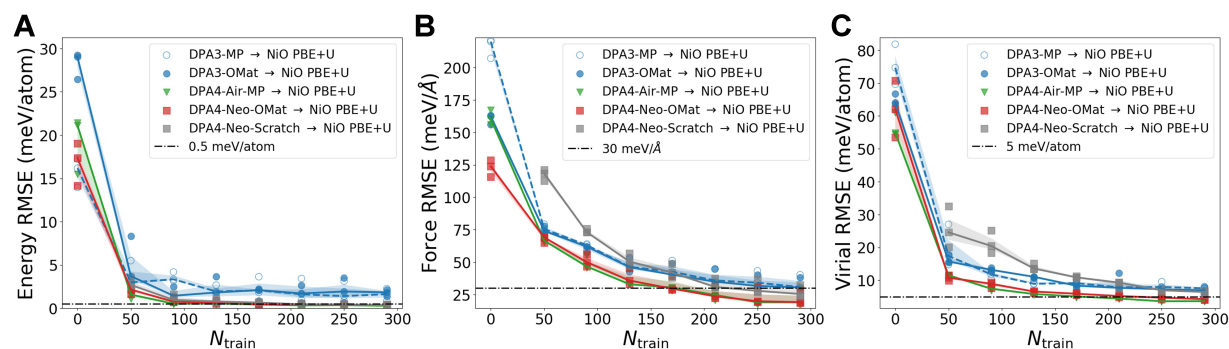


Figure 2. Data-efficient adaptation of pretrained MLFFs to the NiO PBE+*U* reference surface. Test-set RMSEs for (A) energies, (B) forces, and (C) virials are shown as functions of the number of PBE+*U* training configurations, N_{train} . Results at $N_{\text{train}} = 0$ correspond to zero-shot predictions of the pretrained models. The zero-shot energy errors are reported as test-set energy-offset-aligned zero-shot RMSEs, using a uniform per-atom energy shift determined from the corresponding test set as described in the Methods section. Symbols represent individual train–test splits. Open blue circles connected by dashed lines denote DPA3-MP, while the filled blue circles, green triangles, and red squares along with solid lines denote DPA3-OMat, DPA4-Air-MP, and DPA4-Neo-OMat, respectively. Filled grey squares represent DPA4-Neo models trained from random initialization using the same training-set sizes and data splits. Shaded regions indicate the interquartile range across independent train–test splits at each N_{train} . The horizontal dash-dotted lines indicate RMSEs of 0.5 meV/atom in (A), 30 meV/Å in (B), and 5 meV/atom in (C). MLFFs: Machine-learned force fields; PBE: Perdew–Burke–Ernzerhof; RMSEs: root-mean-square errors.

configurations and improve further with additional training data. Their virial RMSEs decrease from approximately 50–80 meV/atom to around 5 meV/atom. Exact split-wise RMSEs at $N_{\text{train}} = 170$, together with three-split summary statistics, are provided in [Supplementary Section 5.1](#). Compared with training DPA4-Neo from scratch, fine-tuning from the pretrained DPA4-Neo-OMat initialization consistently yields lower force and virial RMSEs, particularly in the low-data regime. These results show that foundation MLFFs pretrained on large materials datasets can be transferred efficiently to a consistently defined, system-specific PBE+*U* surface, despite differences between the electronic-structure conventions represented during pretraining and the consistently defined target level used for fine-tuning.

Under the present fine-tuning protocol (see Section “Foundation MLFFs and fine-tuning”), the tested DPA-4 models exhibit greater data efficiency than the tested DPA-3 models. Although the DPA-3 models also benefit from target-protocol fine-tuning, their energy RMSEs remain approximately 1–2 meV/atom over much of the investigated range, and their force RMSEs reach approximately 30 meV/Å only for the largest training subsets ($N_{\text{train}} = 290$), while their virial RMSEs also remain substantially higher. In comparison, both DPA-4 models reach lower energy, force, and virial RMSEs with fewer PBE+*U* labels.

Recovery after no-*U* fine-tuning

We further test a more demanding transfer scenario in which the pretrained models are first fine-tuned on non-spin-polarized PBE data without *U* and subsequently transferred to the FM-initialized PBE+*U* target surface. The intermediate fine-tuning stage deliberately fits the models to the potential-energy surface that favors *Sqr* over *Oct*, allowing us to test whether prior learning of a qualitatively distinct energetic preference impedes subsequent target-protocol adaptation.

