ACS APPLIED MATERIALS & INTERFACES

Forum Article

Cite This: ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

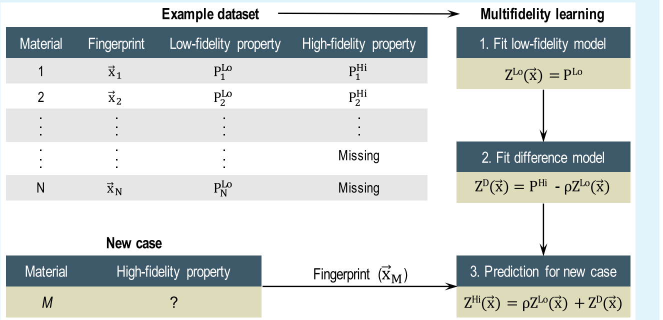
www.acsami.org

Multifidelity Information Fusion with Machine Learning: A Case Study of Dopant Formation Energies in Hafnia

Rohit Batra,[†,‡] Ghanshyam Pilania,[‡] Blas P. Uberuaga,[‡] and Rampi Ramprasad[*,†]

[†] Department of Materials Science & Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States
[‡] Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

ABSTRACT: Cost versus accuracy trade-offs are frequently encountered in materials science and engineering, where a particular property of interest can be measured/computed at different levels of accuracy or fidelity. Naturally, the most accurate measurement is also the most resource and time intensive, while the inexpensive quicker alternatives tend to be noisy. In such situations, a number of machine learning (ML) based multifidelity information fusion (MFIF) strategies can be employed to fuse information accessible from varying sources of fidelity and make predictions at the highest level of accuracy. In this work, we perform a comparative study on traditionally employed single-fidelity and three MFIF strategies, namely, (1) Δ-learning, (2) low-fidelity as a feature, and (3) multifidelity cokriging (CK) to compare their relative prediction accuracies and efficiencies for accelerated property predictions and high throughput chemical space explorations. We perform our analysis using a dopant formation energy data set for hafnia, which is a well-known high-*k* material and is being extensively studied for its promising ferroelectric, piezoelectric, and pyroelectric properties. We use a dopant formation energy data set of 42 dopants in hafnia—each studied in six different hafnia phases—computed at two levels of fidelities to find merits and limitations of these ML strategies. The findings of this work indicate that the MFIF based learning schemes outperform the traditional SF machine learning methods, such as Gaussian process regression and CK provides an accurate, inexpensive and flexible alternative to other MFIF strategies. While the results presented here are for the case study of hafnia, they are expected to be general. Therefore, materials discovery problems that involve huge chemical space explorations can be studied efficiently (or even made feasible in some situations) through a combination of a large number of low-fidelity and a few high-fidelity measurements/computations, in conjunction with the CK approach.



Material	Fingerprint	Low-fidelity property	High-fidelity property
1	$\vec{x}_1$	P_1^{Lo}	P_1^{Hi}
2	$\vec{x}_2$	P_2^{Lo}	P_2^{Hi}
⋮	⋮	⋮	⋮
⋮	⋮	⋮	Missing
N	$\vec{x}_N$	P_N^{Lo}	Missing

1. Fit low-fidelity model

$Z^{Lo}(\vec{x}) = P^{Lo}$

2. Fit difference model

$Z^{D}(\vec{x}) = P^{Hi} - \rho Z^{Lo}(\vec{x})$

New case

Material	High-fidelity property
M	?

3. Prediction for new case

$Z^{Hi}(\vec{x}) = \rho Z^{Lo}(\vec{x}) + Z^{D}(\vec{x})$

KEYWORDS: *multifidelity learning, hafnia, machine learning, density functional theory, dopant formation energy, materials informatics*

1. INTRODUCTION

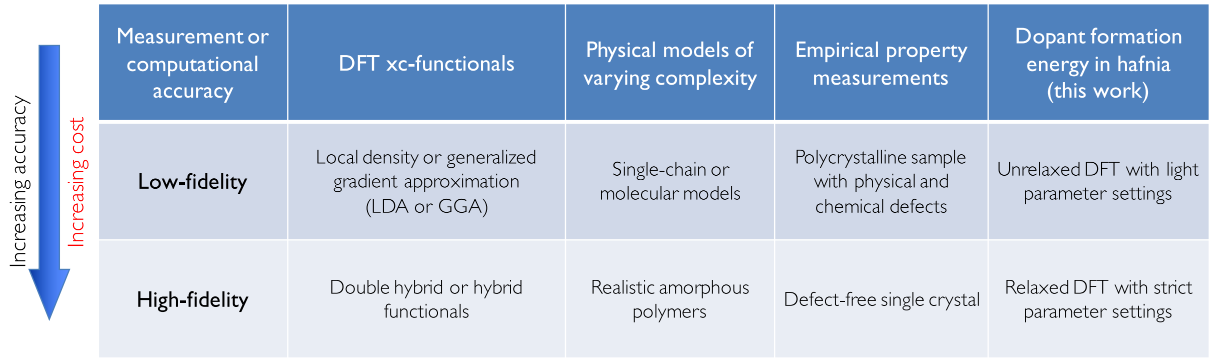
After witnessing transformative feats in distinct areas of artificial intelligence (AI),[1−3] such as computer vision,[4] AI-played games,[5] speech recognition, and natural language processing,[6] machine learning (ML) based methods are gradually finding inroads in the physical and chemical sciences.[7−9] In fact, the data-enabled statistical learning approach has been recently recognized as the fourth paradigm in materials science—following empirical, theoretical, and computational paradigms.[10,11] A vast portfolio of materials classification, regression, and design problems are being approached using ML-based methods, including accelerating materials property prediction, assisting autonomous high-throughput experiments, emulating first-principles computations, and more.[8,12−25]

A common practice in materials informatics is to utilize past data (or "training set"), generated at a consistent level of experiment (or theory), to build efficient surrogate models of diverse material properties, which are otherwise expensive to obtain from fresh experiments or simulations. For example, ML models utilizing Gaussian process regression (GPR), kernel ridge regression (KRR), neural networks, etc., have been made to quickly estimate several macro- or microscopic quantities.[26−33] A notable observation among all such commonly availed ML models is that they are built using training data obtained from a single and consistent source, and thus, can be referred to as single-fidelity (SF) models. However, from a practical standpoint, it is common to find a material property being estimated through multiple sources (experiments/physical models) with varying levels of fidelity. Further, the accuracy of the measurement or simulation also tends to be proportional to its cost, making the volume of high-fidelity data relatively low. A few such examples are discussed next using Figure 1.

