Linking the polyatomic arrangements of interstitial H_2O and cations to bonding within Prussian blue analogues *ab initio* using gradient-boosted machine learning

Sizhe Liu[1] and Kyle C. Smith[1,2,3,4,*]

[1]*Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[2]*Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[3]*Computational Science and Engineering Program, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[4]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

(Received 1 December 2020; accepted 16 February 2021; published 8 March 2021)

Prussian blue analogues (PBAs) are model host compounds for the intercalation of monovalent cations for electrochemical energy storage and separations. However, the interactions among interstitial species and their effects on atomic arrangements therein are understood mainly at a phenomenological level. Analyzing correlations between electronic interactions and polyatomic arrangements in hydrated Prussian blue analogues is complicated by the nonlocal hydrogen-bonding interactions between zeolitic water and framework lattices. Here, we train machine-learning (ML) models to learn DFT-calculated energy landscapes of nickel hexacyanoferrate PBA lattices with various lattice hydration degrees, oxidation states, and types of intercalated alkali cations based on various three-particle feature parameters. This ML approach is enabled by using gradient-boosted regression trees with features that are rotationally invariant geometric parameters. ML model accuracy is shown to be a cation-specific indicator of correlations between energy and polyatomic arrangements. Overlap population analysis among correlated atoms further confirms that such correlations are caused by the competition for dative bonding between Lewis-acid intercalated cations and Lewis bases (cyanide and oxygen in H_2O). Examination of lowest-energy structures reveals that cation hydrophilicity and bare ionic radius determine dative-bonding strength, resulting in cation-H_2O ordering in interstitial space. The projected energy landscapes of hydrated PBA lattices is also explored in subspaces spanned by certain many-particle feature parameters inspired by ML analysis. The downhill traces in such landscapes indicate that lattice distortion is accompanied by two kinds of collective movements: (1) rearrangements in the hydration shells around small and hydrophilic cations and (2) collective attack of H_2O molecules on nickel-cyanide bonds promoted by large, hydrophobic cations.

DOI: 10.1103/PhysRevMaterials.5.035003

I. INTRODUCTION

Prussian blue analogues (PBAs) have been and are being used as redox-active host materials in aqueous electrochemical devices [1,2], including electrochemical cation intercalation desalination cells [3–6] and alkali-ion rechargeable batteries [7–9]. PBAs intercalate/deintercalate cations from/into solution when transition metal centers therein are reduced/oxidized. Unlike typical Li-ion host materials, PBAs are hygroscopic as a result of H_2O sorption into defect sites where $Fe(CN)_6$ units are missing [10] and within interstitial vacancies [11] that are also accessible to cations at body-centered sites [12]. While water molecules at such $Fe(CN)_6$ defect sites are relatively inert during electrochemical cycling [10,13], zeolitic water molecules in interstitial vacancies actively interact with intercalated cations and the PBA framework itself [14,15], resulting in lattice distortion [16,17] and disordered spatial arrangement of interstitial species [18]. These effects are important because PBA framework distortion affects long-term cycling performance [19], while the interstitial arrangement of water molecules influences cation storage capacity and intercalation kinetics [20]. Therefore, mechanistic understanding of the role that interstitial H_2O plays in mediating cation intercalation is essential to the development of PBA electrodes with simultaneously high capacity, long cycle life, and high rate capability. Beyond these attributes that are especially important for energy storage applications, PBAs exhibit cation-specific reduction potential [21], resulting in selectivity bias toward monovalent cations with smallest hydrated ionic radius [22]. This attribute of PBAs has enabled their application in selective electrochemical separations for Cs^+ over Na^+ [23,24], K^+ over Na^+ [4], and NH_4^+ over Na^+ [25]. While the hydration of intercalated cations is commonly believed to play a role in determining such selectivity preferences, understanding of bonding between and the arrangements of H_2O, intercalated cations, and PBA frameworks is needed in order to enable the design of future redox-active host materials with molecular recognition toward particular ions of interest.

Past studies have attempted to characterize the interactions between zeolitic water molecules and PBA frameworks using first-principles calculations [14] and *ex situ* experiments [15,26,27]. The x-ray diffraction (XRD) pattern for nickel hexacyanoferrate (NiFe-PBA) has shown evidence of water cointercalation with Na^+ ions and of water's interactions

*Corresponding author: kcsmith@illinois.edu

with cyanide ligands that induce reversible cubic-monoclinic lattice transformations [28]. Recent first-principles calculations have predicted the electronic structure within PBA supercells having two distinct redox-active centers and have shown that the affinity of water molecules to the framework dilates interstitial space, facilitating low-strain cation intercalation [29]. Contrarily, others have found that the presence of water in PBAs deteriorates cycling performance. Fourier-transform infrared spectroscopy (FTIR) spectra have revealed that MnCoNi-PBA alleviates water-induced capacity fading by intercalating dehydrated Na^+ ions [17]. Also, strong Zn-water coordination in a ZnFe-PBA causes the formation of $Fe(CN)_6$ defects during PBA synthesis, exacerbating long-term capacity fading [19]. In the MnCr-PBA lattice interstitial water stretches cyanide ligands, making framework lattice unstable [20]. These conflicting results indicate that water-lattice interactions alone are not sufficient to explain the effects of interstitial H_2O on the electrochemical activity of PBAs. Instead, we show here that water-cation and cation-lattice interactions must be understood in concert due to the confined environment of the interstitial space within PBAs.

Past work has also linked cation-water interactions to cation intercalation kinetics at solid/electrolyte interfaces. Various PBAs show preference for intercalating K^+ ions [15,30] over Li^+ ions [31] as K^+ ions are less coordinated by solvent molecules. Aside from alkali cations, the selectivity of multivalent cation intercalation into PBAs has been characterized [32] with CuFe-PBA, in particular, showing selectivity for Al^{3+} intercalation over Mg^{2+} due to the compact hydration shell around Al^{3+} [33]. Earlier research using NiFe-PBA to intercalate Ca^{2+} from both aqueous and nonaqueous electrolytes showed that reversible capacity is only achievable when the Ca^{2+} are dehydrated [34,35]. The suppressed activation energy of redox kinetics at PBA/electrolyte also implicates interfacial dynamics to the intercalation preferences of PBAs [36].

Bulk interactions between intercalated cations and zeolitic H_2O inside PBA lattices have also been implicated in their electrochemical behavior. H_2O coabsorption has been shown to increase reduction potential for intercalation of Li^+ and Na^+ into PBAs, while intercalation of K^+ is not accompanied by H_2O coabsorption [30]. Recent observations of switchable thermal expansion in PBAs suggest that ordering of cations and water can occur in the bulk of PBAs [37,38] thus differentiating the interactions between cations and water in such interstitial environments from those in bulk liquid electrolytes. Since the spatial confinement of water has recently been shown to suppress its dielectric screening [39], confinement of coabsorbed water within the moderately sized interstitial space of PBAs (∼5 Å) is likely to screen cation-framework interactions to a lesser extent than water would in bulk form. Water coabsorption has been studied in other electrosorption materials, as well. Non-Faradaic MXene materials have shown $\sim 1\,H_2O$ per Li^+ and $\sim 0\,H_2O$ per Na^+ [40]. In contrast, poly(vinylferrocene) films have shown coabsorption of $>9\,H_2O$ per electron during electrosorption of organic anions [41], and after electrosorption of ReO_4^- poly(vinylferrocene) shows relatively weak ReO_4^-/vinyl interactions and no direct ReO_4^-/Fc^+ interactions [42]. We posit here, and later demonstrate using first principles modeling, that cation-specific ordering of interstitial water molecules affects the cation-framework interactions that are responsible for ion recognition by PBAs.

Previous atomistic modeling of PBAs has focused on linking cation-framework interactions to macroscopic phenomena in the absence of water. Supported by the Landau theory and XRD data, the magnetic ordering temperature of the MnMn-PBA lattice has recently been shown to increase with the strength of cation-cyanide interactions [43]. The tautomerism of Fe-CN-metal chains in PBAs has been used to show that the active redox center switches from Fe to Mn when intercalated Rb^+ ions are present to strengthen orbital interactions between Fe and Mn [44]. Building on DFT calculations that revealed the mutual repulsion of interstitial vacancies, we used grand canonical ensemble theory to predict the variation between equilibrium intercalation potential with the degree of Na^+ ion intercalation in an anhydrous NiFe-PBA [11], in agreement with experiment.

To date atomistic modeling of hydrated PBA lattices has focused on particular aspects of cation-water-framework interactions that largely neglect many-particle correlations [38,45] due to the challenge of disentangling polyatomic orbital overlap from atom-specific interactions. To circumvent cumbersome analysis of atomic correlations, ligand field stabilization energy (LFSE) and redox center ionization energy concepts have been applied to PBAs [46], where zeolitic water has been shown to switch the redox sequence of metal centers by increasing LFSE [46]. Lattice stiffness has also been shown to increase by weakening the interaction between water molecules and metal-nitrogen bonds [47]. However, LFSE has limited application to understanding the interstitial interactions of PBAs that contain octahedral low-spin complexes [48] since only half of metal centers therein are redox active. Among the broader class of metal organic frameworks (MOFs) the principles of hard-soft acid-base (HSAB) theory have been applied to rationalize the interactions of interstitial cations with MOF ligands [49]. By surveying the dissociation energies of different metal-ligand bonds, both hard-base/hard-acid and soft-base/soft-acid framework-cation interactions were shown to promote MOF stability, while simultaneously making cation-exchange processes reversible. However, HSAB theory can only predict qualitative trends in the reversibility of cation exchange, as its precise application depends heavily on ligand relaxation energy [50], an empirical parameter that is material/ion specific.

Theories of many-particle correlations have been developed based on density functional theory (DFT) calculations [51], but have not yet been applied to PBAs. Among them, the theory of many-body expansion (MBE) is often used to investigate the interactions between molecular clusters [51–53], where a certain n-body interaction energy is defined as total cluster energy minus the energies of independent n-atom components [51]. Such n-body energies can be calculated using DFT, and each such energy is a characteristic measure of the atomic/electronic interactions outside of nuclear cores. Notably, MBE has been applied to investigate many-body effects in deformed cluster structures [54,55] and in noncovalent bonding among charged molecules [56]. Alternatively, methods other than MBE have used partitioning of DFT electron density fields to analyze many-body interactions

[58,59], including a refinement of the Density Derived Electrostatic and Chemical method (DDEC6) [57] that accounts for both itinerant and localized electrons. Here, the number of partitioned electrons is used to calculate the electron overlap population (OP) and the bond order for distinct pairs of atoms. Using this approach, bond orders have recently been calculated on a wide range of organometallic lattices and have shown that the DDEC6 method reproduces the chemically expected trend of electron transferability [60]. However, basis-set superposition effects in MBE [61] and partitioning of each electron to a single nucleus with DDEC6 [57] limit the effective application of both methods to understanding three-body interactions. Therefore, other methods are needed to analyze the many-particle correlations between the structures of and interactions between interstitial cations, H_2O, and PBA frameworks.

In the current study, we investigate the many-particle correlations in the hydrated NiFe-PBA framework using a method that combines DFT calculation, overlap population analysis, and a machine-learning (ML) model. We first perform lattice relaxation calculations of NiFe-PBA unit cells by varying the type of intercalated cation, the oxidation state of the PBA, and the number of zeolitic water molecules within the PBA. Based on the atomic configurations obtained from DFT calculations, we prepare atomic feature vectors to train an ensemble of regression trees using the Xgboost method [62]. We introduce an importance index (IMP) that enables us to rank the sensitivity of total energy to many-body arrangements. This analysis reveals that the type of intercalated cation, small/hydrophilic or large/hydrophobic, is a critical factor that affects the variation of many-body structural correlations with the degree of hydration. We then use overlap population analysis to link structural correlations between the PBA framework and interstitial species to the extent of dative bonding between (1) intercalated cations and framework ligands and (2) intercalated cations and oxygen in H_2O. Inspection of the lowest-energy structures for a certain cation type, hydration number, and oxidation state enables us to categorize cations of interest into three groups based on their bare ionic radius and hydrophilicity. In particular, we show that bare ionic radius determines the degree of cation-water coordination and that cation hydrophilicity dictates bonding within clusters formed by multiple water molecules. Identification of the most important features from analysis of ML results subsequently enables us to investigate the corresponding DFT-calculated energy landscape of PBAs in subspaces defined by many-body structural features. The downhill traces therein illustrate how atoms in two different clusters move collectively and how the orientation of zeolitic water molecules is affected by its surrounding atoms.

II. METHODOLOGY

In this section we first summarize the challenges of simulating hydrated PBA materials using DFT, while also highlighting certain details of ML model training that relate to DFT. Subsequently, we present details of our DFT calculations and ML modeling methods. We focus our analysis on the NiFe-PBA that exhibits reversible intercalation of alkali cations from aqueous solution [63,64]. DFT calculations for hydrated PBA lattices are particularly challenging for two reasons: (1) they require knowledge of positions for interstitial H_2O that are unknown prior to structural relaxation, and (2) nonlocal dispersion interaction energy functionals that capture hydrogen bonding converge slowly during self-consistent iteration sequences [65,66]. We address the first challenge by sampling H_2O positions over Voronoi tessellation points within the pristine PBA and by sampling different random orientations for each water molecule at those points, as described in Sec. II A. However, the second challenge stems from the nonlocal nature of H-bonding interactions. Because such bonding depends on electron densities and their gradients in a complex fashion [67], DFT simulation of lattices filled with water molecules using dispersion-corrected functionals is computationally expensive. As a result, relatively permissive termination criteria have sometimes been used to relax such structures with reduced computational time. As an alternative, ML models trained using DFT-calculated data have been coupled to global minimizers to relax structures. Once adequately trained, such ML models are used to predict energies rapidly. The artificial neural network (ANN) has been used widely in simulating crystalline materials [68], biological molecules [69], and amorphous materials [70]. The versatility of ANNs can be attributed to their network structure that maps high-dimensional atomic features to target data, and little-to-no prior knowledge is needed during the model training process. We argue, however, that other ML models are better suited to learning the energy landscape of hydrated PBAs for which DFT's computational expense prohibits collection of data sets large enough to train ANNs. Also, the descriptors that prepare input vectors (i.e., feature vectors) for ANNs use kernel functions to normalize vector elements [71,72] that can cause nonunique mapping between feature vectors and atomic configurations [73]. With small training data sets and inaccurate descriptors, ANN-based models are likely to suffer from overfitting [74], which is a problem that cannot be fixed easily by network regularization methods [75]. Compared to ANN-based models, we show here that training of an Xgboost-based model requires much smaller data sets (see Table V) with fewer hyperparameters that need to be fine-tuned [62]. In the following sections, we exploit the unique qualities of the Xgboost method to develop a simple, yet efficient, descriptor using physically interpretable many-particle features to represent atomic configurations in hydrated, cation-intercalated PBAs.