To verify accurate fitting to the no-*U* reference surface, we examine DPA4-Neo-OMat, which was first fine-tuned using 660 non-spin-polarized PBE configurations. The corresponding DFT-label standard deviations are reported in the parity plots to provide the scale of variation in each target quantity. As shown in [Figure 3A–C](#), the resulting model reproduces the test-set no-*U* energies, forces, and virials with RMSEs of 0.12 meV/atom, 6.49 meV/Å and 4.88 meV/atom, respectively. The intermediate model therefore reproduces the no-*U* surface accurately, rather than merely acquiring only a weak source-protocol bias.

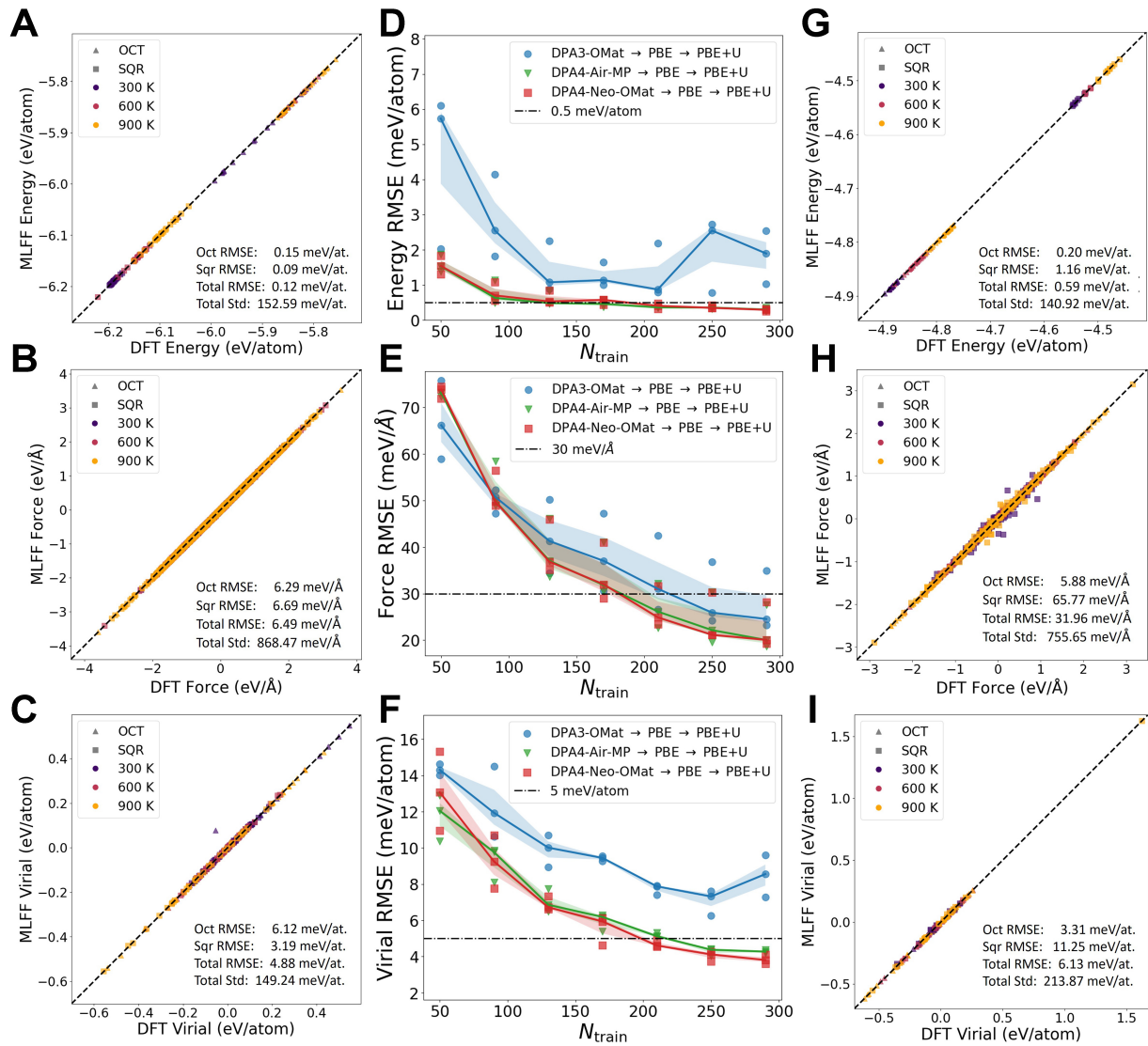


Figure 3. Second-stage adaptation from the non-spin-polarized PBE surface without U to the FM-initialized PBE+ U target surface. (A–C) Parity plots of MLFF predictions against the corresponding DFT labels for DPA4-Neo-OMat after the first-stage fine-tuning on 660 non-spin-polarized PBE configurations without U : (A) energy, (B) force, and (C) virial. The dashed diagonal lines indicate ideal agreement; (D–F) Test-set RMSEs for (D) energies, (E) forces, and (F) virials as functions of the number of FM-initialized PBE+ U training configurations, N_{train} , used in the second fine-tuning stage. Before this stage, each model was first fine-tuned on non-spin-polarized NiO PBE data without U . Symbols denote results from independent train–test splits. Blue circles, green triangles, and red squares denote DPA3-OMat, DPA4-Air-MP, and DPA4-Neo-OMat models, respectively, each first fine-tuned on 660 non-spin-polarized PBE configurations without U and subsequently fine-tuned on FM-initialized PBE+ U configurations; solid lines show the corresponding trends. Shaded regions indicate the interquartile range across independent train–test splits at each N_{train} . The horizontal dash-dotted lines mark RMSEs of 0.5 meV/atom in (D), 30 meV/Å in (E), and 5 meV/atom in (F); (G–I) Parity plots of MLFF predictions against the corresponding DFT labels for DPA4-Neo after the second-stage fine-tuning on 170 FM-initialized PBE+ U configurations for data split 0: (G) energy, (H) force, and (I) virial. The dashed diagonal lines indicate ideal agreement. The parity panels report phase-resolved and total RMSEs together with the standard deviation of the corresponding DFT reference quantity. Split 0 is shown as a representative split; statistics across all three splits are reported in [Supplementary Section 5.1](#). PBE: Perdew–Burke–Ernzerhof; FM: ferromagnetically; MLFF: machine-learned force field; DFT: density-functional theory; RMSEs: root-mean-square errors.