Special Issue: Materials Discovery and Design

Received: February 1, 2019
Accepted: April 2, 2019

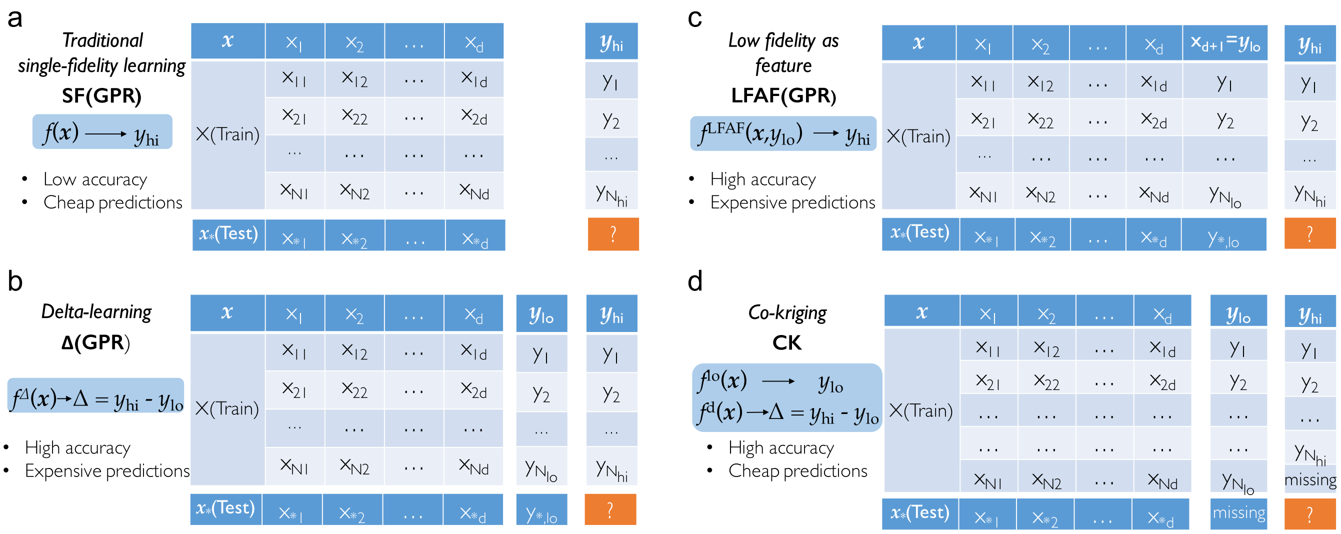
DOI: 10.1021/acsami.9b02174



Increasing accuracy
Increasing cost

Measurement or computational accuracy	DFT xc-functionals	Physical models of varying complexity	Empirical property measurements	Dopant formation energy in hafnia (this work)
Low-fidelity	Local density or generalized gradient approximation (LDA or GGA)	Single-chain or molecular models	Polycrystalline sample with physical and chemical defects	Unrelaxed DFT with light parameter settings
High-fidelity	Double hybrid or hybrid functionals	Realistic amorphous polymers	Defect-free single crystal	Relaxed DFT with strict parameter settings

Figure 1. General trade-off between accuracy and computational/experimental cost prevalent across different domains in materials science. The presence of increasingly complex xc functionals in DFT computations, construction of physical models with varying levels of morphological complexity, and empirical measurements of samples with increasing purity allow estimation of materials properties at varying levels of accuracy. The general trend of increasing time and cost with higher-fidelity measurements can also be observed.



a
Traditional single-fidelity learning
SF(GPR)
$f(x) \longrightarrow y_{hi}$
• Low accuracy
• Cheap predictions
x x_1 x_2 … x_d y_{hi}
X(Train)
x_{11} x_{12} … x_{1d} y_1
x_{21} x_{22} … x_{2d} y_2
x_{N1} x_{N2} … x_{Nd} $y_{N_{hi}}$
x_*(Test) x_{*1} x_{*2} … x_{*d} ?
b
Delta-learning
Δ(GPR)
$f^{\Delta}(x) \rightarrow \Delta = y_{hi} - y_{lo}$
• High accuracy
• Expensive predictions
y_{lo} y_{hi}
$y_{N_{lo}}$ $y_{N_{hi}}$
$y_{*,lo}$?
c
Low fidelity as feature
LFAF(GPR)
$f^{LFAF}(x, y_{lo}) \rightarrow y_{hi}$
• High accuracy
• Expensive predictions
$x_{d+1} = y_{lo}$
d
Co-kriging
CK
$f^{lo}(x) \longrightarrow y_{lo}$
$f^{d}(x) \rightarrow \Delta = y_{hi} - y_{lo}$
• High accuracy
• Cheap predictions
$y_{N_{hi}}$ missing
missing

Figure 2. Comparative description of various ML models, *i.e.*, (a) SF(GPR), (b) Δ(GPR), (c) LFAF(GPR), and (d) CK, employed in the present study to learn dopant formation energies in hafnia. $f(\boldsymbol{x}) \rightarrow y$ represents a machine-learned mapping function that relates the feature space $\boldsymbol{x}$ to the property space y. In this work, the feature vector $\boldsymbol{x}$ comprised a set of chemically informed elemental features and identities of different hafnia phases, while y_{hi} and y_{lo} represent the DFT-computed high-fidelity and low-fidelity values of dopant formation energies in hafnia, respectively. GPR denotes Gaussian process regression. See the text for details on definitions of $f^{\Delta}(\boldsymbol{x})$, $f^{LFAF}(\boldsymbol{x}, y_{lo})$, $f^{lo}(\boldsymbol{x})$, and $f^{d}(\boldsymbol{x})$.