A. DFT calculations for a hydrated NiFe-PBA

The unit-cell configurations that we adopt in all DFT calculations here have a formula of $A_x\mathrm{NiFe(CN)_6}\cdot n\mathrm{H_2O}$, where A is an intercalated alkali cation ($A = \mathrm{Li^+}$, $\mathrm{Na^+}$, $\mathrm{K^+}$, $\mathrm{Rb^+}$, or $\mathrm{Cs^+}$). When electrochemically oxidized the NiFe-PBA unit cell is half intercalated with cations ($x = 1$) and when reduced it is fully intercalated ($x = 2$), as shown in Fig. 1(a). All DFT calculations were performed using the GGA + U approach implemented in the Quantum ESPRESSO (QE) package [76]. The strong on-site Coulombic interactions on Fe and Ni atoms were corrected by setting their respective U parameters to 1 and 3 eV to reproduce Na-intercalated lattice constants

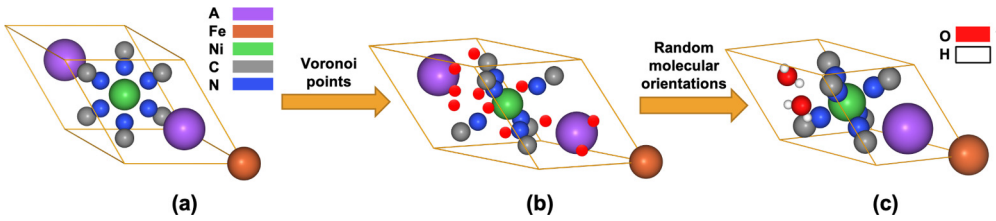

FIG. 1. The three-step scheme used to generate initial configurations for DFT calculations. The first step in (a) determines the number of A cations in the anhydrous unit cell. A collection of potential locations for oxygen atoms shown in (b) is found by performing Voronoi tessellation based on the nuclear coordinates of the relaxed anhydrous lattice. In the last step (c), the locations for oxygen atoms are randomly chosen from candidate locations obtained in (b), and two hydrogen atoms are subsequently attached to each oxygen atom.

reported in previous experiments [77] and to maintain the low-spin state of the Fe redox center [11,18]. The kinetic cutoff energy for wave functions was set to 600 eV, and self-consistent iterations were required to converge within 0.136 μeV. Vanderbilt ultrasoft pseudopotentials [78] were adopted for all atoms except for Cs, for which we used a norm-conserving pseudopotential [79] to avoid the negative charge density caused by the Ultrasoft density interpolation scheme [78]. For all calculations we use the PBEsol+rVV10 functional to approximate exchange-correlation effects and van der Waals interactions to make the calculations of both anhydrous and hydrated configurations consistent. We note that, unlike local dispersion functionals [80], the nonlocal functional rVV10 provides a better correction to molecular binding energy while keeping computational cost modest [81].

We use a sequential scheme to add water molecules within each unit cell with certain degree of intercalation x and cation type A. We first assume body-center occupancy of intercalated cations and subsequently perform variable-cell relaxation calculations on anhydrous unit cells of A_xNiFe(CN)$_6$ (i.e., $n = 0$) in the limit of vanishing solid pressure. We then add water molecules to the relaxed anhydrous configurations to simulate hydrated configurations, as described later. The resulting lattice constants of the anhydrous systems with different kinds of cations and oxidation states are shown in Table I. The NiFe-PBA lattice undergoes isotropic expansion when the bare ionic radius of the cation increases from that of Li^+ to Cs^+. For the Na^+-intercalated configurations, switching from half to fully intercalated unit cells has no effect on lattice constant, consistent with previous experiments [21,77]. For cations other than Na^+, fully intercalated configurations have lattice constants that are slightly smaller than that of half intercalated configurations due to cation-framework interactions. Recent experiments have shown that the lattice constant of a hydrated NiFe-PBA slightly increases or unaffected during K^+ intercalation [77,82,83]. Thus, the results in Table I indicate that the cation-framework interaction is mitigated by interstitial water molecules.

TABLE I. Predicted lattice constant a (in Å) of anhydrous A_xNiFe(CN)$_6$.

	$A = Li^+$	$A = Na^+$	$A = K^+$	$A = Rb^+$	$A = Cs^+$
$x = 1$	5.128	5.100	5.125	5.133	5.139
$x = 2$	5.110	5.100	5.105	5.120	5.133

After performing such variable-cell calculations, we use Voronoi tessellation based on the coordinates of atoms in relaxed anhydrous configurations to define the candidate locations for the oxygen atom of each water molecule to be added to create hydrated initial configurations [red spheres in Fig. 1(b)]. In practice, among all such candidate locations we use those that are at least 0.8 Å away from anhydrous nuclei to limit the number of potential locations for oxygen atoms. For a given configuration we randomly choose n locations among all candidate locations, and we placed a corresponding oxygen atom there. We then attach two hydrogen atoms to each oxygen by setting O-H distances and H-O-H angle to 0.9572 Å and 104.5°, respectively. To do so we sample the orientations of H_2O about each oxygen using random quaternions [see Fig. 1(c)].

Using such initial hydrated configurations we conducted fixed-cell relaxation calculations to obtain relaxed coordinates and associated DFT energies for subsequent Xgboost model training. We constrained unit cell shape and volume during such calculations for two reasons. For the initial configurations that have H_2O molecules in close proximity to atoms on the NiFe-PBA lattice, variable-cell optimization can induce excessive lattice distortion during relaxation sequences. When initial configurations are assembled using the approach already described, we find that such lattice distortion can produce configurations that are potentially far away from equilibrium. Also, fixed-cell calculations constrain the motion of metal centers, enabling the translation of interstitial cations and H_2O and the rotation of cyanide ligands.

We use the nomenclature $A\delta n$ to denote unit cells based on the type of intercalated cation A, the extent of cation intercalation δ (either $\delta = h$ for half intercalated or $\delta = f$ for fully intercalated), and the number of water molecules per formula unit or hydration degree n. For example, Naf4 and Csh2 denote unit cells for $Na_2NiFe(CN)_6 \cdot 4H_2O$ and $CsNiFe(CN)_6 \cdot 2H_2O$, respectively. DFT calculations were performed with as many as four water molecules, resulting in 40 different types of hydrated NiFe-PBA configurations among the five different intercalated cations tested. For each such configuration type, we sampled 50 random realizations of water molecules on Voronoi sites, each of which was subjected to fixed-cell relaxation using the methods already described. The resulting atomic coordinates and self-consistent field energies obtained during the course of each fixed-cell relaxation sequence were subsequently used to train the ML model (i.e., not only those of the terminal relaxed structure itself).

TABLE II. Meta-algorithm of the gradient boosting training method.

Input: $\{(\vec{x}_i, y_i)\}$ and a differentiable loss function $L[y_i, F(\vec{x}_i)]$.
Step 1: Initialize the model with the average energy value $F_0(\vec{x}_i)$.
Step 2: for $m = 1$ to N:
a. Compute residual $r_{im} = \frac{\partial L_m}{\partial F(\vec{x}_i)}\|_{F_{m-1}(\vec{x}_i)}\ \forall \vec{x}_i \in \{(\vec{x}_i, y_i)\}$.
b. Fit a regression tree to r_{im} values and create leaf regions R_{jm}.
c. For $j = 1, \ldots, J_m$, compute $\gamma_{jm} = \underset{\gamma}{\text{argmin}} \sum_{\vec{x}_i \in R_{jm}} L_m[y_i, F_{m-1}(\vec{x}_i) + \gamma]$.
d. Update $F_m(\vec{x}_i) = F_{m-1}(\vec{x}_i) + \nu \sum_j \gamma_{jm} \delta_{\vec{x}_i}(R_{jm})$.
Output: $F(\vec{x}_i)$.

B. Xgboost model

We now summarize the gradient boosting ML method used and subsequently introduce the descriptor that we use to generate the feature vectors that serve as input to our Xgboost ML model. In general, gradient-boosting ML methods train many independent weak regressors (or weak learners) using the same feature vectors, but using different target data from those feature vectors. Such methods subsequently combine such weaker learners to yield an ensemble of learners that spans all feature-vector elements. Table II shows the meta-algorithm of a gradient boosting method used to train an ML model. The algorithm starts with a training data set that is comprised of a set of pairs of feature vectors $\vec{x}_i$ and corresponding target data y_i [i.e., $(\vec{x}_i,\ y_i)$ at the first line of Table II]. In the context of energy-landscape learning, each feature vector $\vec{x}_i$ represents the atomic structure of a certain unit-cell configuration, while y_i is its corresponding energy. The algorithm then trains the first weak learner F_0 that predicts energies using the average of y_i [i.e., $F_0(\vec{x}_i) = \bar{y}$ in Table II]. *Gradient boosting* is accomplished by using the *gradient* of an associated *loss function* to reduce error by training *new weak learners* on *gradient values* after performing training on its output predictions. Accordingly for the mth iteration at step 2 in the algorithm (see Table II), the gradient r_{im} of a loss function L is calculated with respect to its prediction of the ith configuration's output value $F_{m-1}(\vec{x}_i)$ obtained during the previous iteration: $r_{im} = \frac{\partial L}{\partial F(\vec{x}_i)}\|_{F_{m-1}(\vec{x}_i)}$. Presently, we use a loss function defined by the sum of squared errors (SSEs). At the mth iteration we have

$$L_m = \sum_i \frac{1}{2}[y_i - F_{m-1}(\vec{x}_i)]^2, \tag{1}$$

with an associated gradient r_{im} that is readily expressed analytically:

$$r_{im} = y_i - F_{m-1}(\vec{x}_i). \tag{2}$$

From Eq. (2), we observe that r_{im} is equal to the error that results from use of the ensemble of trained weak learners obtained at the $(m-1)$th iteration.

After the values of r_{im} are computed, a new weak learner is trained by taking those r_{im} values as the targeted training data set. Specifically, a regression tree is trained as a new weak learner based on residual errors r_{im} from the previously trained trees in steps 2b and 2c of the algorithm (Table II). To illustrate, Fig. 2 shows an example regression tree for the particular instance in which the mth learner feeds on three features, a, b, and c. In this example we choose the tree to have $J_m = 5$ different predicted output values γ_{jm} for the feature vectors that fall into the associated leaf regions, shown as green blocks denoted by R_{jm}, where $j = 1, 2, \ldots J_m$ enumerates each such leaf. Each node on such a tree subdivides the associated target r_{im} values into two groups based on certain conditions that are feature-specific (blue blocks). In this example, the targets that belong to R_{2m} have corresponding feature vector values that satisfy $a < 1$, $b \leqslant 3$, and $c \geqslant 7$. The morphology of each regression tree is defined by its maximum allowable tree depth $D_{\max}$, its upper limit of leaf-region size n_R, and its leaf algebra Θ. During the training process, a tree is allowed to grow $D_{\max}$ nodes at most before it reaches a leaf region. The tree in Fig. 2 has a depth of 3, which would be produced for $D_{\max} \geqslant 3$. The number of target r_{im} values that feed into a given leaf region defines its leaf region size. Thus, a small n_R value reduces leaf region size and forces the training algorithm to create more leaf regions. The leaf algebra is the particular formula used for calculating the outputs γ_{jm} based on the target r_{im} values. We presently use an average of target r_{im} values to determine γ_{jm}: $\gamma_{jm} = \sum_i r_{im}\delta_{\vec{x}_i}(R_{jm}) / \sum_i \delta_{\vec{x}_i}(R_{jm})$. Here, the Kronecker delta $\delta_{\vec{x}_i}(R_{jm})$ is unity when the vector $\vec{x}_i$ belongs to the leaf region R_{jm} and is otherwise zero. Training such a regression tree is accomplished by changing the threshold values at each node, splitting new branches and pruning redundant branches, so as to minimize the loss function. We will not discuss the numerical details of tree fitting here as they can readily be found elsewhere [62,84]. In practice, we use the open-source Scikit-learn package [85] to train our Xgboost model.

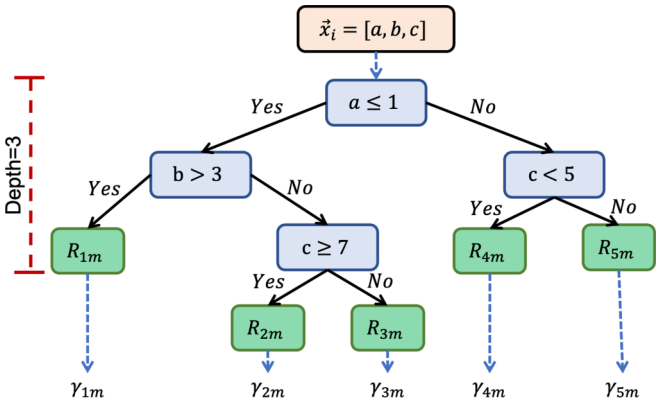

FIG. 2. Illustrative example of a regression tree.

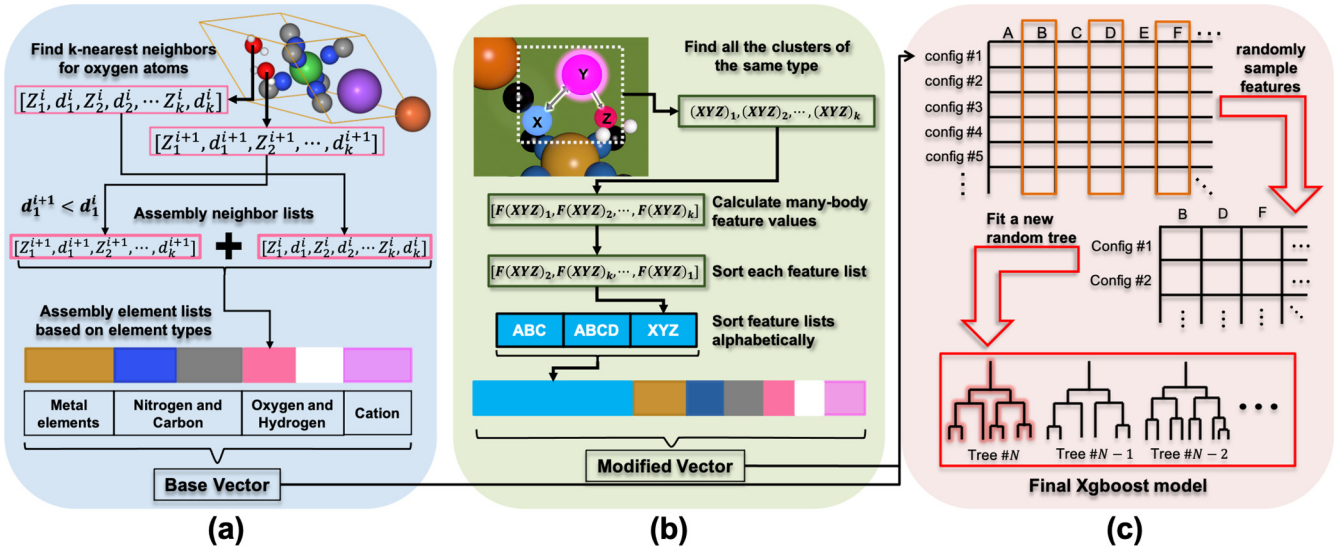

FIG. 3. Schematic of descriptor used for preparing (a) base and (b) modified feature vectors and of the training process for the present Xgboost-based model (c).

The gradient boosting method updates the tree ensemble for the mth iteration by appending the new leaf outcomes to the ensemble for the $(m-1)$th iteration. Accordingly at step 2d of the meta-algorithm, the values of γ_{jm} are weighted by a learning rate parameter ν and the Kronecker delta $\delta_{\vec{x}_i}(R_{jm})$ in updating leaf outcomes. A small ν value increases the accuracy of the ML model by reducing the importance of each tree relative to the entire ensemble of regression trees. The final outcome of the algorithm is an ensemble of N regression trees that yields predicted output values $F(\vec{x})$ for a given feature vector $\vec{x}$.