Figure 3D–F show the test errors obtained when the no- U -adapted models are further fine-tuned using FM-initialized PBE+ U data. Despite their prior alignment with the opposing energetic preference, both DPA-4 models adapt rapidly to the PBE+ U surface. Their RMSEs decrease with increasing training-set size at rates comparable to those obtained by direct fine-tuning from the corresponding foundation models in Figure 2. For the representative split 0 at $N_{\text{train}} = 170$, the sequentially adapted DPA4-Neo reaches energy, force, and

virial RMSEs of 0.59 meV/atom, 31.96 meV/Å, and 6.13 meV/atom, respectively, as further shown in [Figure 3G–I](#). Statistics across all three splits are provided in [Supplementary Section 5.1](#). The *Sqr* configurations exhibit consistently larger errors than the *Oct* configurations across 300, 600, and 900 K despite comprising approximately two-thirds of the $N_{\text{train}} = 170$ training subsets; detailed phase compositions and phase-resolved RMSEs are provided in [Supplementary Section 5.2](#). Therefore, prior alignment with the no-*U* surface does not substantially increase the number of target-protocol configurations required by DPA-4. We further evaluated the directly and sequentially adapted DPA4-Neo models on independently initialized 600 K PBE+*U* AIMD trajectories that were not used during fine-tuning [[Supplementary Figure 3](#)]. The resulting errors are higher than those obtained from the original trajectory-based test splits, with total energy, force, and virial RMSEs of 3.15 meV/atom, 61.62 meV/Å, and 11.16 meV/atom for direct fine-tuning, and 3.73 meV/atom, 65.09 meV/Å, and 11.02 meV/atom for sequential adaptation. Because 600 K is represented in the training data and does not show systematically larger errors in the original held-out test sets, the increase on the independent trajectories cannot be attributed primarily to a temperature-distribution shift. Instead, the degradation is localized mainly to *Sqr* configurations and is consistent with a configurational/electronic-state distribution shift in the multistable *Sqr* region [[Supplementary Sections 1.3 and 5.3](#)]. Nevertheless, both adapted models benefit substantially from fine-tuning compared with the zero-shot model, and the sequentially adapted model recovers nearly the same accuracy as direct fine-tuning, providing further evidence that transfer between the two electronic-structure levels remains effective beyond configurations directly correlated with the original AIMD dataset.