The first example highlights a hierarchy of increasingly accurate (and expensive) calculations that can be performed within the framework of density functional theory (DFT). While results from computations involving inexpensive local density and generalized gradient approximations (LDA or GGA) to the electronic exchange-correlation (xc) interaction may constitute low-fidelity estimates, the use of more expensive hybrid (e.g., HSE06) and double hybrid functionals can provide high-fidelity results or data. Similarly, if one aims to estimate a property of a "realistic" amorphous polymer, a simple physical model based on a single infinite chain or an oligomer molecule can provide an inexpensive first-order low-fidelity estimate. As another example in the context of experiments, characterization of a scintillator compound can be performed with light yield measurements on high-quality single crystals or on polycrystalline ceramics. While the former may represent a high-fidelity expensive measurement only available after laborious and time-consuming single crystal synthesis process, the latter can be considered as a cheaper and relatively quick low-fidelity estimate. The above examples reflect the prevalence of materials information available at multiple levels of fidelity across different domains. Naturally, the important question to be asked is *Can we fuse information from such multiple sources to make ML models that predict at the highest level of fidelity?*

Besides utilizing information available from varying sources, one also needs to be able to use partial information present at different levels. This is particularly critical in materials science, wherein availability of high quality experimental data is often limited while relatively large volumes of low-fidelity data can be accessed. However, despite their practical relevance, development of ML methods that can handle multifidelity information fusion (MFIF) for materials property predictions remains scarce. In fact, to the best of our knowledge, only three such models, namely, the Δ-learning,[34,35] regression models explicitly using low-fidelity as a feature (LFAF)[36,37] and multifidelity cokriging (CK),[38] have been put forward thus far.

Figure 2 highlights the fundamental differences between the three MFIF approaches and the traditionally employed SF models and the nature of data sets utilized in each case. In the

DOI: 10.1021/acsami.9b02174

SF ML approach, a surrogate model maps easily accessible material attributes or features ($\boldsymbol{x}$) to a target property computed or measured at a single level of fidelity (usually the highest available fidelity, *i.e.* y_{hi}). In contrast, the Δ-learning and LFAF approaches fuse information from different sources in a closely related, although nonequivalent, manner. The former learns the differences or Δ between the high-fidelity (y_{hi}) and the low-fidelity (y_{lo}) estimates of the target property, with the final high fidelity ML estimate made by summing the learned Δ to the known low fidelity value (*i.e.*, $y_{hi} = y_{lo} + f^{\Delta}(\boldsymbol{x})$). On the other hand, the LFAF approach explicitly augments the feature $\boldsymbol{x}$ with the low fidelity value to directly learn the corresponding high fidelity estimate (*i*.e., $y_{hi} = f^{\mathrm{LFAF}}(\boldsymbol{x}, y_{lo})$. An additional point to be noted in the SF, Δ-learning, and LFAF approaches is that all of them are general learning schemes independent of the ML algorithm—such as GPR, KRR, or neural networks. Lastly, the relatively advanced CK model learns the available data as two independent Gaussian processes—one for the low fidelity data ($f^{\mathrm{lo}}(\boldsymbol{x})$) and one for the difference between the two fidelities ($f^{\mathrm{d}}(\boldsymbol{x})$)—while explicitly taking into account the pairwise correlations between features, low-fidelity data, and high-fidelity data (see Figure 2).[39] We also note that while each of the three MFIF methods can, in principle, account for more than two fidelities by resorting to a recursive formulation, the CK is most flexible, as it does not require the knowledge of low-fidelity estimates of a new case to make a prediction—whereas both the Δ-learning and LFAF approaches strictly *require* the availability of all the fidelities during prediction.

Although rare, a few examples of aforementioned MFIF approaches can be found in materials science. While the Δ-learning method has been applied to learn formation enthalpies in small molecules,[34,35] LFAF[36] and CK have been utilized to learn bandgap in solids.[40] In a systematic study on LFAF (titled as, crude approximation of property or CEP), Zhang and co-workers clearly showed the advantage of MFIF learning over the traditional ML method, especially for small data sets, using three distinct examples of learning band gap of binary semiconductors, lattice thermal conductivities, and bulk and shear moduli of zeolites.[37] A frequentist analogue of the CK approach has also been studied very recently by employing recursive KRR.[41] In all these examples, many low-fidelity DFT predictions obtained from relatively simple (LDA or GGA) xc functionals were combined with a few complex (HSE) high-fidelity xc functional predictions to make ML models at the highest-level of accuracy. However, a study that systematically explores the comparative performance, merits, and limitations of different SF and MFIF methods remains absent.

In this work, we assess the ability of different SF and MFIF approaches to predict the dopant formation energy in hafnia (HfO_2), with formation energy data obtained at different levels of cost and accuracy. The motivation behind the specific material choice stems from the fact that hafnia is a well-known high dielectric constant material, in which dopants have been extensively utilized to enhance the dielectric permittivity.[42−48] Moreover, there has been a renewed interest in this material owing to the recent observations of ferroelectricity in thin films of doped hafnia.[49−51] Utilization of dopants to enhance its ferroelectric, piezoelectric, and pyroelectric behavior owing to stabilization of particular phases in hafnia is also being studied extensively.[52,53] An improved understanding of the dopant formation energies in different HfO_2 phases is naturally desirable in all the applications mentioned above. Further, this problem acts as an ideal case study to evaluate comparative performance of different SF and MFIF approaches since the combined chemical space of dopants and hafnia phases is modest enough to draw reasonable conclusions, and yet, the cost for the involved DFT computations is practically feasible.

The data set employed for training of different ML models consists of first-principles dopant formation energies computed for a set of 42 substitutional dopants, each studied in six different hafnia phases. Different choices for the plane wave basis set size, sampling density in the reciprocal space and atomic relaxations were utilized to generate low and high-fidelity values for the DFT-computed dopant formation energies, as discussed further in section 2.1. This information is then combined with a set of physically relevant elemental features ($\boldsymbol{x}$) based on dopant chemistry and identity of the different hafnia phases to form a benchmark training data set used to compare the different SF and MFIF regression models illustrated in Figure 2. Since SF, Δ-learning, and LFAF are compatible with any regression scheme, we use a consistent and commonly employed GPR framework across all of them.

Our results suggest that while all MFIF approaches are more accurate than traditional SF methods, CK is the most efficient and flexible among the three MFIF approaches. Models similar in accuracy to Δ-learning and LFAF can be made using CK approach, but at much less training data generation and prediction cost. The ability of the CK model to make prediction for a new case—dopant formation energy of a new dopant in hafnia for this work—solely based on features (and without the knowledge of its respective low-fidelity estimate) renders it a very powerful ML learning approach for property prediction and materials screening. Further, based on the accuracy and cost results obtained on the toy example of hafnia, it can be concluded that the CK approach provides an efficient pathway to tackle problems involving large exploration space (*e.g.* materials discovery), for which obtaining expensive high-fidelity data is impossible.