The Xgboost method uses the gradient-boosting steps described above by adopting the approximate greedy method [86] and a parallel learning scheme [87,88] to build regression trees efficiently. Here, the approximate greedy method grows a regression tree iteratively by evaluating the accuracy gain of subdividing leaf regions at each split-finding step [62]. A new leaf region is only created when it improves the accuracy of the regression tree. Therefore, trees with a larger number of leaf regions provide more accurate predictions of the associated target values. Parallel learning schemes assemble all feature vectors for training into a feature matrix with each row being a feature vector and each column containing the values of the same feature among all feature vectors [see the top matrix in the right subplot of Fig. 3(c)]. The scheme then reproduces N submatrices by randomly sampling $C\%$ distinct columns from the feature matrix [e.g., the matrix at the middle right of Fig. 3(c)]. Each submatrix and its associated targeted output values are later used to train a separate regression tree, while the highly parallelized Xgboost method trains all the trees simultaneously. The final model is an ensemble of regression trees with different structures because they are trained on different features.

1. Descriptor for atomic configurations

We now introduce the descriptor that we use to generate atomic feature vector values on the basis of atomic numbers and polyatomic arrangements. Figures 3(a) and 3(b) show the workflow for assembling feature vectors using the descriptor that we adopt here. For the ith atom in an atomic configuration, the descriptor builds a neighbor list by including the atomic number of each neighbor Z_l^i and its corresponding distance d_l^i, where the integer l ranges from 1 to a prespecifed value k [e.g., the two neighbor lists shown at the top of Fig. 3(a)]. These Z_l^i-d_l^i pairs are subsequently sorted in ascending sequence of d_l^i, where d_1^i is the distance from the ith atom to its nearest neighbor. The descriptor then sorts the neighbor lists for the same atom type in an *element list* based on the values of d_1^i. In Fig. 3(a), the neighbor list of the ith atom is appended behind that of the $(i+1)$th atom because of the condition $d_1^{i+1} < d_1^i$. The descriptor then assembles element lists to prepare a *base vector* based on each element's type. For the NiFe-PBA unit-cell configuration in Fig. 3(a), the lists for Fe and Ni (brown block) are followed by the lists for N (navy), C (grey), O (red), H (white), and intercalated cations (pink).

The use of such base vectors has a number of beneficial attributes. Such base vectors contain the single-particle (i.e., atomic numbers) and two-particle (i.e., atomic distance) information around each atom in an atomic configuration. The size of each neighbor list, $2k$, determines the length of the base vector. The computational cost of preparing base vectors grows linearly with k, which is much cheaper scaling than that produced by the neighboring matrix method that is widely used in ANN-based models [89,90]. The descriptor used here introduces different sorting algorithms for neighbor lists, element lists, and the whole base vector to avoid nonunique mapping between feature vectors and atomic configurations. The preparation of both base and modified vectors does not require normalization, such that finding a suitable set of kernel functions to make each feature differentiable is unnecessary. This convenience is attributed to the fact that we can train a regression tree on both discrete and continuous features.

To prepare modified feature vectors we prepend certain many-particle features to base vectors, as shown in Fig. 3(b). Specifically, we consider modified feature vectors here that include three-particle and four-particle polyatomic structural features, in addition to base vector features. We define three-particle features for a cluster comprised of atoms X, Y, and Z, $F(XYZ)$, in two different ways: (1) the cosine of the bond angle $\angle XYZ$ formed at atom Y or (2) the product $d_{XY}d_{YZ}[1-\cos(\angle XYZ)]$ where d_{XY} is the distance between atoms X and Y. We ultimately use the particular three-particle feature definition that leads to a lower MAE among our Xgboost-based models, as described in the Results and Discussion section. We define a four-particle feature for a G-X-Y-Z cluster, $F(GXYZ)$, using the tangent of half of the solid angle subtended by an X-Y-Z cluster that emanates from atom G:

$$F(GXYZ)=\tan(\Omega_{GXYZ}/2)=\left|\frac{\overrightarrow{GX}\cdot(\overrightarrow{GY}\times\overrightarrow{GZ})}{|\overrightarrow{GX}|\cdot|\overrightarrow{GY}|\cdot|\overrightarrow{GZ}|+(\overrightarrow{GX}\cdot\overrightarrow{GY})|\overrightarrow{GZ}|+(\overrightarrow{GX}\cdot\overrightarrow{GZ})|\overrightarrow{GY}|+(\overrightarrow{GY}\cdot\overrightarrow{GZ})|\overrightarrow{GX}|}\right|. \quad (3)$$

In our subsequent analysis of four-particle features, we only consider Ni and intercalated cations as the G atoms of interest in Eq. (3), and almost exclusively we consider HOH as XYZ in $GXYZ$ clusters. These choices for G atoms are made with inspiration from hard-soft acid-base (HSAB) theory [49], since Ni^{2+} and A^{+} are Lewis acids that exhibit different affinities to basic oxygen in H_2O. For this reason, H_2O plays a role of electron donor in such clusters and causes the formation of dative bonds [91,92] among interstitial species. Indeed, our charge distribution analysis in the Results and Discussion section shows that cations and zeolitic H_2O share electrons in interstitial space. For each such polyatomic cluster of the same kind in a given configuration, the associated ML descriptor calculates a separate many-particle feature and sorts the corresponding values of all such features in a many-particle feature list [e.g., the feature list at the middle of Fig. 3(b)]. If a modified feature vector contains multiple many-particle feature lists, the descriptor assembles them alphabetically based on their associated polyatomic clusters' atomic symbols. In Fig. 3(b), the descriptor appends feature lists of ABC and $ABCD$ clusters before the list of XYZ cluster to yield ascending alphabetical order. By considering different polyatomic arrangements, we use the descriptor to prepare different feature matrices that contain different many-particle feature columns, while fixing the total number of feature vectors [i.e., the row number of feature matrices shown in Fig. 3(c)].

2. *Fine-tuning hyperparameters of Xgboost-based models*

Compared to ANN-based models, the present Xgboost model requires only a small set of hyperparameters whose values must be chosen in order to perform training: (1) the total number of regression trees in the Xgboost ensemble N, (2) the maximum depth of each regression tree $D_{\max}$, (3) the percentage of sampled columns from the total feature matrix for training each regression tree $C\%$, (4) the learning rate ν, (5) the size of neighbor lists k, and (6) the maximum allowable size of each leaf region n_R. However, different choices of hyperparameters also affect the minimum level of error that can be achieved. Thus, we fine-tune hyperparameters to optimize the accuracy of our Xgboost-based model by using a grid search method [85] to find the combination of hyperparameters that minimizes the mean absolute error (MAE) of a separately trained model for a given configuration type. Starting from a coordinate grid in the space spanned by all hyperparameters, the grid search method optimizes a certain hyperparameter by applying the steepest descent algorithm while keeping other hyperparameters fixed. Presently, we optimize hyperparameters in the sequence of k, ν, N, $C\%$, $D_{\max}$, and n_R because we find that the sensitivity of MAE to each such parameter decreases in the same sequence. At each iteration step, the method follows an 80-20 rule [84] to split base vectors and their associated targets into two data sets for training and testing, respectively. After training a different Xgboost ensemble for each configuration type, we evaluate MAE based on the testing data set. We terminate the search and confirm the best hyperparameter combination when the MAE difference between two consecutive iteration steps is less than 0.001 eV/atom. The optimal values for N, $D_{\max}$, $C\%$, ν, and n_R found for all configuration types are the same (Table III), but their optimal k values are specific to each configuration type (Table IV).

Two factors affect the magnitude of k values in Table IV. For half intercalated configurations, the optimal k value generally increases with increasing number of water molecules n. For example, compare the variation of k values from Lih1 to Lih4 with their variation from Nah1 to Nah4. Such correlation of k with n is not observed for fully intercalated configurations, however. The number of base vectors in different training data sets also affects the k values determined by finite-tuning. For Af1 configurations, the fine-tuned k value decreases from 11 for Naf1 to 6 for Rbf1, while the total number of base vectors decreases from 1538 to 836. The descriptor used here prepares modified vectors by appending many-particle features to their corresponding base vectors that have their lengths defined by the k values listed in Table IV.

We now briefly compare and contrast the use of Xgboost for learning energy landscapes with other kernel regression and ANN methods. Table V shows the results of a qualitative survey of the typical MAE levels that have been demonstrated in the past using such methods, revealing that Xgboost is able to achieve similar MAE to such methods while using a smaller data set to train without overfitting or underfitting. However, we acknowledge that such methods were used in the past for a variety of atomic systems that are different from the focus of

TABLE III. Fine-tuned values of hyperparameters.

$C\%$	ν	N	$D_{\max}$	n_R
60%	0.05	900	6	2

TABLE IV. Fine-tuned k values for all kinds of unit-cell configurations.

	$A = Li^+$	$A = Na^+$	$A = K^+$	$A = Rb^+$	$A = Cs^+$
Ah1	9	8	9	9	8
Ah2	7	6	10	10	12
Ah3	11	9	10	9	9
Ah4	11	9	9	12	8
Af1	9	11	9	6	7
Af2	10	8	9	6	10
Af3	8	10	7	9	7
Af4	9	11	9	6	7

the present study, namely hydrated PBAs. Kernel regression methods, such as kernel ridge regression and support vector regression (SVR), have been applied to learn the energy landscapes of organic molecules [93] and amorphous alloys [94]. To achieve sufficient accuracy using kernel regression, atomic features have been constructed in the past using smoothed basis functions [95], which is a practice that can lead to nonunique mapping between feature parameters and atomic configurations. On the other hand, the intricate networks of ANN-based models have yielded accurate energy predictions for small organic molecules [97], metal oxide lattices [72], and amorphous glassy systems [70], but in general a large pool of feature vectors or feature matrices is required to train ANNs properly relative to other methods. Recent work has also shown that supplementation of certain potentials with ANNs has yielded improved accuracy of energy prediction for systems with nonlocal interactions [100]. To directly benchmark the performance of Xgboost against SVR and ANN models we also trained such models using data from each of the 40 different hydrated PBA unit cells investigated here by incorporating the same base features as in our Xgboost models. To do so the hyperparameters of each SVR model were fined-tuned using the grid search method, while we adopted the structure of each ANN based on its recent application to heterogeneous crystalline materials [101,102]. Among all unit cells for which ML models were created, the MAEs of SVR and ANN models on average were respectively 0.174 and 0.132 eV/atom, while Xgboost achieved 0.016 eV/atom on average. Thus, we contend that Xgboost is highly suitable for learning the energy landscapes of hydrated PBAs and other condensed-matter systems, though its past use was limited to organic molecules [87].

III. RESULTS AND DISCUSSION

The aim of this work is to discover and elucidate the electronic and atomic interactions that affect the energetics and interstitial structures formed within hydrated NiFe-PBAs that are intercalated with alkali cations. To that end we have sampled the configurational space for NiFe-PBAs with varying degrees of hydration, oxidation state, and intercalated cation type. We then trained a gradient-boosted ML model to identify the particular many-body structural features with highest correlation to DFT-predicted energy using that model. Descent paths in the corresponding DFT-energy landscapes are subsequently analyzed using such many-body structural feature parameters. We use overlap population analysis in tandem to characterize the competition for dative bonding between intercalated cations and framework ligands and between intercalated cations and H_2O.

We also note that, subsequently, we use the nomenclature $A\delta n$ to denote unit cells using a certain intercalated alkali cation A, a certain extent of cation intercalation δ (either $\delta = h$ for half or $\delta = f$ for fully intercalated), and a certain number of water molecules per formula unit or hydration degree n.

A. Correlating DFT energy with three-particle features

We now analyze the correlation between the many-body structural features of different polyatomic arrangements within a hydrated PBA and DFT-predicted energy by introducing an importance index (IMP), defined as the relative change in mean absolute error (MAE) of an Xgboost model trained on modified feature vectors incorporating a base vector and certain many-body features from that of an Xgboost model solely trained on a base feature vector:

$$\mathrm{IMP}_m = \frac{\mathrm{MAE}_0 - \mathrm{MAE}_m}{\mathrm{MAE}_0}. \quad (4)$$

Here, MAE_m and MAE_0 respectively correspond to Xgboost models trained using modified feature vectors with many-body feature m appended and using a base feature vector.

TABLE V. Qualitative survey of attributes for different ML models used for energy landscape regression.

Model type	Type of atomic systems	Typical MAE (eV/atom)	Number of feature vectors or matrices for training	Input format
Kernel regression	Organic molecules [89,93] and alloys [94,95]	0.006 [95] ~ 0.043 [89]	1000 [89] ~ 26500 [95]	vectors/matrices
Artificial neural network	Molecules [96,97], crystals [96,98] and interfaces [99]	0.02 [99] ~ 0.075 [98]	10000 [96] ~ 5500000 [97]	vectors/matrices
Xgboost	Molecules [87] and PBAs [present]	0.016 [present]	1300 [present]	vectors

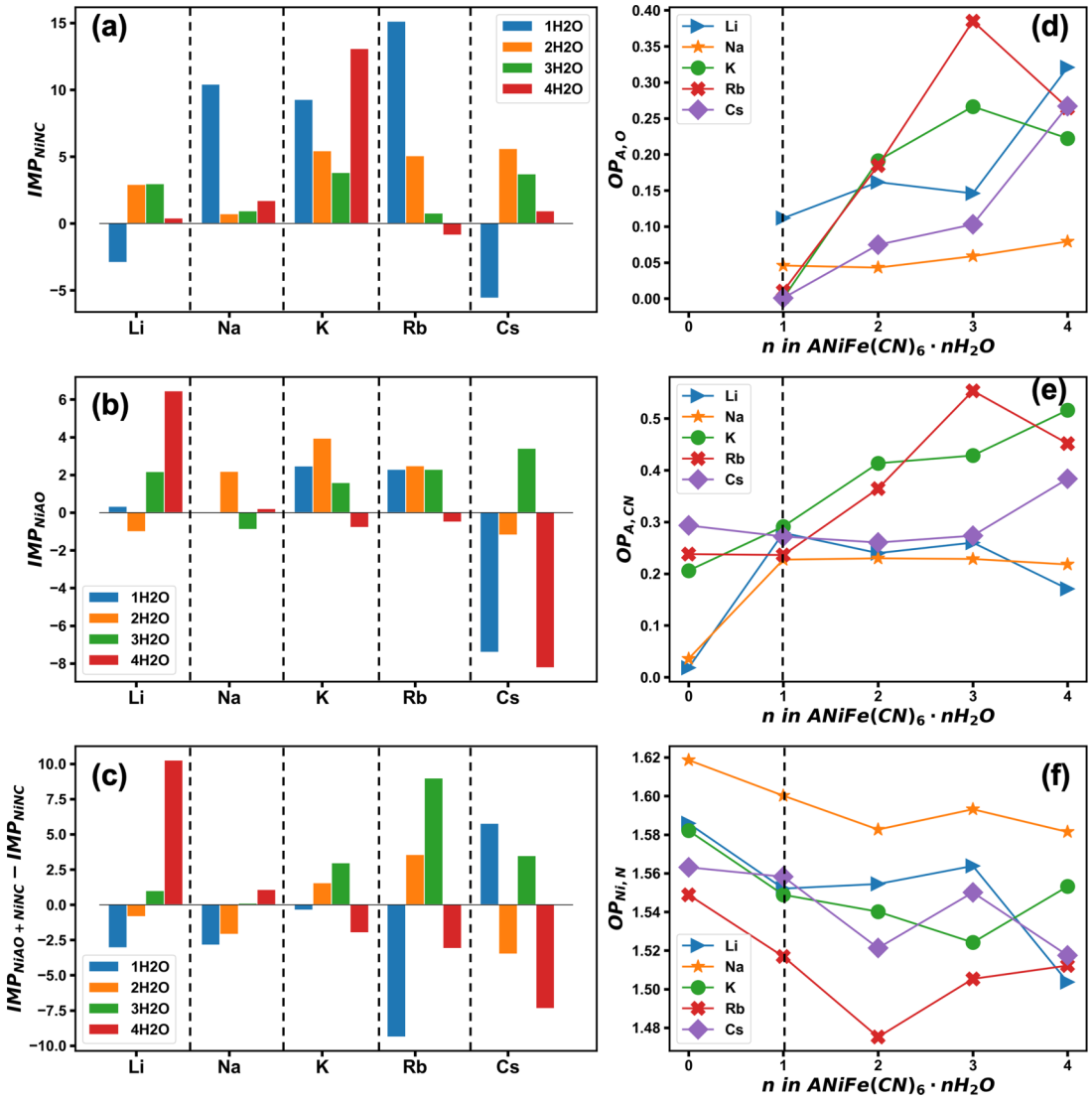

FIG. 4. (a)–(c) Importance index values using three-body feature parameters for Ni-N-C and Ni-*A*-O clusters to train Xgboost models for half intercalated systems. (d)–(f) Results of overlap population analysis for *A*-O, *A*-CN, and Ni-N.