More importantly, second-stage fine-tuning corrects not only the aggregate test errors but also the target-protocol energetic trend. [Figure 4](#) compares the relative-energy profiles predicted by the no-*U*-adapted, directly PBE+*U*-fine-tuned, and sequentially adapted DPA4-Neo models along the common *Oct*–*Sqr* interpolation path. The no-*U*-adapted model places the *Sqr* endpoint 0.796 eV/f.u. below *Oct*. In contrast, after second-stage fine-tuning with PBE+*U*, the recovered model places *Sqr* 0.765 eV/f.u. above *Oct*, reproducing the sign and closely following the profile obtained from direct PBE+*U* fine-tuning. The inherited source-protocol energetic preference can therefore be reversed, and a compact target-protocol dataset is sufficient to redirect DPA-4 toward the opposing energetic preference. Although both the direct and sequential PBE+*U* fine-tuning routes yield a positive *Sqr*–*Oct* energy difference, both underestimate its magnitude, predicting 0.765 eV/f.u. compared with the PBE+*U* value of 0.936 eV/f.u. Our additional analyses indicate that this residual discrepancy is localized primarily near the *Sqr* end of the path and is most plausibly associated with the particularly complex electronic/magnetic landscape of *Sqr* NiO, where closely related or even identical atomic geometries can converge to distinct SCF states with substantially different energies. This multistability persists when the electronic SCF convergence threshold is tightened from 10^{-6} to 10^{-8} eV, indicating that it is not simply a consequence of insufficient SCF convergence [[Supplementary Section 1.3](#)]. Detailed analyses of the structural coverage, electronic/magnetic multistability, and its consequences for a coordinate-only MLFF are provided in [Supplementary Section 1](#).

DPA-3 also benefits from second-stage fine-tuning, but its energy, force, and virial RMSEs remain substantially higher than those of the DPA-4 models over the investigated training-set sizes, with greater variation across the predefined data splits. DPA-4 therefore exhibits greater adaptability and greater stability across the tested data splits. The contrast in recovery behavior of DPA-3 and DPA-4 further indicates that efficient cross-protocol adaptation is not an automatic consequence of fine-tuning, but rather a model-dependent capability of the underlying MLFF.

CONCLUSIONS

In this work, we investigated whether an MLFF carrying qualitatively distinct source-protocol energetics can be efficiently adapted to a target-protocol surface. Using NiO as a case study, we considered transfer from a