2. TECHNICAL DETAILS

2.1. Training Data Set.

We start by building a training database of dopant formation energies in bulk HfO_2. The dopant formation energy (F_D^{ph}) of a dopant D in the ph phase of hafnia is defined as

$$F_D^{ph} = [E^{ph}_{DHfO_2^{\mathrm{bulk}}} + E^{\mathrm{hcp}}_{\mathrm{Hf}}] - [E^{ph}_{HfO_2^{\mathrm{bulk}}} + E^{\mathrm{equi}}_{D}] \tag{1}$$

where $E^{ph}_{DHfO_2^{\mathrm{bulk}}}$ and $E^{ph}_{HfO_2^{\mathrm{bulk}}}$ are the respective DFT computed energies of substitutionally doped and pure hafnia. A 96 atoms supercell with 3.125% concentration of substitutional doping was employed to simulate doped HfO_2. $E^{\mathrm{hcp}}_{\mathrm{Hf}}$ and E^{equi}_{D} represent the DFT computed per atom energies of commonly observed bulk elemental phases of hafnium (*i.e.*, the hcp phase) and the dopant D, respectively. These represent the chemical potential of Hf and D, respectively. For the substitutional doping, six different phases of hafnia, namely, monoclinic (M) $P2_1/c$, tetragonal (T) $P4_2/nmc$, cubic (C) $Fm\bar{3}m$, orthorhombic (OA) $Pbca$, polar orthorhombic (P−O1) $Pca2_1$, and another polar orthorhombic (P−O2) $Pmn2_1$, were considered as these are either empirically known to occur under reasonable temperature−pressure variations or are theoretically predicted to have energy close to the equilibrium M phase.[54−58] For the choice of substitutional dopants in the different phases of bulk HfO_2, we selected 42 elemental species, highlighted in Figure 3, from

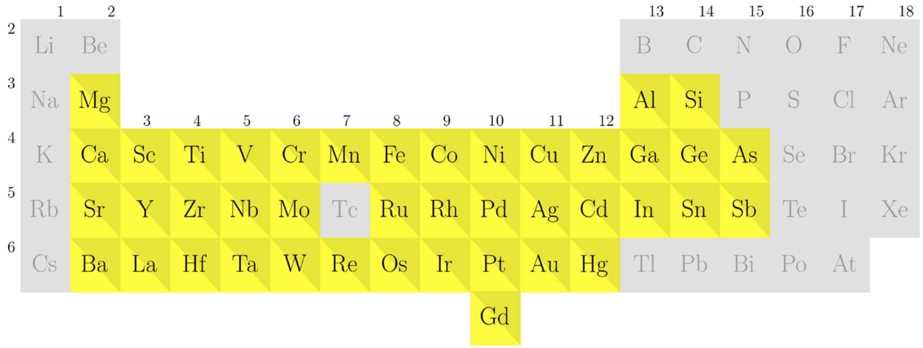


Figure 3. Periodic table of elements highlighting the 42 dopants utilized in this work in yellow.

the 3rd, 4th, and 5th rows of the Periodic Table, with the exception of Al, Si, Mg, and Gd. The rationale for these dopant choices was largely based on the recent reports of interesting ferroelectric observations in hafnia films doped with these elements.[59−61] All of the computational DFT data is made available at our online repository https://khazana.gatech.edu.

The various energy terms described in eq 1 were computed using electronic structure DFT calculations, performed using the Vienna *Ab Initio* Simulation Package[62] (VASP) employing the Perdew−Burke−Ernzerhof (PBE) exchange-correlation functional[63] and the projector-augmented wave methodology.[64] Two data sets with varying levels of fidelity were constructed using the different DFT computation settings as described in Table 1. We note that the last three terms of eq 1 are relatively inexpensive to compute owing to the relatively small bulk supercell size (as compared to the doped supercells) and therefore only high fidelity settings were used to compute these bulk energies of pure phases. On the other hand, the energy term, $E^{ph}_{\mathrm{DHfO}_2^{\mathrm{bulk}}}$, is significantly more computationally demanding, as this quantity needs to be computed for all combinatorial possibilities of dopants in different hafnia phases for a relatively large supercell. Thus, two different settings (*cf.* Table 1) corresponding to the employed *k*-point sampling densities and plane-wave basis set size were used to evaluate this term. The low-fidelity data set was generated using a single *k*-point (at Γ-point) and a basis set of plane waves with kinetic energies only up to 250 eV, while the high-fidelity computations were performed using a 3 × 3 × 3 Monkhorst−Pack *k*-point mesh[65] for the reciprocal space integrations and a basis set of plane waves with kinetic energies up to 500 eV. Furthermore, only in the case of the high-fidelity data set, internal coordinates of the atoms were allowed to relax until all components of the atomic forces along the Cartesian axes were smaller than 10^{-2} eV/Å. In the low-fidelity data set, however, atoms were fixed at their equilibrium bulk positions of the respective hafnia phases. For all doped hafnia computations, the supercell lattice parameters were fixed at their corresponding equilibrium bulk values. Spin polarized computations were performed in each case. The low-fidelity and high-fidelity data sets are henceforth collectively referred to as $\mathbf{y}_{\mathrm{lo}}$ and $\mathbf{y}_{\mathrm{hi}}$, respectively, and the constituting individual formation energy values, corresponding to a given dopant, are indicated with symbols y_{lo} and y_{hi}. We also note that the high-fidelity energy values were found to be consistent with the previous reports.[58,66,67]

Table 1. Details of the Computational Parameters Used during DFT Calculations to Generate Low-Fidelity $\mathbf{y}_{\mathrm{lo}}$ and High-Fidelity $\mathbf{y}_{\mathrm{hi}}$ Datasets

DFT-parameters	low-fidelity	high-fidelity
DFT xc functional	PBE	PBE
plane wave cutoff	250 eV	500 eV
k-points mesh	Γ-point only	3 × 3 × 3
atomic relaxation	false	true

Considering 42 dopants in 6 phases of hafnia results in a total of 252 data points within each of the $\mathbf{y}_{\mathrm{lo}}$ and $\mathbf{y}_{\mathrm{hi}}$ data sets. However, a careful analysis of the final relaxed geometries of the doped hafnia supercells revealed that in four specific cases the doped supercells went through a strain-induced phase transformation during the course of atomic relaxation (for the high-fidelity case) and eventually collapse to a different phase, rendering the definition of F^{ph}_{D} defined in eq 1 invalid. These cases correspond to Cu- and W-doped hafnia T-phase, and Mn- and Si-doped hafnia C-phase, which were excluded from the remaining analysis, and the updated $\mathbf{y}_{\mathrm{lo}}$ and $\mathbf{y}_{\mathrm{hi}}$ data sets of computed dopant formation energies consisted of 248 points each.