Subsequently, we rank the correlation between energy and certain many-body feature parameters based on the magnitude of their corresponding IMP_m values, where the polyatomic arrangements of certain m clusters are *important* if IMP_m is positive.

We select candidate three-particle clusters that could be important in either half or fully intercalated configurations as pairs of Lewis acid cations ($Ni^{2+} > A^+ = Li^+$, Na^+ K^+, Rb^+, or Cs^+) and Lewis base atoms (O > N > C) that have varying degrees of chemical hardness. These clusters can contain either two basic atoms and one acidic atom or one basic atom and two acidic atoms: Ni-N-C, Ni-*A*-O, *A*-O-*A*, N-*A*-N, and O-*A*-O. This approach enables us to later interpret the structural correlations observed on the basis of HSAB theory [49]. We analyze the polyatomic arrangements of half and fully intercalated configurations separately in the next two subsections because the many-body feature parameters that correlate with energy mostly depend on the degree of cation intercalation, as we now show.

1. *Half intercalated configurations*

Figures 4(a)–4(c) show the IMP values for different cation types, arranged in ascending order of the number of water molecules used. Among the three-body feature parameters tested (NiNC, Ni*A*O, N*A*N, *A*O*A*, and O*A*O), most half intercalated configurations only showed positive IMP_m values for NiNC and Ni*A*O [Figs. 4(a) and 4(b)], indicating high sensitivity of DFT-calculated energy to three-body features of $F(\mathrm{NiNC}) = \cos\angle\mathrm{NiNC}$ and $F(\mathrm{NiAO}) = d_{\mathrm{NiA}}d_{AO}\cos\angle\mathrm{NiAO}$. In the configurations with the same hydration degree n [i.e., bars with the same color in Figs. 4(a)–4(c)], the influence of cation type on the importance of NiNC and Ni*A*O cluster features is not evident, as $\mathrm{IMP}_{\mathrm{NiNC}}$ and $\mathrm{IMP}_{\mathrm{NiAO}}$ fluctuate between negative and positive values when the intercalated cation type changes from Li^+ to Cs^+. While Rb-intercalated configurations show $\mathrm{IMP}_{\mathrm{NiNC}}$ that increases monotonically with increasing hydration degree n, no other intercalated cations show monotonic trends of $\mathrm{IMP}_{\mathrm{NiNC}}$ or of $\mathrm{IMP}_{\mathrm{NiAO}}$ with n. The lack of clear trends for $\mathrm{IMP}_{\mathrm{NiNC}}$ and

IMP_{NiAO} with n and cation type suggests that the correlation between the corresponding three-particle features and DFT energy is affected not only by lattice hydration degree and cation types, but also by other polyatomic features. In order to study the correlation between DFT energy and multiple many-particle features simultaneously, we use the present ML approach to incorporate feature parameters for NiNC and Ni*A*O clusters in a single model.

Figure 4(c) shows the difference between the corresponding IMP values obtained using both three-body features $IMP_{NiNC+NiAO}$ and IMP_{NiNC}, where $IMP_{NiNC+NiAO}$ values were obtained from training and testing of an Xgboost model using three-particle features for NiNC and Ni*A*O clusters. We argue here and later demonstrate by examining DFT energy landscapes that a positive value of $IMP_{NiNC+NiAO} - IMP_{NiNC}$ implies that combining the features of both such clusters improves ML accuracy to a greater extent than including only the feature of NiNC clusters. A positive $IMP_{NiNC+NiAO} - IMP_{NiNC}$ value also indicates possible correlation between the polyatomic arrangements of NiNC and Ni*A*O clusters.

The value of $IMP_{NiNC+NiAO} - IMP_{NiNC}$ increases with hydration degree for Na- and Li-intercalated PBA lattices [Fig. 4(c)], suggesting increasing correlation between the arrangements of NiNC clusters and the NiAO clusters with increasing n. For K- and Rb-intercalated lattices, however, such correlation is only significant with intermediate water content, i.e., $n = 2$ and 3. Cs-intercalated PBA lattices show no apparent trend of $IMP_{NiNC+NiAO} - IMP_{NiNC}$ with n, fluctuating between positive and negative values with increasing n. On the basis of these observations we group these five cations into three distinct categories: light cations Li^+ and Na^+, intermediate cations K^+ and Rb^+, and heavy cations Cs^+. We show later that such behavior is caused by two qualities that vary between these categories: cation hydrophilicity and bare ionic radius.

The cation-dependent correlation between the feature parameters for NiNC and Ni*A*O clusters suggests that cation type affects lattice distortion (e.g., tilting of metal-ligand octahedra [103]) and cation-water arrangement in half intercalated configurations. On this basis we assert that the dominant physiochemical interactions within both clusters are cation specific. Accordingly, we now investigate the relationships between cation type and dative bonding (i.e., coordinate covalent bonding) between Lewis acid and base atoms in PBAs by calculating electronic overlap population (OP), which is a measure of the shared electronic density between two kinds of atoms [57] and is defined for a pair of atoms or clusters, P and Q:

$$OP_{P,Q} = \sum_{i\in P, j\in Q} \int \frac{\rho(\vec{r}_i)\rho(\vec{r}_j)}{\rho(\vec{r})} d\vec{r}. \quad (5)$$

Here, $\rho(\vec{r})$ and $\rho(\vec{r}_i)$ are respectively the total charge density and the charge density assigned to the ith P atom at $\vec{r}$ using the DDEC6 charge density partition method [60]. For example, dative bonding between intercalated cations and oxygen atoms in H_2O is quantified by $OP_{A,O}$ and between intercalated cations and cyanide ligands is quantified by $OP_{A,CN}$. Figures 4(d) and 4(e) show the corresponding results of $OP_{A,O}$ and $OP_{A,CN}$ for the lowest-energy configurations for each half intercalated unit cell with certain hydration degree n. In the anhydrous configuration (i.e., $n = 0$), the values of $OP_{A,CN}$ follow a sequence of $OP_{Li,CN} \approx OP_{Na,CN} < OP_{K,CN} \approx OP_{Rb,CN} < OP_{Cs,CN}$, which groups the cations into three distinct categories that coincide with the results of our more exhaustive IMP analysis. From $n = 0$ to $n = 1$, the values of $OP_{Li,CN}$, $OP_{Na,CN}$, and $OP_{K,CN}$ increase while $OP_{Rb,CN}$ and $OP_{Cs,CN}$ remain nearly unchanged. This observation implies that zeolitic H_2O in less hydrated framework intensifies cation-framework interaction for Li-, Na-, and K-intercalated lattices, but it has a negligible effect on heavy ions. Compared with results with $n \leqslant 1$, $OP_{K,CN}$, $OP_{Rb,CN}$, $OP_{K,O}$, and $OP_{Rb,O}$ rise to larger values at higher hydration degrees. We thus conclude that adding zeolitic water molecules to a PBA intercalated with either K^+ or Rb^+—in contrast with Li^+ and Na^+ as we subsequently show—enhances dative bonding between intercalated cations and framework ligands. In contrast, $OP_{Li,CN}$ and $OP_{Na,CN}$ either decrease or are practically constant with increasing n [Fig. 4(e)], while $OP_{Li,O}$ and $OP_{Na,O}$ increase with increasing hydration degree n [Fig. 4(d)]. These results indicate that the ascending NiNC-Ni*A*O correlation observed in Fig. 4(c) for Li- and Na-intercalated configurations therefore results both from enhanced cation-water interactions and from diminishing cation-octahedron ligand interactions as zeolitic water content increases. For Cs-intercalated configurations, however, the trends of $OP_{Cs,O}$ and $OP_{Cs,CN}$ are not correlated with the variations of $IMP_{NiNC+NiCsO} - IMP_{NiNC}$ with n.

To understand such anomalous behavior for Cs-intercalated configurations, we also calculated $OP_{Ni,N}$ values to reveal changes in Ni-N bonding in the PBA lattice as a function of hydration degree n [Fig. 4(f)]. When $n \leqslant 1$, $OP_{Ni,N}$ reduces with hydration degree noticeably for all the cases except for Cs-intercalated lattice. Starting from $n = 1$, $OP_{Ni,N}$ varies synchronously with $IMP_{NiNC+NiAO} - IMP_{NiNC}$ in Fig. 4(c) for Cs-intercalated configurations, both of which are linked to Ni-N bonding that stretches and contracts in an alternating manner depending on hydration degree. For K- and Rb-intercalated configurations, in contrast, the single local minimum observed for $OP_{Ni,N}$ varies opposite to the variations of $IMP_{NiNC+NiAO} - IMP_{NiNC}$ in Fig. 4(c), where the minimum value of $OP_{Ni,N}$ coincides with positive values of $IMP_{NiNC+NiAO} - IMP_{NiNC}$. For light cations, however, $OP_{Ni,N}$ varies weakly for $n \leqslant 3$, while for $n = 4$ $OP_{Ni,N}$ decreases. Because the correlations between NiNC and Ni*A*O clusters are most significant at $n = 4$ for both Li- and Na-intercalated configurations [see Fig. 4(c)], the drop in $OP_{Ni,N}$ at $n = 4$ implies that sufficiently large cation-water clusters act to stretch Ni-N bonds and cause lattice distortion in half intercalated configurations.

To explain the different variations of $OP_{Ni,N}$ for different cation types at high water content, we now analyze the structures of lowest-energy configurations for Lih3, Lih4, Kh3, Kh4, Csh3, and Csh4 unit-cell types in Fig. 5. Observation of the arrangement of interstitial cations and water in Fig. 5 enables us to reconcile the contrasting results deduced for different intercalation cations from variations $OP_{Ni,N}$ with n by considering a competition between two factors: cation bare ionic radius and hydrophilicity. Because light cations are relatively small and hydrophilic, Li^+ and Na^+ ions are

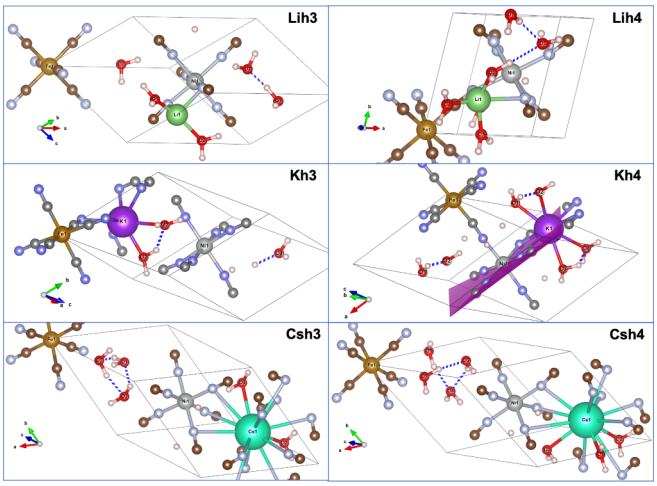

FIG. 5. Lowest-energy structures for Lih3, Lih4, Kh3, Kh4, Csh3, and Csh4 configurations.

firmly coordinated by zeolitic water molecules (e.g., Lih3 and Lih4 configurations in Fig. 5). As hydration degree increases, a given cation-H_2O cluster grows in apparent size, so as to cause stronger interactions between that cluster and the surrounding framework's lattice. For this reason, a monotonically increasing trend of $\mathrm{IMP}_{\mathrm{NiNC+NiAO}} - \mathrm{IMP}_{\mathrm{NiNC}}$ with increasing n is observed for light cations. From Lih3 to Lih4 in Fig. 5, we observe that water molecules form hydration shells around Li^+, causing the Ni-N bond to stretch from 2.02 to 2.06 Å while the average value of ∠NiNC (i.e., ⟨∠NiNC⟩) reduces from 173° to 169°. These observations are consistent with previous experiments that confirm hydration of Li^+ in the interstitial space of ternary PBA frameworks [104]. In contrast, K^+ and Rb^+ are hydrophobic, and Fig. 5 reveals that these ions are displaced to body centers to adjacent sites with minimal energic costs. In the Kh3 configuration, water molecules and nitrogen both closely coordinate to K^+ ions with average values of $d_{\mathrm{KO}} = 2.44$ Å and $d_{\mathrm{KN}} = 3.09$ Å, resulting in enhanced NiNC-NiKO structural correlation. In the Kh4 configuration, K-O coordination becomes weaker with $d_{\mathrm{KO}} = 2.93$ Å, while ⟨∠NiNC⟩ shifts from 170.9° in Kh3 to 171.2° in Kh4. Therefore, the suppressed NiNC-NiKO structural correlation for $n = 4$ in K- and Rb-intercalated lattices results both from interstitial water molecules detaching from intercalated cations and from enhanced cation-ligand coordination. We also note that the Kh4 configuration exhibits K-N coordination in the same plane as Ni-N coordination (i.e., purple plane in Fig. 5), resulting in off body-center occupation by K^+. For Cs-intercalated configurations, both cation-water and cation-cyanide interactions become stronger as hydration degree increases. Because Cs^+ has a large bare ionic radius and is hydrophobic, it pushes all zeolitic water molecules into the vacant body-center site while still retaining coordination to those molecules. The average distance of d_{CsO} is 2.87 Å in Csh3 and 2.98 Å in the Csh4 configurations shown in Fig. 5, both of which are shorter than the distance between Cs-ion and its first hydration shell in aqueous solution [105]. The zeolitic water molecules in the Cs-intercalated lattice interact strongly with each other through hydrogen bonding, indicated using blue-dashed lines in Fig. 5. Notably, the characteristic diameter of three-water clusters is significantly smaller than that of four-water clusters in Cs-intercalated systems (i.e., 2.61 Å vs 3.19 Å). The concomitant reduction of ⟨∠NiNC⟩ reduces from 179.3° in Csh3 to 173.8° in Csh4, which suggests that the fluctuating nature of NiNC-NiCsO correlation results from the formation of a kind of amorphous phase transition of water clusters [106] with increasing hydration degree.

Our analysis of overlap population shows that NiNC-NiAO structural correlation is linked to the competition for dative bonding between intercalated Lewis-acid cations and the Lewis-base groups surrounding them (O in H_2O and N in CN). In addition, analysis of the correlation between the associated three-body feature parameters and DFT energy, quantified using the accuracy of ML models incorporating such features, reveals three distinct categories of cations. Further, inspection of lowest-energy structures reveals that the bare ionic size and hydrophilicity of different cations determine how cations are classified. While these results point to the key bonding interactions that occur among different classes of cations at equilibrium positions, we next seek to interpret how the concerted motions of atoms occur in the vicinity of those local minima on energy landscapes.