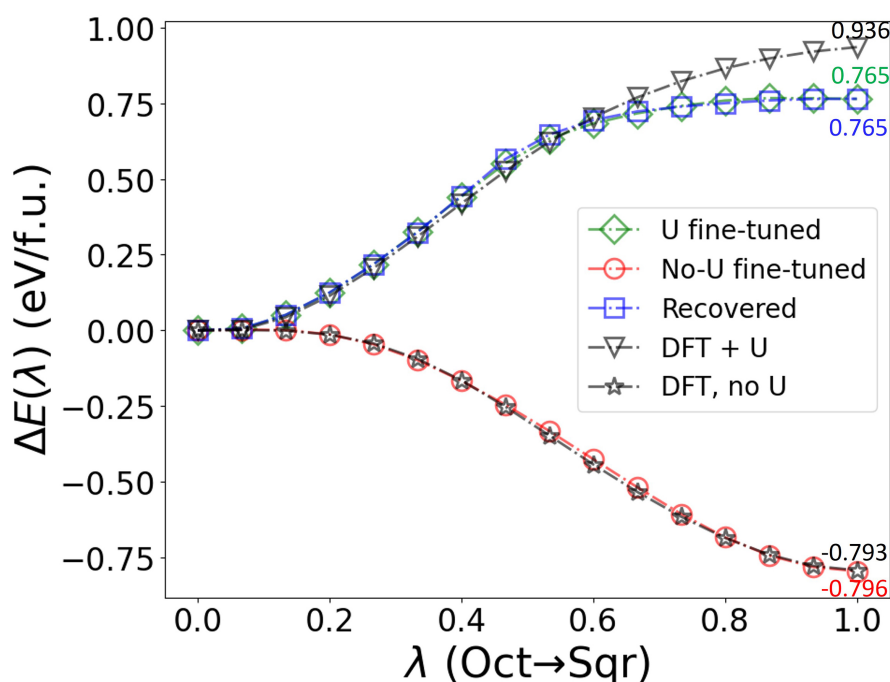


Figure 4. MLFF-predicted and DFT-computed relative energies, $\Delta E(\lambda)$, along the common Oct–Sqr structural interpolation path (λ), where $\lambda = 0$ and 1 correspond to Oct and Sqr, respectively. For each model, energies are referenced to its prediction for Oct at $\lambda = 0$, such that $\Delta E(\lambda) = E(\lambda) - E(0)$. Red circles denote predictions from the model fine-tuned on non-spin-polarized PBE data without U , while green diamonds and blue squares denote the model fine-tuned directly from its pretrained initialization to PBE+ U and the model first fine-tuned on PBE without U and subsequently fine-tuned on PBE+ U , respectively. All predictions are obtained with DPA4-Neo. Black triangles and stars represent static DFT-calculated energies from PBE+ U and non-spin-polarized PBE, respectively. MLFF: Machine-learned force field; DFT: density-functional theory; PBE: Perdew–Burke–Ernzerhof.

non-spin-polarized PBE surface, which favors Sqr over Oct, to an FM-initialized PBE+ U surface with the opposite Sqr–Oct ordering along a fixed structural path. When fine-tuned directly, the pretrained DPA-4 models adapted rapidly to the target energies, forces, and virials, reaching energy, force, and virial RMSEs of approximately 0.4–0.5 meV/atom, 30–35 meV/Å, and 5–6 meV/atom, respectively, using 170 PBE+ U training configurations. More importantly, models that had first been fine-tuned to the no- U surface could subsequently recover the qualitative opposing PBE+ U energetic ordering with nearly the same data efficiency as models fine-tuned directly from pretraining. In the separate sequential-adaptation experiment, DPA4-Neo-OMat reached 0.59 meV/atom, 31.96 meV/Å, and 6.13 meV/atom for representative split 0 at the same target training size. Prior learning of a distinct source-protocol energetic ordering therefore does not preclude efficient downstream adaptation. The same conclusion is supported by independent 600 K AIMD trajectories, on which sequential adaptation retains nearly the same accuracy as direct fine-tuning and substantially improves over the zero-shot model [Supplementary Section 2].

Our results also indicate that downstream adaptability is an important criterion for evaluating foundation MLFFs. Foundation MLFFs should be assessed not only by their zero-shot errors but also by how efficiently they can be fine-tuned with limited target data and how stable their performance is across data splits. Under the present fine-tuning protocol, the tested DPA-4 models reach lower target errors with fewer configurations and show less variation among independent data splits than the tested DPA-3 models. Because the models differ in both architecture and pretraining history, the observed difference cannot be attributed uniquely to architecture. The tested DPA-4 models are more adaptive in the present NiO benchmark.

Several limitations define the scope of these conclusions. First, the source and target datasets differ in both correlation treatment and magnetic polarization. The source surface is defined by non-spin-polarized PBE, whereas the target surface is defined by FM-initialized PBE+ U . The process demonstrated here must therefore be interpreted as transfer between two consistently defined electronic-structure reference surfaces, rather than as an isolated correction from DFT to DFT+ U . Second, the present results do not establish that arbitrary source-protocol errors can always be corrected across model architectures, data-selection schemes, and material systems. Transfer may become more difficult when the source data lack relevant coordination environments, the target surface introduces new bonding motifs, or multiple electronic or magnetic states must be represented simultaneously. Finally, an MLFF based only on atomic structure cannot determine the electronic or magnetic state independently. Instead, it learns the coordinate-to-energy mapping represented by the labels generated under the prescribed electronic-structure protocol, while a common initialization protocol does not necessarily guarantee convergence to a unique electronic or magnetic branch for every configuration. The residual near- Sqr discrepancy is best understood in the context of the pronounced electronic/magnetic multistability of Sqr -NiO. This does not alter the central conclusion of the controlled sequential-adaptation experiment, which demonstrates MLFF transferability between prescribed electronic-structure protocols rather than establishing a definitive NiO phase diagram or microscopic mechanism. Consistent magnetic initialization and convergence are therefore essential when constructing target-protocol datasets for correlated materials.