Figure 4 shows a good correlation between the calculated dopant formation energies using the aforementioned low- and high-fidelity settings within the DFT computations for the above 248 cases. Interestingly, the dopant formation energy values from the $\mathbf{y}_{\mathrm{lo}}$ data set appear consistently higher than the corresponding values in the $\mathbf{y}_{\mathrm{hi}}$ data set. This is potentially due to the mechanical stresses inherently present in the atomically constrained structures of $\mathbf{y}_{\mathrm{lo}}$ data set. However, a few

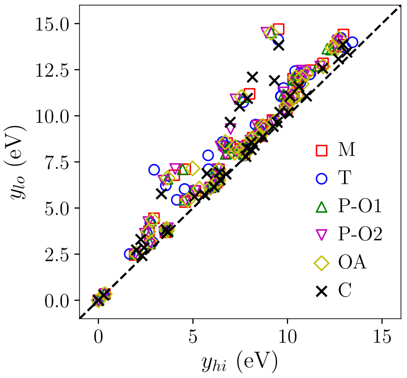


Figure 4. Correlation between the high-fidelity ($\mathbf{y}_{\mathrm{hi}}$) and the low-fidelity ($\mathbf{y}_{\mathrm{lo}}$) dopant formation energies obtained for a total of 42 dopants in six phases of hafnia using DFT computations. The plot contains 248 points, with 4 cases excluded as described in the text.

occurrences of data points for which y_{hi} is slightly greater than y_{lo}—not obvious from Figure 4 owing to relatively large energy scale of the dopant formation energies—confirms additional sources of error owing to the *light* computational settings used to create the low-fidelity data set. It is important to note that a good correlation between the low- and the high-fidelity data sets can be highly beneficial in a MFIF approach as it signifies presence of relevant information at the low-fidelity level as well.

2.2. Details of the Feature Set. Identification of easily accessible and relevant features is an integral part of any statistical learning exercise. Since we are interested in learning dopant formation energies in different phases of hafnia, we select features ($\boldsymbol{x}$) that uniquely represent different dopants and diverse hafnia phases, are readily available and carry information pertaining to the chemical similarity/dissimilarity of the 42 elemental dopant species considered in this work. In particular, we use five elemental features, namely, empirical radii,[68] electronegativity,[69] valency or nominal oxidation state, electron affinity[70] and ionization potential[70] to capture the chemistry of the dopant elemental species. The six hafnia phases were denoted by a categorical variable, referred to as phase id, which varied from 0 to 5. Further discussion on the correlation of these features with the target property $\boldsymbol{y}_{\text{hi}}$ is provided in section 5.1. We note that although additional features can further improve the performance of the ML models shown here, we expect the major conclusions of this study to remain valid, since the feature set $\boldsymbol{x}$ is consistent across all the ML strategies compared in this work.

2.3. Machine Learning Models. As stated earlier, the GPR scheme was utilized for each of the SF, Δ-learning, and LFAF approaches. GPR uses a Bayesian framework, wherein a Gaussian process is used to obtain the aforementioned functional mapping $f(\boldsymbol{x}) \rightarrow y$ based on the available training set and the Bayesian prior, incorporated using the covariance (or kernel) function. The square exponential kernel with three hyper-parameters, *i.e.*, σ_f, σ_l, and σ_n, was chosen for this work, details of which can be found in section 5.2. This choice of kernel is quite standard and is known to work well for a variety of materials problems.[7,13,71] Further, it facilitates a fair comparison with the CK approach, wherein a similar kernel was employed.

In case of the SF models, the set of six features discussed in section 2.2 were used as $\boldsymbol{x}$, and since the standard GPR scheme can only incorporate a single level of fidelity, $y = y_{\text{hi}}$ was used (*cf.* Figure 2a). As alluded to earlier, the Δ-learning and the LFAF are straightforward extensions of traditional GPR scheme, wherein both y_{lo} and y_{hi} data is utilized simultaneously. While in Δ-learning a $f^{\Delta}(\boldsymbol{x}) \rightarrow \Delta$ mapping is established by setting $y = \Delta = y_{\text{hi}} - y_{\text{lo}}$,[34,35] in LFAF a $f^{\text{LFAF}}(\boldsymbol{x}, y_{\text{lo}}) \rightarrow y_{\text{hi}}$ mapping is learned.[36] Both of these approaches are schematically captured in Figure 2b and c, wherein it's important to note in both cases the knowledge of y_{lo} is *essential* for all points in the training and test sets. This particular constraint makes these two MFIF approaches expensive, as will be demonstrated later. These ML models are, henceforth, referred to with labels SF(GPR), Δ(GPR), and LFAF(GPR) to remind that each of these three approaches were implemented using GPR, although any other regression scheme could have been employed.

The flexibility of the CK approach allows it to have a variable number of low- (N_{lo}) and high-fidelity (N_{hi}) data points, as shown in Figure 2d. Given that it is inexpensive to compute y_{lo}, we will constrain the nature of our problem such that $N_{\text{hi}} \subseteq N_{\text{lo}}$, *i.e.*, for all cases whose high-fidelity value y_{hi} is known, the respective low-fidelity value y_{lo} is also available (*cf.* Figure 2d). In analogy to GPR, the CK model assumes the high-fidelity data to be a realization of the Gaussian process $Z_{\text{hi}}(.)$, which is further defined as the sum of a low-fidelity process $Z_{\text{lo}}(.)$ scaled by a factor ρ plus another independent Gaussian process $Z_d(.)$, capturing the difference between the available low- and high-fidelity data points. Thus

$$Z_{\text{hi}}(\boldsymbol{x}) = \rho Z_{\text{lo}}(\boldsymbol{x}) + Z_d(\boldsymbol{x}) \tag{2}$$

Note that the Gaussian processes $Z_{\text{lo}}(\boldsymbol{x})$ and $Z_d(\boldsymbol{x})$ represent the functional mapping $f^{\text{lo}}(\boldsymbol{x})$ and $f^{d}(\boldsymbol{x})$ in Figure 2d, respectively. Further, in order to make a prediction for a new case $\boldsymbol{x}_*$, no knowledge of the low-fidelity data is required. Again, similar to GPR, we describe these two Gaussian processes using the squared exponential kernel defined in section 5.2. The details associated with hyper-parameter optimization, and the mean estimate of the CK approach are also included.