We now attempt to investigate the implicit dynamics occurring during the collective motions of polyatomic clusters in the vicinity of local minima. To do this we project high-dimensional DFT-predicted energy landscapes into two-dimensional (2D) subspaces spanned by the most important three-particle feature parameters, as determined from our preceding analysis. In particular, the values of $\mathrm{IMP}_{\mathrm{NiNC}}$ and $\mathrm{IMP}_{\mathrm{NiAO}}$ were found to be positive for most unit-cell types. Therefore, we project the energy landscapes using the subspace spanned by the average values of cos(∠NiNC) and $d_{\mathrm{ANi}}d_{AO}[1 - \cos(\angle \mathrm{NiAO})]$ for a given structure. Each such projected landscape in Fig. 6 is shown with data from the 100 structures with lowest energy for a given unit-cell type. The hue of each data point represents the energy difference ΔE between a given configuration and the lowest-energy configuration in the unit of eV. By following the downhill traces on these projections, we can infer the collective movements of the NiNC and NiAO clusters in the quasistatic limit.

Based on the three categories of different cations that we defined earlier, we group the different DFT energy landscapes, as shown in Fig. 6. For small, hydrophilic cations (red rectangle: Li- and Na-intercalated configurations), the labeled downhill traces start from an energy with $\Delta E \sim 0.3$ eV and descend toward certain local minima. By connecting the tail to the head of each such trace using a straight line, we calculate the secant slope of each such trace $|\Delta y/\Delta x|$. We identify two kinds of traces based on such secant slope values. Steep traces with $|\Delta y/\Delta x| > 0.05$ are red, and they indicate significant that changes in the degree of Ni octahedron tilting are accompanied by mild changes in cation-water coordination along the corresponding trace. In contrast, shallow traces with $|\Delta y/\Delta x| < 0.01$ are black, and they indicate that water molecules rearrange themselves within a given cation-water cluster without causing significant octahedral tilting. As water content increases in Li- and Na-intercalated configurations (i.e., by traversing from the left to the right side of cases in the red rectangle), we observe that shallow (black) traces vanish at

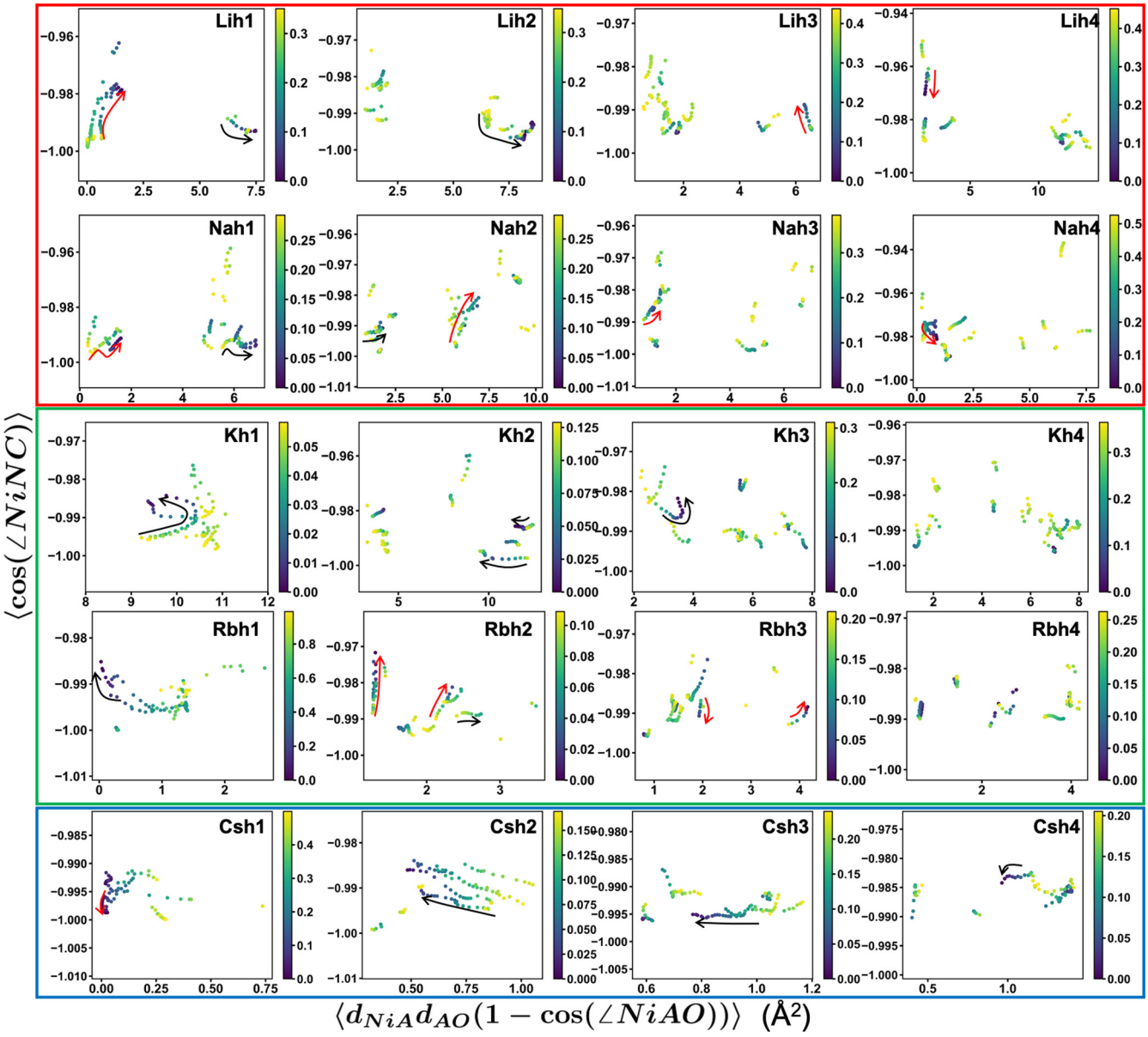

FIG. 6. 2D projections of DFT-predicted energy landscapes for half intercalated configurations near local minima. Color bars indicate DFT energy in eV per formula unit, relative to that of the lowest-energy structure for a given configuration type.

certain n. In contrast, the evolution of octahedral tilting along steep traces involves virtually no rearrangements of cation-H_2O clusters for a sufficiently large hydration degree ($n \geqslant 3$).

For intermediate-size, hydrophobic cations (green rectangle: K- and Rb-intercalated configurations), downhill traces start from energies with $\Delta E \sim 0.08$ eV and follow more circuitous paths than those on the landscapes for Li- and Na-intercalated configurations. The mild variations of $\langle\cos(\angle\text{NiNC})\rangle$ along such traces show that interactions between the framework and interstitial speces in K- and Rb-intercalated configurations do not cause significant lattice distortion during relaxation toward local minima. The shallow (black) traces in the landscapes of K-intercalated configurations demonstrate that lattice distortion evolves in concert with K-H_2O coordination. The steep (red) traces in the landscapes of Rb-intercalated configurations indicate that Rb-H_2O coordination remains unaffected during relaxation toward local minima. At $n = 4$, the landscapes of Kh4 and Rbh4 are absent of downhill traces, however. Instead, all configurations are concentrated near certain local minima, indicating that the motion of interstitial species is impeded by crowding of H_2O in the PBA framework.

The energy landscapes of large, hydrophobic cations (blue rectangle: Cs-intercalated configurations) also show negligible octahedral tilting [i.e., $\langle\cos(\angle\text{NiNC})\rangle \leqslant -0.98$]. However, large changes of cation-water coordination occur along traces. These changes are consistent with the atomic structures of Cs-intercalated configurations (Fig. 5), where cation-water coordination is coupled to the amorphous ordering of water clusters in adjacent interstitial sites. These results therefore confirm that bare ionic radius and the hydrophilicity of intercalated cations not only affect bonding at local minima, they also affect the collective movement of interstitial species in the vicinity of local minima.

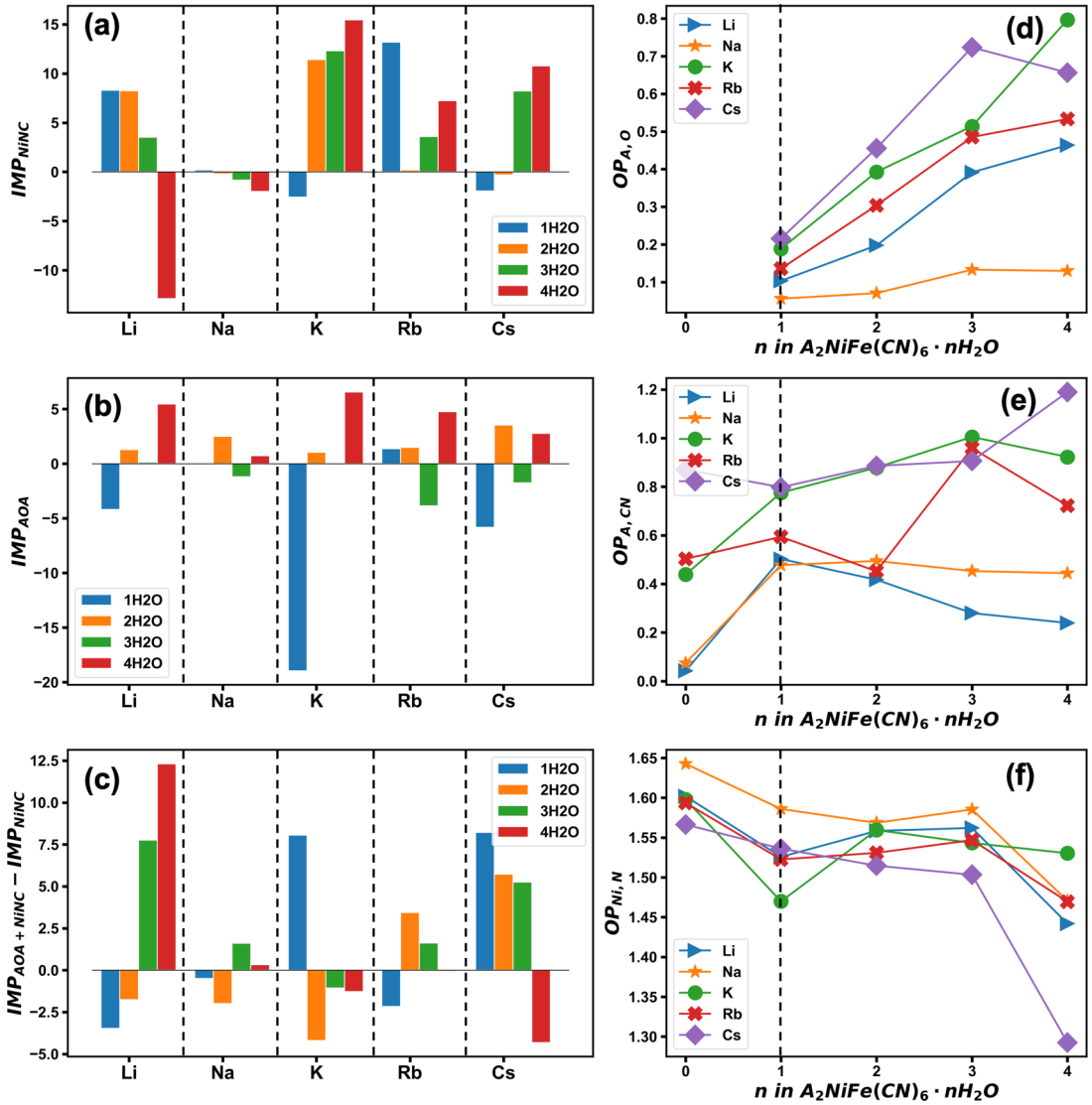

FIG. 7. (a)–(c) Importance index values using three-body feature parameters for Ni-N-C and *A*-O-*A* clusters to train Xgboost models for fully intercalated systems. (d)–(f) Results of overlap population analysis for *A*-O, *A*-CN, and Ni-N.

2. *Fully intercalated configurations*

As we now show, cations in fully intercalated configurations occupy the two body-center sites of the PBA lattice, irrespective of intercalated-cation type or the degree of hydration. This effect occurs because much less space is available for the occupation of interstitial species when two cations per formula unit are intercalated. While for half intercalated configurations we found that the NiNC and Ni*AO* were the three-body clusters showing most significant correlation to DFT energy, for fully intercalated configurations we find that NiNC and *AOA* clusters show most significant correlation. This change in correlation for fully intercalated configurations arises from increased confinement experienced by cations and zeolitic water molecules that enhances interactions between the framework and those interstitial species. Accordingly, we use IMP values for NiNC clusters to analyze the effects of hydration degree and cation types on octahedral tilting. Figure 7(a) shows that the IMP value obtained using the feature $F(\mathrm{NiNC}) = \cos\angle\mathrm{NiNC}$ decreases with increasing hydration degree for Li- and Na-intercalated configurations, whereas it increases with n for K- and Cs-intercalated cells. The increasing trend of $\mathrm{IMP_{NiNC}}$ for Rb-intercalated cells does not persist for all n, as the value of $\mathrm{IMP_{NiNC}}$ for the Rbf1 cell is much larger than those of Rbf2, Rbf3, and Rbf4 cells. Our earlier findings for the half intercalated configurations showed that K and Rb ions belong to the same category of intermediate cations. However, the difference in the trends of $\mathrm{IMP_{NiNC}}$ between K- and Rb-intercalated configurations that we observe here is inconsistent with such classification.

To examine whether the presence of an additional Rb ion makes Rb-water-framework interactions patently different from K-water-framework interactions, we calculated the corresponding values of $\mathrm{OP_{A,O}}$ and $\mathrm{OP_{A,CN}}$ for the associated lowest-energy configurations. Comparing between Rb- and K-intercalated configurations, we observe that both $\mathrm{OP_{K,O}}$ and $\mathrm{OP_{Rb,O}}$ increase monotonically with increasing n [Fig. 7(d)]. Also, the values of $\mathrm{OP_{K,CN}}$ and $\mathrm{OP_{Rb,CN}}$ both decrease significantly for $n \geqslant 3$ [Fig. 7(e)]. The lowest-energy

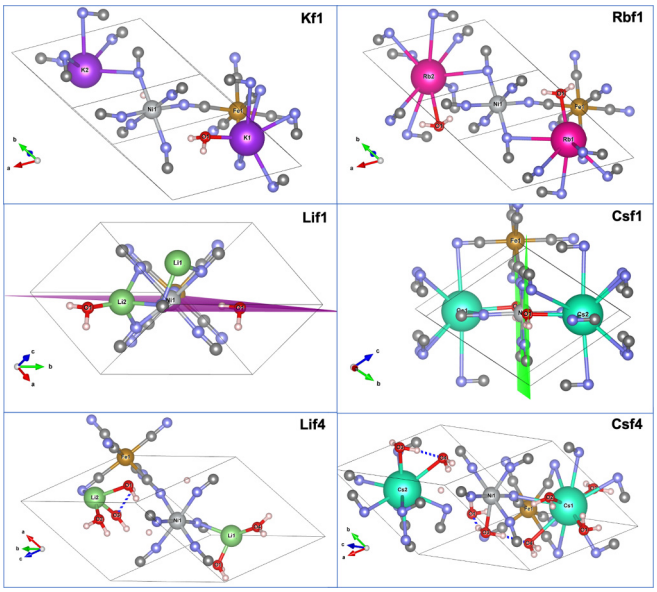

FIG. 8. Lowest-energy structures for Kf1, Rbf1, Lif1, Lif4, Csf1, and Csf4 configurations.

configurations, whose structures are shown in Fig. 8, also show similar cation-water and cation-nitrogen coordination among both cation types with similar $\langle\angle\text{NiNC}\rangle$ values of 169.7° for Rbf1 and 170.3° for Kf1. Therefore, we attribute the abnormal value of IMP_{NiNC} for the Rbf1 cell to the insufficient data of Rbf1 configurations for ML model training (690 trajectory images) compared to other configuration types (approximately 1300 trajectory images), rather than to K- and Rb-intercalated configurations experiencing patently different interactions.