Together, these observations suggest a division of roles between universal pretraining and application-specific fine-tuning. Recent work has emphasized that mixing composition-dependent DFT and DFT+ U conventions within a pretraining dataset can compromise label consistency and has advocated consistent pretraining without U ^[13]. Our results complement this argument by showing that omitting U at the source level need not prevent the resulting model from later adapting to a PBE+ U target surface. Pretraining may therefore prioritize broad chemical and configurational coverage, label consistency, and computational affordability, rather than quantitative agreement with the target-protocol energetic ordering of every possible target material. Before system-specific applications, however, a foundation MLFF should be validated against a consistently generated target-protocol dataset and, when necessary, fine-tuned using a compact set of higher-fidelity calculations. Foundation MLFFs should therefore be viewed not as universally accurate zero-shot calculators but as transferable priors whose practical value depends both on broad prior knowledge and efficient target adaptability.

DECLARATIONS

Acknowledgments

We thank Dr. Yuxiang Gao, Dr. Ruoyu Wu and Dr. Hongyu Wu for helpful discussions regarding the DFT calculations. We also thank DP Technology for providing computational resources through the Bohrium platform and data hosting services via AIS Square.

Author's contributions

Made major contributions to conception and design of the study and performed data analysis and interpretation: Xie, F.

Provided supervision as well as administrative, technical, and material support: Jiang, P.

Provided supervision and revised the manuscript: Zhong, Z.

Availability of data and materials

All DFT data and fine-tuned MLFFs are available on Figshare at <https://doi.org/10.6084/m9.figshare.33191403>.

AI and AI-assisted tools statement

The authors used the Hermes agent (version 0.20.6, released 2026-08-27), powered by the DeepSeek-V4-Flash large language model, solely for language polishing and grammatical checking. All scientific analyses,

calculations, datasets, interpretations, and the original manuscript were produced by the authors. The authors reviewed and verified all AI-assisted revisions and take full responsibility for the content of the manuscript.

Financial support and sponsorship

This work was supported by the National Key R&D Program of China (Grant No. 2021YFA0718900), the National Natural Science Foundation of China (Grant Nos. 12574069, 92477114, and 12374096) and the Suzhou Municipal Science and Technology Program (Gusu Innovation and Entrepreneurship Leading Talent Program, Grant No. ZXL2025304).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

Supplementary Materials

[Supplementary Materials](#)

REFERENCES

1. Anisimov, V. I.; Zaanen, J.; Andersen, O. K. Band theory and Mott insulators: Hubbard U instead of Stoner I . *Phys. Rev. B*. **1991**, *44*, 943. [DOI](#)
2. Cococcioni, M.; de Gironcoli, S. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B*. **2005**, *71*, 035105. [DOI](#)
3. Jain, A.; Hautier, G.; Ong, S. P.; et al. Formation enthalpies by mixing GGA and GGA + U calculations. *Phys. Rev. B*. **2011**, *84*, 045115. [DOI](#)
4. Spagnoli, D.; Refson, K.; Wright, K.; Gale, J. D. Density functional theory study of the relative stability of the iron disulfide polymorphs pyrite and marcasite. *Phys. Rev. B*. **2010**, *81*, 094106. [DOI](#)
5. Long, O. Y.; Sai Gautam, G.; Carter, E. A. Evaluating optimal U for 3d transition-metal oxides within the SCAN+ U framework. *Phys. Rev. Mater.* **2020**, *4*, 045401. [DOI](#)
6. Kam, R. L.; Binci, L.; Kaplan, A. D.; Persson, K. A.; Marzari, N.; Ceder, G. Interplay between electron localization, magnetic order, and Jahn-Teller distortion dictates LiMnO₂ phase stability. *Phys. Rev. B*. **2025**, *111*, 245132. [DOI](#)
7. Jain, A.; Ong, S. P.; Hautier, G.; et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. [DOI](#)
8. Horton, M. K.; Huck, P.; Yang, R. X.; et al. Accelerated data-driven materials science with the Materials Project. *Nat. Mater.* **2025**, *24*, 1522–32. [DOI](#)
9. Schmidt, J.; Wang, H. C.; Cerqueira, T. F. T.; Botti, S.; Marques, M. A. L. A dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals. *Sci. Data*. **2022**, *9*, 64. [DOI](#)
10. Schmidt, J.; Hoffmann, N.; Wang, H. C.; et al. Machine-learning-assisted determination of the global zero-temperature phase diagram of materials. *Adv. Mater.* **2023**, *35*, e2210788. [DOI PubMed](#)
11. Barros-Luque, L.; Shuaibi, M.; Fu, X.; et al. The Open Materials 2024 (OMat24) inorganic materials dataset and models. *Nat. Comput. Sci.* **2026**, *6*, 642–52. [DOI](#)
12. Zhou, F.; Cococcioni, M.; Marianetti, C. A.; Morgan, D.; Ceder, G. First-principles prediction of redox potentials in transition-metal compounds with LDA + U . *Phys. Rev. B*. **2004**, *70*, 235121. [DOI](#)
13. Warford, T.; Thiemann, F. L.; Csányi, G. Better without U : impact of selective Hubbard U correction on foundational MLIPs. *Mach. Learn. Sci. Technol.* **2026**, *7*, 035033. [DOI](#)