The CK approach can be extended to account for the possible noise present in the high-fidelity data. This involves introduction of a noise parameter in the diagonal of the associated covariance matrices, details of which can be found elsewhere.[38] For this work, a variety of CK models were built by varying the values of N_{lo} and N_{hi} during training, using the aforementioned six-dimensional feature vector $\boldsymbol{x}$ and the noise in data set $\boldsymbol{y}_{\text{hi}}$ set to 0.1 eV.

For all the four ML models, *i.e.*, SF(GPR), Δ(GPR), LFAF(GPR), and CK, the prediction accuracy was computed on a completely unseen and randomly chosen test set consisting of 48 data points (~20% of the total data set of 248 points). This allows us to compare the accuracy of different ML methods and track the learninability of dopant formation energies in hafnia using the feature $\boldsymbol{x}$. Further, in each case, statistically meaningful results were obtained by averaging the performance of various ML models over 50 runs. To avoid overfitting, 5-fold cross validation was adopted in the GPR-based approaches.

3. RESULTS AND DISCUSSION

3.1. SF, Δ-Learning, and LFAF-Learning Models. To estimate the learnability of the regression problem at hand, *i.e.*, the dopant formation energy in hafnia, we start by building the aforementioned SF(GPR), Δ(GPR), and LFAF(GPR) models. Figure 5a presents the average root-mean-square error (RMSE) for all the three approaches as a function of training set size. As expected, the prediction accuracy of all the ML models can be seen to increase with increasing training set size. However, the MFIF based Δ(GPR) and LFAF(GPR) models that use both $\boldsymbol{x}$ and y_{lo} information outperform the SF(GPR) model, demonstrating that the low-fidelity data, although not accurate, contains a significant amount of relevant information. Note that the observed performance disparity is particularly pronounced at relatively small training set sizes, which are often encountered in materials science problems. While the SF(GPR) RMSE converges around 0.70 eV for the largest training size, the corresponding RMSEs for Δ(GPR) and LFAF(GPR) are 0.51 and 0.47 eV, respectively. Further, both the Δ(GPR) and LFAF(GPR) models show similar performance since the two approaches utilize exactly the same amount of information, although the learning problem is cast in a different way. The results presented in Figure 5a, clearly

DOI: 10.1021/acsami.9b02174
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

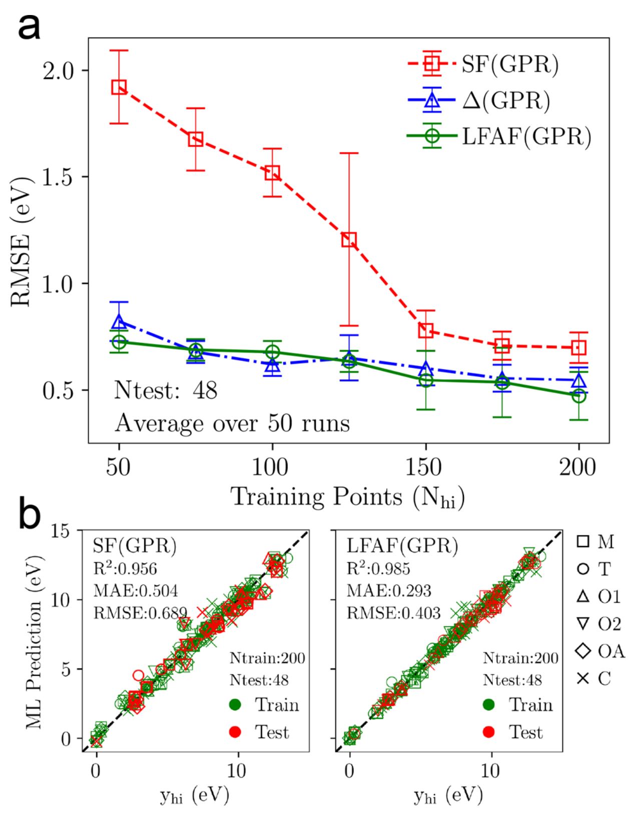


Figure 5. Prediction accuracy for GPR models trained using SF, Δ-learning, and LFAF methods. All results in part a are for test sets consisting of 48 randomly selected points and averaged over 50 runs, with error bars illustrating 1σ deviation. Panels in part b illustrate example parity plots with train and test sets of 200 and 48 points, respectively, in each case. Colors are used to represent training (green) and test (red) sets, and distinct symbols are used to indicate different hafnia phases. Several error measures (R^2, MAE, RMSE) are also provided.

demonstrate the advantage of utilizing information available at lower levels of fidelity to learn target property at a higher level of fidelity. As another benchmark, we evaluated the role played by the feature set $\boldsymbol{x}$ to enable the learning process. Using just the low-fidelity data as feature and linear regression as the ML method, we tried to learn the high-fidelity data. Test error of ~0.95 eV, in contrast to that of 0.47 eV for the MFIF strategies, were obtained. This could be because of (1) use of chemically meaningful features to uniquely represent the dopant-hafnia phase pair and (2) use of nonlinear (GPR) ML algorithms to allow better learning.