The cation-dependent IMP_{NiNC} trends for light, intermediate, and heavy cations are consistent with the results based on the values of $\text{OP}_{A,\text{CN}}$ [Fig. 7(e)]. At $n = 0$, $\text{OP}_{A,\text{CN}}$ follows a sequence of $\text{Li}^+ \approx \text{Na}^+ < \text{K}^+ \approx \text{Rb}^+ < \text{Cs}^+$, identical to the results in half intercalated lattices. From $n = 0$ to $n = 1$, $\text{OP}_{A,\text{CN}}$ is weakened only for the Cs-intercalated lattice. When either Li^+ or Na^+ are fully intercalated in a framework, cation-ligand interactions are screened by zeolitic water molecules, as evidenced by decreases in $\text{OP}_{\text{Li,CN}}$ and $\text{OP}_{\text{Na,CN}}$ with increasing n [Fig. 7(e)]. When heavier cations are intercalated, large hydration degrees promote cation-ligand interactions. Further, our analysis shows that the Ni-N bond is stretched at $n = 4$ irrespective of which cation is intercalated [Fig. 7(f)]. In particular, we find that Ni-N bonds are broken in the lowest-energy configuration of Csf4 cases, resulting in the Ni atom coordinating only to five nitrogen atoms and to one oxygen atom (i.e., "O3" in Fig. 8). The results in Figs. 7(e) and 7(f) have important implications regarding cation intercalation processes in PBAs. On the one hand, the PBA framework, like other metal-organic frameworks [107], is likely to experience bond breakage at high degrees of hydration because intercalated cations force zeolitic water to attack weak bonds near metal centers, such as Ni-N bonds here. On the other hand, to accommodate the large size of heavy ions, such cations are likely to expel water molecules from the PBA lattice during cation intercalation, as a result of the energetic cost associated with extreme confinement of both the interstitial cation and H_2O. This conclusion is consistent with recent neutron powder diffraction and extended x-ray absorption fine-structure spectroscopy experiments that have revealed that K^+ intercalation induces the expulsion of zeolitic water molecules from the CuFe-PBA lattice [15].

By following the same method elaborated in the previous subsection, we find that atomic arrangements of *AOA* clusters are also important for predicting energies of fully intercalated configurations [Fig. 7(b)]. We also evaluate the differences of $\text{IMP}_{\text{NiNC}+AOA} - \text{IMP}_{\text{NiNC}}$ to investigate correlations between cation-water coordination and octahedral tilting [Fig. 7(c)]. We observe an increasing trend of $\text{IMP}_{\text{NiNC+LiOLi}} - \text{IMP}_{\text{NiNC}}$ with increasing n, while the value of $\text{IMP}_{\text{NiNC+CsOCs}} - \text{IMP}_{\text{NiNC}}$ decreases monotonically with increasing hydration degree [Fig. 7(c)]. Similar to our results for half intercalated configurations, the increasing trend of $\text{IMP}_{\text{NiNC+LiOLi}} - \text{IMP}_{\text{NiNC}}$ corresponds to the enhanced correlation between NiNC clusters and Li-water coordination due to the growing size of Li-water clusters with increasing hydration degree. In the lowest-energy configurations of fully Li-intercalated cells, we further find that large Li-water clusters exacerbate octahedral tilting, as $\langle\angle\text{NiNC}\rangle$ reduces from 177° for Lif1 to 167° for Lif4 (Fig. 8). Meanwhile, the arrangement of CsOCs clusters decouples from the NiNC arrangement at large n [Fig. 7(c)] because Cs^+ ions, which persistently occupy body-center sites, confine water molecules in such a way that they attack Ni-N bonds. However, we observe no clear trends of $\text{IMP}_{\text{NiNC}+AOA} - \text{IMP}_{\text{NiNC}}$ with n for $A = \text{Na}^+$, K^+, and Rb^+, indicating that structural correlations in these systems involve more clusters than NiNC and *AOA* clusters.

Because the arrangement of *AOA* clusters is essential to understanding energetics in most of the present cases [Fig. 7(b)], we define the energy landscape subspace for fully intercalated configurations using $\langle d_{AO} d_{OA'}[1 - \cos(\angle AOA')]\rangle$ and $\langle\cos(\angle\text{NiNC})\rangle$ as coordinates. We use the same conditions as for half intercalated configurations to classify steep (red) and shallow (black) downhill traces. For the landscapes of configurations intercalated with light cations (red rectangle: Li- and Na-intercalated configurations), most steep traces are nearly perpendicular to the horizontal axis, implying that total energy relaxes by lattice distortion via Ni-octahedron tilting while maintaining cation-water coordination. Unlike the results for half-intercalated configurations (see Fig. 6 and discussion thereof), the downhill traces of fully intercalated configurations are accompanied by increases in the tilting angle of metal octahedra (i.e., decreases in the angle of $\angle$NiNC), indicating that the addition of a light cation destabilizes the pristine NiFe-PBA lattice to a certain degree.

The landscapes of K- and Rb-intercalated configurations contain downhill traces similar to those of Li- and Na-intercalated cells, except in the case of Kf1 (see Fig. 9). The landscapes of all K-intercalated configurations show downhill traces that are concentrated around several distinct local minima, indicating that atomic movements are confined in interstitial space. For Rb-intercalated cells, however, extended downhill traces with $|\Delta y| \sim 0.01$ are found even when the hydration degree is large ($n \geqslant 3$). Furthermore, most steep traces for Rb-intercalated configurations show an increase in the angle of $\angle$NiNC, implying that the addition of Rb^+

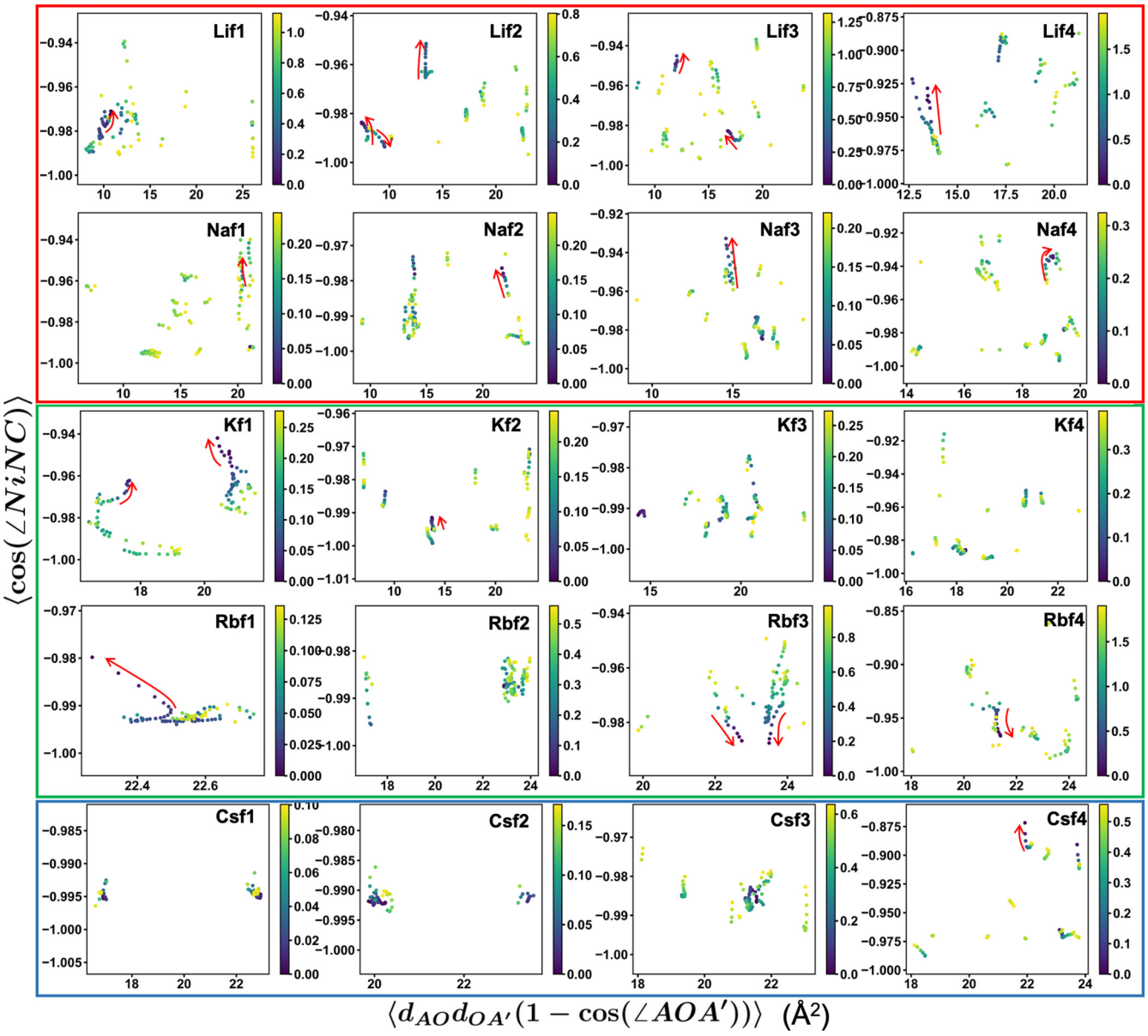

FIG. 9. 2D projections of DFT-predicted energy landscapes for fully intercalated configurations near local minima. Color bars indicate DFT energy in eV per formula unit, relative to that of the lowest-energy structure for a given configuration type.

stabilizes the PBA lattice by decreasing octahedral tilting. Because Cs^+ is especially large and hydrophobic, the fully Cs-intercalated lattice has limited interstitial space for cations and zeolitic H_2O molecules to move collectively. As a result, the corresponding landscapes of Cs^+ (blue rectangle) are concentrated around distinct local minima.

Thus far, our analysis of correlation between energy and the arrangement of polyatomic clusters has focused on atoms other than hydrogen atoms. However, we have observed that hydrogen bonding determines the morphology of cation-water clusters in both half and fully intercalated configurations (Figs. 5 and 8), as a result of which the nature of framework-interstitial interactions is affected indirectly. In the subsequent subsection we explore these effects in more detail by analyzing the orientation of H_2O molecules in the interstitial space of cation-intercalated NiFe-PBAs.

B. Correlating DFT energy with four-particle features

We now use certain four-particle feature parameters based on the solid angles subtended by zeolitic water molecules that extend either from Ni^{2+} (Ω_{Ni-H_2O}) or A^+ (Ω_{A-H_2O}) cations to characterize the effect of H_2O's orientation on DFT energy. We identify downhill traces in the energy landscapes of Li- and Cs-intercalated configurations with $n \leqslant 2$ using $\langle \tan(\Omega_{A-H_2O}/2) \rangle$ and $\langle \tan(\Omega_{Ni-H_2O}/2) \rangle$ to define an associated subspace. In the landscapes shown in Fig. 10, shallow traces (black) have secant slopes with $|\Delta y/\Delta x| \leqslant 0.01$ that are rendered only if $|\Delta x| \geqslant 0.1$.

The definition of the solid angle [Eq. (5)] shows that its tangent for a *G-XYZ* cluster can only be zero if $\overrightarrow{GX} \perp (\overrightarrow{GY} \times \overrightarrow{GZ})$ is satisfied, in which case *G*, *X*, *Y*, and *Z* must be coplanar. When water content is low either in small or in large cation-intercalated configurations ($n = 1$), shallow traces

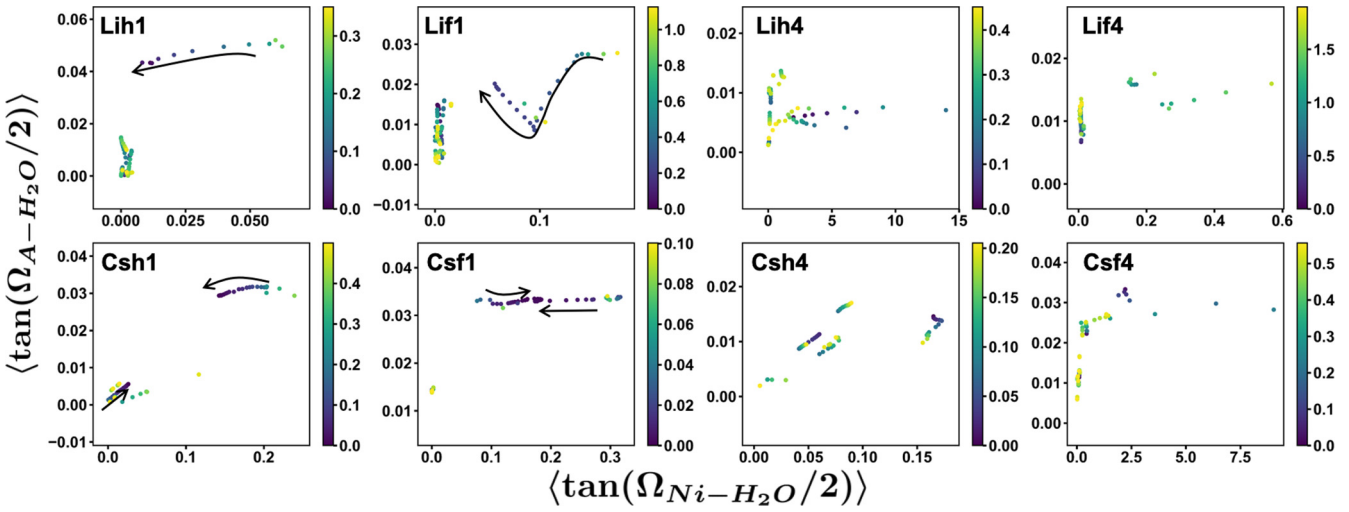

FIG. 10. 2D landscape projections spanned by solid angles for H_2O with respect to intercalated cation A^+ (y axes) and to lattice cation Ni^{2+} (x axes). Such landscapes are shown for eight different Li- and Cs-intercalated configurations.

indicate dramatic changes in H_2O's orientation relative to an observer at each Ni^{2+} center. However, along such a trace H_2O's orientation does not change significantly relative to an observer at each A^+ center. When water content is high however ($n = 4$), the associated energy landscapes are concentrated around certain local minima because confined interstitial space constrains the librational movement of each water molecule. The local minima that appear in the landscapes of Cs-intercalated configurations are found with $|\langle \tan(\Omega_{Cs-H_2O}/2)\rangle| \sim 0.03$, whereas $|\langle \tan(\Omega_{Li-H_2O}/2)\rangle|$ values at local minima vary from 0.4 for Lih1 to 0.005 for Lif4 for Li-intercalated configurations. Differences in the values of $|\langle \tan(\Omega_{A-H_2O}/2)\rangle|$ are evident from the structures of Lif1, Lif4, Csf1, and Csf4 (Fig. 8). Cs-O coordination in the lowest-energy Csf1 configuration forms a 90° angle with the lattice plane on which the corresponding water molecule and Ni atom reside, indicated by the purple plane for Csf1 in Fig. 8. In the Lif1 configuration, however, such coordination is frustrated. Repeating the unit cell of Csf1 in 3D space, we recognize that an extended cation-water network exists in its interstitial space, where a water molecule bridges between each pair of adjacent Cs^+ ions. In contrast, due to the small size and the hydrophilicity of Li^+, the formation of a hydration shell around Li^+ prohibits such ordering of cation-water clusters. Further, when hydration degree is large ($n \geqslant 3$), cation-water ordering is frustrated for all configuration types.