14. Wang, R.; Gao, Y.; Wu, H.; Zhong, Z. Pre-training, fine-tuning, and distillation (PFD): automatically generating machine learning force fields from universal models. *Phys. Rev. Mater.* **2025**, *9*, 113802. DOI
15. Zhang, D.; Peng, A.; Cai, C.; et al. A graph neural network for the era of large atomistic models. *npj. Comput. Mater.* **2026**, *12*, 276. DOI
16. Li, T.; Li, W.; Peng, A.; et al. DPA4: pushing the accuracy-cost frontier of interatomic potentials with EMFA SO(2) convolution. *arXiv* **2026**, arXiv:2606.02419. Available online: <https://doi.org/10.48550/arXiv.2606.02419>. (accessed 2026-10-21).
17. Kresse, G.; Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B.* **1993**, *47*, 558. DOI
18. Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15-50. DOI
19. Kresse, G.; Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B.* **1996**, *54*, 11169. DOI
20. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B.* **1999**, *59*, 1758. DOI
21. Ong, S. P.; Richards, W. D.; Jain, A.; et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314-9. DOI
22. Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. *Phys. Rev. B.* **1998**, *57*, 1505. DOI
23. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865. DOI
24. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396. DOI
25. Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and numerically efficient r²SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208-15. DOI
26. Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106. DOI
27. QUESTS 2024.06.21 documentation. github.io. <https://dskoda.github.io/quests/>. (accessed 2026-10-21).
28. Schwalbe-Koda, D.; Hamel, S.; Sadigh, B.; Zhou, F.; Lordi, V. Model-free estimation of completeness, uncertainties, and outliers in atomistic machine learning using information theory. *Nat. Commun.* **2025**, *16*, 4014. DOI PubMed PMC
29. Zhang, D.; Peng, A.; Cai, C. DPA-3.1-3M model. AIS Square model repository. <https://www.aissquare.com/models/detail?pageType=models&name=DPA-3.1-3M&id=343>. (accessed 2026-10-21).
30. Zeng, J.; Zhang, D.; Peng, A.; et al. DeePMD-kit v3: a multiple-backend framework for machine learning potentials. *J. Chem. Theory. Comput.* **2025**, *21*, 4375-85. DOI
31. Kaplan, A. D.; Liu, R.; Qi, J.; et al. A foundational potential energy surface dataset for materials. *arXiv* **2025**, arXiv:2503.04070. Available online: <https://doi.org/10.48550/arXiv.2503.04070>. (accessed 2026-10-21).
32. Kuner, M. C.; Kaplan, A. D.; Persson, K. A.; Asta, M.; Chrzan, D. C. MP-ALOE: an r²SCAN dataset for universal machine learning interatomic potentials. *npj. Comput. Mater.* **2025**, *11*, 352. DOI
33. Shinagawa, C.; Takamoto, S.; Shintani, D.; et al. Matlantis-PFP v8: universal machine learning interatomic potential with better experimental agreements via r²SCAN functional. *arXiv* **2026**, arXiv:2603.11063. Available online: <https://doi.org/10.48550/arXiv.2603.11063>. (accessed 2026-10-21).

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.