While Figure 5a captures the average performance of different models as a function of training set size, panels in Figure 5b compare prediction performance on individual cases included in the training and test sets for the SF(GPR) and LFAF(GPR) models—all for a training size of 200 data points. The results for Δ-learning are quite similar to LFAF, as can be seen from Figure 5a, and are therefore not explicitly shown here. The test set RMSE, mean absolute error (MAE), and R^2 coefficients, frequently referred to as *goodness of fit*, reported in each panel illustrate the improvement in the prediction accuracy moving from SF to MFIF based models; for example, while SF converges at R^2 ≃0.96, the MFIF approaches R^2 ≃0.99. Furthermore, the prediction accuracy can be observed to be similar for all phases of hafnia, represented using different symbols in Figure 5b. As another error measure besides RMSE, we also compared mean absolute relative error (defined as the average of $\left(\left|\frac{y_*^{\mathrm{ML}} - y_{\mathrm{hi}}}{y_{\mathrm{hi}}}\right|\right)$ for the test set) across all the approaches. Similar to RMSE, this was also found to be highest for SF models, in comparison to MFIF models—with errors of 9.5% for SF(GPR), 8.90% for Δ(GPR), 5.22% for LFAF(GPR), and 5.28% for CK models (to be discussed later). Additionally, we note that while the GPR scheme was used to estimate performance of SF, Δ-learning, and LFAF, similar results are expected using any other regression method. We corroborate this using KRR in section 5.3.

An important limitation to emphasize in the context of both the Δ-learning and LFAF approaches is that in order to make predictions for a new case, the corresponding low-fidelity $y_{\mathrm{lo},*}$

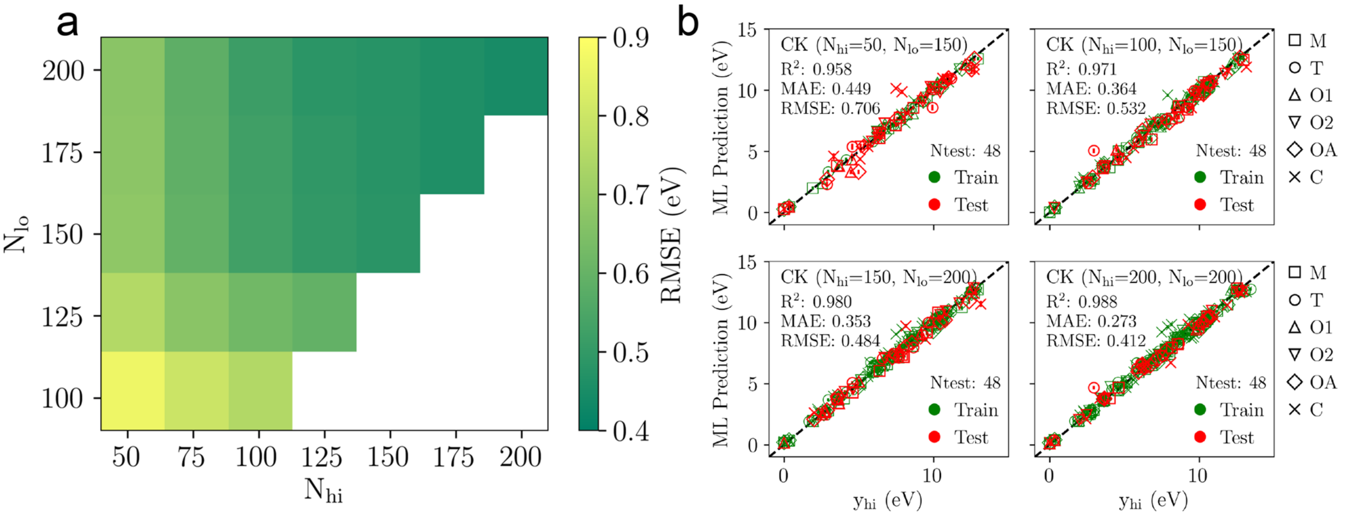


Figure 6. (a) Prediction accuracy for CK models trained using N_{hi} high-fidelity (y_{hi}) and N_{lo} low-fidelity (y_{lo}) data points. All results in part a are for test sets consisting of 48 randomly selected points and averaged over 50 runs. Panels in part b illustrate example parity plots with varying training set and a test set of 48 points. Colors are used to represent training (green) and test (red) sets, and distinct symbols are used to indicate different hafnia phases. Several error measures (R^2, MAE, RMSE) are also provided.

DOI: 10.1021/acsami.9b02174

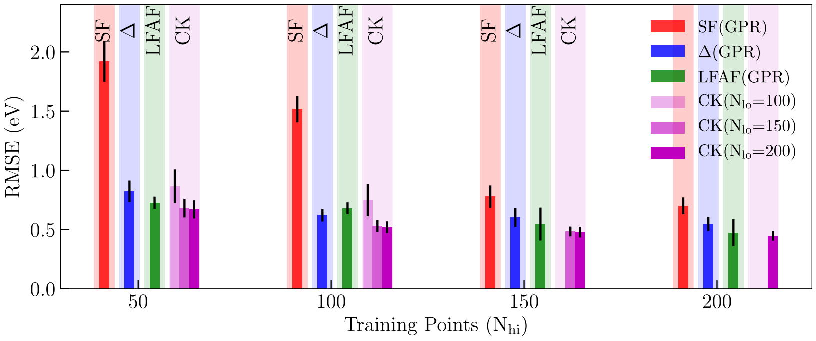


Figure 7. Comparison of prediction accuracy of four different ML models trained to estimate dopant formation energies in hafnia. Results for three flavors of GPR models are shown; the SF based on six chemical features $\boldsymbol{x}$, the Δ(GPR) which maps the difference between low and high fidelity data, and the LFAF(GPR) that uses the low-fidelity value y_{lo} in addition to $\boldsymbol{x}$ as features. The results for three kinds of CK models, varying in number (N_{lo} = 100, 150, or 200) of low-fidelity values y_{lo}, are also included. Cases where $N_{\text{hi}} \nsubseteq N_{\text{lo}}$ are omitted for the CK approach. All results are averaged over 50 runs, with error bars illustrating 1σ deviation.

value should be known beforehand. As referred in Figure 2b and c, this is due to the fact that y_{lo} is used as a feature (in addition to $\boldsymbol{x}$) in the LFAF models, while for the Δ-learning the final prediction is obtained by adding the y_{lo} value to the predicted difference. This requirement renders these approaches particularly ineffective for problems that involve exploration of large chemical spaces, since to make high-fidelity predictions one has to first compute the respective low-fidelity values, which can be either computationally demanding or practically infeasible if the search space is truly vast. Note that, while the low fidelity properties are much less computationally demanding to obtain than the high fidelity properties, they are still much more costly than the other features, which are simply elemental properties requiring no computation. In this sense, Figure 5a serves as a benchmark for two extremes. SF(GPR) represents a scenario where only y_{hi} is known for the training set, while the Δ(GPR) and LFAF(GPR) models exemplify a 2-level case where besides the knowledge of y_{hi} for the training set, y_{lo} is known for both the training and test set. As we will see next, the CK approach is useful for an intermediate, yet practically much more relevant, situation wherein many—but not all—instances of y_{lo} and only some instances of y_{hi} are known.