IV. CONCLUSIONS

The present investigation reveals that the interactions between PBA lattices and interstitial species significantly affect lattice stability and the hydration degree of cations in interstitial space. To elucidate the associated correlations between polyatomic arrangements and atomic interactions in PBA frameworks, we performed DFT calculations of hydrated NiFe-PBA lattices with various oxidation states, hydration degrees, and types of intercalated cations, using Voronoi tessellation to sample the configurational space of H_2O occupation. We used a simple descriptor to extract many-particle features from polyatomic arrangements in DFT-calculated configurations. These features were then incorporated into different data sets to train ML models, resulting in different model accuracy levels. Based on the accuracy of ML models using different features, we ranked the sensitivity of DFT-calculated energy to various many-particle features. The correlation between two features is identified when including both features in training data improves ML accuracy to a greater extent than including only one of the two features. Because the many-particle features are generated from their associated polyatomic arrangements, statistical correlations in the feature space bijectively map to polyatomic correlations in configurational space.

Using this technique we found a significant correlation between NiNC and Ni*A*O clusters for most of the half intercalated lattices. The variation of these ML model correlations with hydration degree indicates that correlation between NiNC and Ni*A*O arrangements depends on the type of intercalated cation. Inspired by these results from ML model training, we analyzed the overlap population between intercalated cations and their surrounding atoms in the lowest-energy configurations to determine which atomic interactions were responsible for such cation-dependent correlation. This overlap population analysis showed that the strong dative bonds between intercalated cations and water molecules are accompanied by weaker dative bonding between cations and cyanide ligands, and vice versa. This analysis therefore demonstrates that NiNC-Ni*A*O correlation results from the competition between two kinds of dative bonds around intercalated cations. By studying the arrangement of interstitial species within lowest-energy structures in conjunction with overlap population analysis, we deduced that the bias toward dative bonding of cations with either H_2O or cyanide ligands is dependent on the bare ionic radius and hydrophilicity of intercalated cations. We thus group Na^+, Li^+, K^+, Rb^+, and Cs^+ ions into three categories. When light, hydrophilic cations (Li^+ and Na^+) intercalate into PBA lattices, zeolitic water molecules firmly coordinate to cations and form a hydration shell, albeit with a hydration shell structure that is confined relative to that

of cations in bulk liquid water. For K^+ and Rb^+ ions that are hydrophobic with intermediate ionic size, zeolitic water molecules push them off of body-center sites, such that those ions coordinate to cyanide ligands to an increasing degree. If intercalated cations are large and hydrophobic, such as Cs^+, they displace zeolitic water molecules into adjacent vacant sites, where water clusters form an amorphous phase that is more dense than bulk liquid water.

We used a similar method to analyze polyatomic correlations in fully intercalated lattices and found that the extent of Ni-octahedron tilting is cation specific, as analyzed using feature parameters for NiNC clusters. The variation of $\mathrm{IMP}_{\mathrm{NiNC}+AOA} - \mathrm{IMP}_{\mathrm{NiNC}}$ with increasing degree of hydration, where IMP is an importance index associated with certain three-body features, further suggests that zeolitic water promotes correlation between NiNC and *AOA* clusters in Li- and Na-intercalated lattices. In contrast, such correlation disappears at high hydration degrees in K-, Rb-, and Cs-intercalated lattices. A corresponding sharp decrease in the overlap population between nickel and cyanide ligands at the highest hydration degree occurs for all kinds of lattices. This result suggests that Ni-N bonds stretch due to confinement of interstitial space. We also showed that the dative bonding between zeolitic water molecules and cyanide ligands is strengthened at high hydration degree in the lowest-energy structures. At the highest hydration degree, we even discovered the breakage of such bonds in Cs-intercalated lattices. We thus conclude that the NiFe-PBA lattice is likely to expel zeolitic water molecules during the intercalation of large and hydrophobic cations.

To explore features of DFT-calculated energy landscapes, we projected these landscapes onto subspaces spanned using two different three-particle feature parameters found using ML analysis to have highest correlation to energy. The downhill traces in such landscapes reveal the correlations between collective movements in interstitial space in the vicinity of local minima in energy. When PBA lattices are half intercalated with Li^+ or Na^+ ions, such downhill traces indicate two patterns of collective movement: (1) the tilting of Ni octahedra accompanied by minimal rearrangement of cation-H_2O clusters, and (2) significant rearrangement of cation-H_2O clusters without distorting the PBA lattice. Pattern (1) prevails for fully intercalated lattices, however. For K- and Rb-intercalated lattices, downhill traces show that both patterns still coexist on the landscapes of half intercalated lattices. In lattices that are intercalated fully with K^+ or Rb^+ ions, the tilting of metal octahedra is constrained and even reversed. Intercalation of Cs^+ ions also suppresses lattice distortion in both half and fully intercalated lattices. For Cs^+ intercalated configurations the collective movements of water clusters are evidenced by landscapes projected using four-particle feature parameters. Downhill traces on those landscapes also suggest a coplanar arrangement of water molecules with Ni-metal centers at low hydration degrees. The zeolitic water molecules in Cs-intercalated lattices bridge between adjacent Cs^+ ions to form a Cs-H_2O network in interstitial space. Such a network is frustrated in Li-intercalated lattices because hydration shells separate adjacent Li^+ cations.

While the present investigation focuses on PBAs, we note that the ML model and analytical method used in this study can be applied to other crystalline solids. The descriptor proposed here effectively maps atomic configurations to feature space spanned by both discrete and continuous features. Therefore, the present method does not require normalization or differentiable features for ML model training and is suitable for fast prototyping of accurate ML models. Despite these benefits, we note that the training data used here were obtained not only from terminal relaxed structures, but also from the structures obtained along relaxation trajectories. Here, leaf regions are assigned to the feature values of configurations near local minima only at the deepest levels of each regression tree. Along those lines we envision using the present ML model as an estimator of DFT-calculated energy in the search for global optima by introducing bias towards near-minimal structures in training data sets.

ACKNOWLEDGMENTS

This research was supported by the U.S. National Science Foundation (Award No. 1931659) and the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign. Access to the Blue Waters supercomputer was provided through a Director's Award from the National Center for Supercomputing Applications (NCSA). We also wish to thank NCSA staff for their assistance in utilizing the Blue Waters supercomputer to conduct a portion of the DFT calculations presented here.

[1] W. Li, C. Han, G. Cheng, S. Chou, H. Liu, and S. Dou, Chemical properties, structural properties, and energy storage applications of Prussian blue analogues, Small **15**, 1900470 (2019).

[2] S. Klink, Y. Ishige, and W. Schuhmann, Prussian blue analogues: A versatile framework for solid-contact ion-selective electrodes with tunable potentials, ChemElectroChem **4**, 490 (2017).

[3] J. Lee, S. Kim, and J. Yoon, Rocking chair desalination battery based on Prussian blue electrodes, ACS Omega **2**, 1653 (2017).

[4] S. Porada, A. Shrivastava, P. Bukowska, P. M. Biesheuvel, and K. C. Smith, Nickel hexacyanoferrate electrodes for continuous cation intercalation desalination of brackish water, Electrochim. Acta **255**, 369 (2017).

[5] T. Kim, C. A. Gorski, and B. E. Logan, Low energy desalination using battery electrode deionization, Environ. Sci. Technol. Lett. **4**, 444 (2017).

[6] E. R. Reale, A. Shrivastava, and K. C. Smith, Effect of conductive additives on the transport properties of porous flow-through electrodes with insulative particles and their optimization for Faradaic deionization, Water Res. **165**, 114995 (2019).

[7] Z. Wang, Y. Huang, R. Luo, F. Wu, L. Li, M. Xie, J. Huang, and R. Chen, Ion-exchange synthesis of high-energy-density Prussian blue analogues for sodium ion battery cathodes with

fast kinetics and long durability, J. Power Sources **436**, 226868 (2019).

[8] J. Yin, Y. Shen, C. Li, C. Fan, S. Sun, Y. Liu, J. Peng, L. Qing, and J. Han, In situ self-assembly of core–shell multimetal Prussian blue analogues for high-performance sodium-ion batteries, Chem. Sus. Chem. **12**, 4786 (2019).

[9] W. Ren, M. Qin, Z. Zhu, M. Yan, Q. Li, L. Zhang, D. Liu, and L. Mai, Activation of sodium storage sites in Prussian blue analogues via surface etching, Nano Lett. **17**, 4713 (2017).

[10] A. Simonov, T. De Baerdemaeker, H. L. B. Boström, M. L. Ríos Gómez, H. J. Gray, D. Chernyshov, A. Bosak, H. B. Bürgi, and A. L. Goodwin, Hidden diversity of vacancy networks in Prussian blue analogues, Nature (London) **578**, 256 (2020).

[11] S. Liu and K. C. Smith, Intercalated cation disorder in Prussian blue analogues: First-principles and grand canonical analyses, J. Phys. Chem. C **123**, 10191 (2019).

[12] A. Shrivastava, S. Liu, and K. C. Smith, Linking capacity loss and retention of nickel hexacyanoferrate to a two-site intercalation mechanism for aqueous Mg^{2+} and Ca^{2+} ions, Phys. Chem. Chem. Phys. **21**, 20177 (2019).

[13] Z.-Y. Yu, Y. Duan, J.-D. Liu, Y. Chen, X.-K. Liu, W. Liu, T. Ma, Y. Li, X.-S. Zheng, T. Yao, M.-R. Gao, J.-F. Zhu, B.-J. Ye, and S.-H. Yu, Unconventional CN vacancies suppress iron-leaching in Prussian blue analogue pre-catalyst for boosted oxygen evolution catalysis, Nat. Commun. **10**, 2799 (2019).

[14] J. Mink, A. Stirling, D. O. Ojwang, G. Svensson, J. Mihály, C. Németh, M. Drees, and L. Hajba, Vibrational properties and bonding analysis of copper hexacyanoferrate complexes in solid state, Appl. Spectrosc. Rev. **54**, 369 (2019).

[15] D. Wardecki, D. O. Ojwang, J. Grins, and G. Svensson, Neutron diffraction and EXAFS studies of K2x/3Cu[Fe(CN)6]2/3·n H_2O, Cryst. Growth Des. **17**, 1285 (2017).

[16] J. Peng, J. Wang, H. Yi, W. Hu, Y. Yu, J. Yin, Y. Shen, Y. Liu, J. Luo, Y. Xu, P. Wei, Y. Li, Y. Jin, Y. Ding, L. Miao, J. Jiang, J. Han, and Y. Huang, A dual-insertion type sodium-ion full cell based on high-quality ternary-metal Prussian blue analogs, Adv. Energy Mater. **8**, 1702856, (2018).

[17] B. Xie, P. Zuo, L. Wang, J. Wang, H. Huo, M. He, J. Shu, H. Li, S. Lou, and G. Yin, Achieving long-life Prussian blue analogue cathode for Na-ion batteries via triple-cation lattice substitution and coordinated water capture, Nano Energy **61**, 201 (2019).

[18] B. Xie, L. Wang, J. Shu, X. Zhou, Z. Yu, H. Huo, Y. Ma, X. Cheng, G. Yin, and P. Zuo, Understanding the structural evolution and lattice water movement for rhombohedral nickel hexacyanoferrate upon sodium migration, ACS Appl. Mater. Interfaces **11**, 46705 (2019).

[19] M. Oliver-Tolentino, G. Ramos-Sánchez, G. Guzmán, M. Avila, I. González, and E. Reguera, Water effect on sodium mobility in zinc hexacyanoferrate during charge/discharge processes in sodium ion-based battery, Solid State Ion. **312**, 67 (2017).

[20] S. Wheeler, I. Capone, S. Day, C. Tang, and M. Pasta, Low-potential Prussian blue analogues for sodium-ion batteries: Manganese hexacyanochromate, Chem. Mater. **31**, 2619 (2019).

[21] C. D. Wessells, S. V. Peddada, M. T. McDowell, R. A. Huggins, and Y. Cui, The effect of insertion species on nanostructured open framework hexacyanoferrate battery electrodes, J. Electrochem. Soc. **159**, A98 (2012).

[22] T. Ikeshoji, Separation of alkali metal ions by intercalation into a Prussian blue electrode, J. Electrochem. Soc. **133**, 2108 (1986).

[23] S. D. Rassat, J. H. Sukamto, R. J. Orth, M. A. Lilga, and R. T. Hallen, Development of an electrically switched ion exchange process for selective ion separations, Sep. Purif. Technol. **15**, 207 (1999).

[24] B. Sun, X. G. Hao, Z. De Wang, G. Q. Guan, Z. L. Zhang, Y. Bin Li, and S. Bin Liu, Separation of low concentration of cesium ion from wastewater by electrochemically switched ion exchange method: Experimental adsorption kinetics analysis, J. Hazard. Mater. **233–234**, 177 (2012).

[25] T. Kim, C. A. Gorski, and B. E. Logan, Ammonium removal from domestic wastewater using selective battery electrodes, Environ. Sci. Technol. Lett. **5**, 578 (2018).

[26] J. Song, L. Wang, Y. Lu, J. Liu, B. Guo, P. Xiao, J.-J. Lee, X.-Q. Yang, G. Henkelman, and J. B. Goodenough, Removal of interstitial H_2O in hexacyanometallates for a superior cathode of a sodium-ion battery, J. Am. Chem. Soc. **137**, 2658 (2015).

[27] H. Hu, W. Liu, M. Zhu, Y. Lin, Y. Liu, J. Zhang, T. Chen, and K. Liu, Yolk-shell Prussian blue nanoparticles with fast ion diffusion for sodium-ion battery, Mater. Lett. **249**, 206 (2019).

[28] Y. Xu, J. Wan, L. Huang, M. Ou, C. Fan, P. Wei, J. Peng, Y. Liu, Y. Qiu, X. Sun, C. Fang, Q. Li, J. Han, Y. Huang, J. A. Alonso, and Y. Zhao, Structure distortion induced monoclinic nickel hexacyanoferrate as high-performance cathode for na-ion batteries, Adv. Energy Mater. **9**, 1803158 (2019).

[29] X. Guo, Z. Wang, Z. Deng, X. Li, B. Wang, X. Chen, and S. P. Ong, Water contributes to higher energy density and cycling stability of Prussian blue analogue cathodes for aqueous sodium-ion batteries, Chem. Mater. **31**, 5933 (2019).

[30] L. Chen, H. Shao, X. Zhou, G. Liu, J. Jiang, and Z. Liu, Water-mediated cation intercalation of open-framework indium hexacyanoferrate with high voltage and fast kinetics, Nat. Commun. **7**, 11982 (2016).

[31] C.-C. Xie, D.-H. Yang, M. Zhong, and Y.-H. Zhang, Improving the performance of a ternary Prussian blue analogue as cathode of lithium battery via annealing treatment, Z. Anorg. Allg. Chem. **642**, 289 (2016).

[32] P. Canepa, G. Sai Gautam, D. C. Hannah, R. Malik, M. Liu, K. G. Gallagher, K. A. Persson, and G. Ceder, Odyssey of multivalent cathode materials: Open questions and future challenges, Chem. Rev. **117**, 4287 (2017).

[33] Z. Li, K. Xiang, W. Xing, W. C. Carter, and Y.-M. Chiang, Reversible aluminum-ion intercalation in Prussian blue analogs and demonstration of a high-power aluminum-ion asymmetric capacitor, Adv. Energy Mater. **5**, 1401410 (2015).

[34] T. Tojo, Y. Sugiura, R. Inada, and Y. Sakurai, Reversible calcium ion batteries using a dehydrated Prussian blue analogue cathode, Electrochim. Acta **207**, 22 (2016).

[35] N. Kuperman, P. Padigi, G. Goncher, D. Evans, J. Thiebes, and R. Solanki, High performance Prussian Blue cathode for non-aqueous Ca-ion intercalation battery, J. Power Sources **342**, 414 (2017).

[36] Y. Mizuno, M. Okubo, E. Hosono, T. Kudo, H. Zhou, and K. Oh-Ishi, Suppressed activation energy for interfacial charge transfer of a Prussian blue analog thin film electrode with

hydrated ions (Li^+, Na^+, and Mg^{2+}), J. Phys. Chem. C **117**, 10877 (2013).

[37] Q. Gao, J. Chen, Q. Sun, D. Chang, Q. Huang, H. Wu, A. Sanson, R. Milazzo, H. Zhu, Q. Li, Z. Liu, J. Deng, and X. Xing, Switching between giant positive and negative thermal expansions of a YFe(CN)6-based Prussian blue analogue induced by guest species, Angew. Chemie Int. Ed. **56**, 9023 (2017).

[38] M. Itoi, I. Maurin, F. Varret, F. A. Frye, D. R. Talham, D. Chernyshov, and K. Boukheddaden, When local deformations trigger lattice instability: Flow diagram investigations for photoinduced and quenched metastable states in a Prussian blue analog, Phys. Rev. B **88**, 094104 (2013).

[39] L. Fumagalli, A. Esfandiar, R. Fabregas, S. Hu, P. Ares, A. Janardanan, Q. Yang, B. Radha, T. Taniguchi, K. Watanabe, G. Gomila, K. S. Novoselov, and A. K. Geim, Anomalously low dielectric constant of confined water, Science **360**, 1339 (2018).

[40] N. Shpigel, M. D. Levi, S. Sigalov, T. S. Mathis, Y. Gogotsi, and D. Aurbach, Direct assessment of nanoconfined water in 2D Ti3C2 electrode interspaces by a surface acoustic technique, J. Am. Chem. Soc. **140**, 8910 (2018).

[41] S. Bruckenstein, A. T. Fensore, and A. R. Hillman, Time-resolved mono-anion, di-anion, and solvent transfers into a Poly(vinylferrocene)-modified electrode, J. Electrochem. Soc. **145**, L24 (1998).

[42] M. Balasubramanian, M. T. Giacomini, H. S. Lee, J. McBreen, and J. H. Sukamto, X-Ray absorption studies of Poly(vinylferrocene) polymers for anion separation, J. Electrochem. Soc. **149**, D137 (2002).

[43] J. H. Her, P. W. Stephens, C. M. Kareis, J. G. Moore, K. S. Min, J. W. Park, G. Bali, B. S. Kennon, and J. S. Miller, Anomalous non-Prussian blue structures and magnetic ordering of K 2MnII[MnII(CN)6] and Rb 2MnII[MnII(CN)6], Inorg. Chem. **49**, 1524 (2010).

[44] M. Okubo, D. Asakura, Y. Mizuno, J. D. Kim, T. Mizokawa, T. Kudo, and I. Honma, Switching redox-active sites by valence tautomerism in Prussian blue analogues AxMny[Fe(CN)6]· n H 2O (A: K, Rb): Robust frameworks for reversible Li storage, J. Phys. Chem. Lett. **1**, 2063 (2010).

[45] E. P. Alsaç, E. Ülker, S. V. K. Nune, Y. Dede, and F. Karadas, Tuning the electronic properties of Prussian blue analogues for efficient water oxidation electrocatalysis: Experimental and computational studies, Chem. - Eur. J. **24**, 4856 (2018).

[46] J. Wu, J. Song, K. Dai, Z. Zhuo, L. A. Wray, G. Liu, Z. Shen, R. Zeng, Y. Lu, and W. Yang, Modification of transition-metal redox by interstitial water in hexacyanometalate electrodes for sodium-ion batteries, J. Am. Chem. Soc. **139**, 18358 (2017).

[47] J.-C. Tan, P. Jain, and A. K. Cheetham, Influence of ligand field stabilization energy on the elastic properties of multiferroic MOFs with the perovskite architecture, Dalton Trans. **41**, 3949 (2012).

[48] I. Bhowmick, A. J. Roehl, J. R. Neilson, A. K. Rappé, and M. P. Shores, Slow magnetic relaxation in octahedral low-spin Ni(III) complexes, Chem. Sci. **9**, 6564 (2018).

[49] A. M. Hamisu, A. Ariffin, and A. C. Wibowo, Cation exchange in metal-organic frameworks (MOFs): The hard-soft acid-base (HSAB) principle appraisal, Inorg. Chim. Acta **511**, 119801 (2020).

[50] J. L. Reed, Hard and soft acids and bases: Structure and process, J. Phys. Chem. A **116**, 7147 (2012).

[51] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, Interaction energies of large clusters from many-body expansion, J. Chem. Phys. **135**, 224102 (2011).

[52] R. M. Richard and J. M. Herbert, A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory, J. Chem. Phys. **137**, 064113 (2012).

[53] G. C. Lynch, R. Steckler, D. W. Schwenke, A. J. C. Varandas, D. G. Truhlar, and B. C. Garrett, Use of scaled external correlation, a double many-body expansion, and variational transition state theory to calibrate a potential energy surface for FH2, J. Chem. Phys. **94**, 7136 (1991).

[54] M. A. Boles and D. V. Talapin, Many-body effects in nanocrystal superlattices: Departure from sphere packing explains stability of binary phases, J. Am. Chem. Soc. **137**, 4494 (2015).

[55] P. G. Boyd, S. M. Moosavi, M. Witman, and B. Smit, Force-field prediction of materials properties in metal-organic frameworks, J. Phys. Chem. Lett. **8**, 357 (2017).

[56] A. Tkatchenko, D. Alfè, and K. S. Kim, First-principles modeling of non-covalent interactions in supramolecular systems: The role of many-body effects, J. Chem. Theory Comput. **8**, 4317 (2012).

[57] T. A. Manz and N. G. Limas, Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology, RSC Adv. **6**, 47771 (2016).

[58] G. Henkelman, A. Arnaldsson, and H. Jónsson, A fast and robust algorithm for Bader decomposition of charge density, Comput. Mater. Sci. **36**, 354 (2006).

[59] T. A. Manz and D. S. Sholl, Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials, J. Chem. Theory Comput. **8**, 2844 (2012).

[60] T. A. Manz, Introducing DDEC6 atomic population analysis: Part 3. Comprehensive method to compute bond orders, RSC Adv. **7**, 45552 (2017).

[61] J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, Trouble with the many-body expansion, J. Chem. Theory Comput. **10**, 3699 (2014).

[62] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by B. Krishnapuram and M. Shah (Association for Computing Machinery, New York, 2016), p. 785–794.

[63] P. Nie, J. Yuan, J. Wang, Z. Le, G. Xu, L. Hao, G. Pang, Y. Wu, H. Dou, X. Yan, and X. Zhang, Prussian blue analogue with fast kinetics through electronic coupling for sodium ion batteries, ACS Appl. Mater. Interfaces **9**, 20306 (2017).

[64] Y. Gao, H. Yang, X. Wang, Y. Bai, N. Zhu, S. Guo, L. Suo, H. Li, H. Xu, and C. Wu, The compensation effect mechanism of Fe-Ni mixed Prussian blue analogues in aqueous rechargeable aluminum-ion batteries, Chem. Sus. Chem. **13**, 732 (2020).

[65] F. Tran and J. Hutter, Nonlocal van der waals functionals: The case of rare-gas dimers and solids, J. Chem. Phys. **138**, 204103 (2013).

[66] O. A. Vydrov and T. Van Voorhis, Nonlocal van der waals density functional: The simpler the better, J. Chem. Phys. **133**, 244103 (2010).

[67] J. Kong, Z. Gan, E. Proynov, M. Freindorf, and T. R. Furlani, Efficient computation of the dispersion interaction with density-functional theory, Phys. Rev. A **79**, 042510 (2009).

[68] T. Suzuki, R. Tamura, and T. Miyazaki, Machine learning for atomic forces in a crystalline solid: Transferability to various temperatures, Int. J. Quantum Chem. **117**, 33 (2017).

[69] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields, Nat. Commun. **9**, 3887 (2018).

[70] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. R. Nelson, T. Back, D. Hassabis, and P. Kohli, Unveiling the predictive power of static structure in glassy systems, Nat. Phys. **16**, 448 (2020).

[71] R. Kondor, N-body networks: A covariant hierarchical neural network architecture for learning atomic potentials, arXiv:1803.01588.

[72] H. Ji and Y. Jung, Artificial neural network for the configuration problem in solids, J. Chem. Phys. **146**, 064103 (2017).

[73] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Commun. **247**, 106949 (2020).

[74] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande, MoleculeNet: a benchmark for molecular machine learning, Chem. Sci. **9**, 513 (2018).

[75] D. Zhang, L. Lu, L. Guo, and G. E. Karniadakis, Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems, J. Comput. Phys. **397**, 108850 (2019).

[76] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, Quantum ESPRESSO toward the exascale, J. Chem. Phys. **152**, 154105 (2020).

[77] C. D. Wessells, S. V. Peddada, R. A. Huggins, and Y. Cui, Nickel hexacyanoferrate nanoparticle electrodes for aqueous sodium and potassium ion batteries, Nano Lett. **11**, 5421 (2011).

[78] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, Pseudopotentials for high-throughput DFT calculations, Comput. Mater. Sci. **81**, 446 (2014).

[79] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table, Comput. Phys. Commun. **226**, 39 (2018).

[80] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. **132**, 241722 (2010).

[81] A. Patra, S. Jana, and P. Samal, Performance of tao-mo semilocal functional with rvv10 dispersion-correction: Influence of different correlation, J. Phys. Chem. A **123**, 10582 (2019).

[82] S. Chong, Y. Wu, S. Guo, Y. Liu, and G. Cao, Potassium nickel hexacyanoferrate as cathode for high voltage and ultralong life potassium-ion batteries, Energy Storage Mater. **22**, 120 (2019).

[83] Y. Yuan, D. Bin, X. Dong, Y. Wang, C. Wang, and Y. Xia, Intercalation pseudocapacitive nanoscale nickel hexacyanoferrate@carbon nanotubes as a high-rate cathode material for aqueous sodium-ion battery, ACS Sustain. Chem. Eng. **8**, 3655 (2020).

[84] K. P. Murphy, *Machine Learning A Probabilistic Perspective* (MIT Press, Cambridge, MA, 2012).

[85] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, Scikit-learn, GetMobile Mob. Comput. Commun. **19**, 29 (2015).

[86] T. Zhang, A general greedy approximation algorithm with applications, *Advances in Neural Information Processing Systems*, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT press, Vancouver, 2002), pp. 1065–1072.

[87] X. Zhang, T. Deng, and G. Jia, Nuclear spin-spin coupling constants prediction based on XGBoost and LightGBM algorithms, Mol. Phys. **118**, e1696478 (2019).

[88] B. Yu, W. Qiu, C. Chen, A. Ma, J. Jiang, H. Zhou, and Q. Ma, SubMito-XGBoost: Predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting, Bioinformatics **36**, 1074 (2020).

[89] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett. **108**, 058301 (2012).

[90] V. L. Deringer and G. Csányi, Machine learning based interatomic potential for amorphous carbon, Phys. Rev. B **95**094203 (2017).

[91] D. Himmel, I. Krossing, and A. Schnepf, Dative or not dative? Angew. Chem. Int. Ed. **53**, 6047 (2014).

[92] C. Lepetit, V. Maraval, Y. Canac, and R. Chauvin, On the nature of the dative bond: Coordination to metals and beyond. The carbon case, Coord. Chem. Rev. **308**, 59 (2016).

[93] W. Qiao, T. Maximova, X. Fang, E. Plaku, and A. Shehu, Reconstructing and mining protein energy landscape to understand disease, in *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine*, edited by C.-R. Shyu and X. T. Hu (IEEE, Kansas, 2017), p. 22–27.

[94] S. Honrao, B. E. Anthonio, R. Ramanathan, J. J. Gabriel, and R. G. Hennig, Machine learning of ab-initio energy landscapes for crystal structure predictions, Comput. Mater. Sci. **158**, 414 (2019).

[95] G. Hegde and R. C. Bowen, Machine-learned approximations to density functional theory hamiltonians, Sci. Rep. **7**, 42669 (2017).

[96] S. Zeng, Y. Zhao, G. Li, R. Wang, X. Wang, and J. Ni, Atom table convolutional neural networks for an accurate prediction of compounds properties, npj Comput. Mater. **5**, 84 (2019).

[97] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, Less is more: Sampling chemical space with active learning, J. Chem. Phys. **148**, 241733 (2018).

[98] D. Yoo, K. Lee, W. Jeong, D. Lee, S. Watanabe, and S. Han, Atomic energy mapping of neural network potential, Phys. Rev. Mater. **3**, 093802 (2019).

[99] S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran, and Z. W. Ulissi, Convolutional neural network of atomic surface structures to

predict binding energies for high-throughput screening of catalysts, J. Phys. Chem. Lett. **10**, 4401 (2019).

[100] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, Machine learning the physical nonlocal exchange-correlation functional of density-functional theory, J. Phys. Chem. Lett. **10**, 6425 (2019).

[101] N. Artrith, A. Urban, and G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, Phys. Rev. B **96**, 014112 (2017).

[102] P. Vidnerová, RBF-Keras: an RBF Layer for Keras Library, (2019), GitHub repository, https://github.com/PetraVidnerova/rbf_keras.

[103] H. L. B. Boström, I. E. Collings, A. B. Cairns, C. P. Romao, and A. L. Goodwin, High-pressure behaviour of Prussian blue analogues: Interplay of hydration, jahn-teller distortions and vacancies, Dalton Trans. **48**, 1647 (2019).

[104] M. Okubo and I. Honma, Ternary metal Prussian blue analogue nanoparticles as cathode materials for Li-ion batteries, Dalton Trans. **42**, 15881 (2013).

[105] J. Jiang, P. Wang, and D. Hou, The mechanism of cesium ions immobilization in the nanometer channel of calcium silicate hydrate: A molecular dynamics study, Phys. Chem. Chem. Phys. **19**, 27974 (2017).

[106] U. Buck, C. C. Pradzynski, T. Zeuch, J. M. Dieterich, and B. Hartke, A size resolved investigation of large water clusters, Phys. Chem. Chem. Phys. **16**, 6859 (2014).

[107] Q. Zhang, B. Li, and L. Chen, First-principles study of microporous magnets M-MOF-74 (M = Ni, Co, Fe, Mn): The role of metal centers, Inorg. Chem. **52**, 9356 (2013).