3.2. Multifidelity Cokriging Model. Figure 6a presents the performance of CK approach as a function of number of low-fidelity (N_{lo}) and high-fidelity (N_{hi}) data points used during the training process. Since we assumed $N_{\text{hi}} \subseteq N_{\text{lo}}$ during the model construction, the lower region in the figure is inapplicable and is intentionally left blank. As evident from the figure, RMSE on the test set (unseen 48 points) decreases with an increase in the number of low-fidelity or high-fidelity data points, with the lowest RMSE of around 0.45 eV when N_{hi} = N_{lo} = 200. Further, the learning rate can be seen to be more sensitive to N_{hi} than N_{lo}. This is understandable as N_{hi} represents the fraction of high-fidelity data that the model is trying to learn, while N_{lo} signifies rather less relevant low-fidelity data. A more interesting area in Figure 6a is the top-left region where RMSE of ~0.66 eV is achieved using only 50 y_{hi} and 200 inexpensive y_{lo} points. In comparison the RMSE for N_{hi} = 50 is around 1.8 eV with SF(GPR), and 0.75 eV for Δ(GPR) or LFAF(GPR) models. It should, however, be noted that this is not a fair comparison between the Δ-learning and the LFAF approaches, and the CK approach as the former cases utilize additional y_{lo} information for test set as well. Nonetheless, the results suggest that the CK approach is particularly beneficial when exceedingly large amounts of low-fidelity data is available in comparison to the high-fidelity data. Such scenarios can be easily encountered when the cost of a low-fidelity estimate is much less than that of a high-fidelity estimate (*cf.* Figure 1). Further, this also means that when one has to explore large chemical spaces, instead of relying on expensive high-fidelity computations (or experiments) one can consider performing large number of low-fidelity along with a few high-fidelity estimates. Sample parity plots in Figure 6b further showcase the improvement in performance of CK model with increasing values of N_{hi} and N_{lo}. As evident from the figure, the R^2 coefficient approaches ≃0.99 (comparable to previously discussed MFIF approaches) with an increase in both N_{lo} and N_{hi}.

As briefly alluded to above, one should be careful while making a comparison between performance of a traditional SF models against the MFIF approaches as different types of data sets are utilized in each case (*cf.* Figure 2). Moreover, model complexity in terms of underlying hyper-parameters can also vary going from one MFIF approach to the other (as detailed in section 2.3). With that caveat, in Figure 7, we attempt to assess the performance of the four models (distinguished using the background colors) from a practical standpoint. Understandably, the CK approach that incorporates additional information from y_{lo} outperforms traditional ML methods based on just SF, *i.e.*, the high-fidelity data alone. Moreover, the greater the amount of y_{lo} data, the better the CK performance—compare models with varying N_{lo} for a particular N_{hi}. However, the most interesting point in Figure 7 is the comparable (or better) performance of CK models to both the Δ(GPR) and LFAF(GPR) models, especially for cases when N_{lo} is much larger than N_{hi}. This obviates the need for making low-fidelity measurements for new cases, which are essential for both Δ-learning and LFAF approaches, every time a prediction has to be made for new cases. In regard to the problem at hand, this would mean that one can estimate the dopant formation energy of a new dopant in one of the six hafnia phases at the high-fidelity level using a large N_{lo} and a small N_{hi} values, thereby saving lots of computational time as will be discussed next.

So far the discussion has been centered around better accuracy (in comparison to SF) or superior flexibility (in comparison to other MFIF approaches) of the CK model. Next, we touch upon another important aspect of any ML model, *i.e.*, its computational cost.

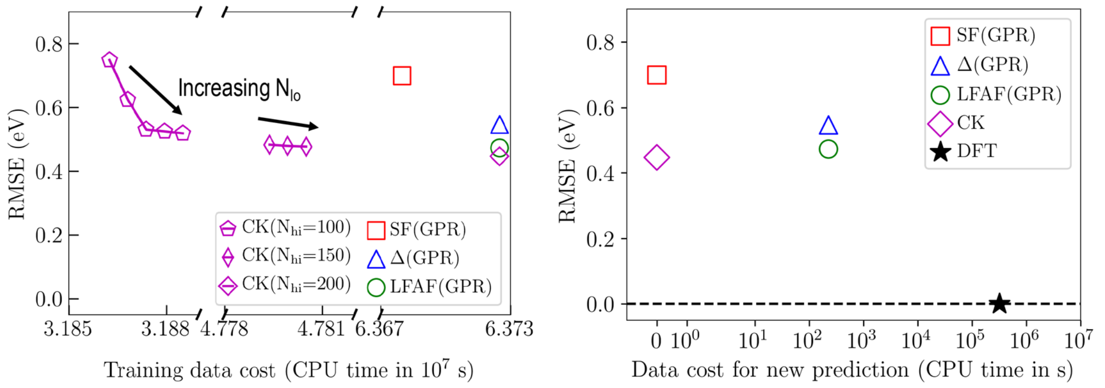


Figure 8. Comparison of prediction accuracy versus computational cost—in terms of training set generation (left) and cost of making a new prediction (right)—for different ML models developed in this work. In part a, the data for SF(GPR) and LFAF(GPR) models corresponds to the scenario with N_{lo} = N_{hi} = 200, while for CK multiple cases with varying N_{lo} and N_{hi} are shown. See text for details.

3.3. Comparison of Accuracy and Cost.

To further establish the advantage of CK approach we plot the accuracy of the best ML models obtained (with N_{hi} = 200) using the different approaches against the computational cost incurred while generating the training data in Figure 8a and while making a new prediction in Figure 8b. It should be noted that cost here refers to the time spent on relevant DFT computations (*i.e.*, computation of y_{lo} or y_{hi} data sets), and not on the ML model evaluation itself, which is relatively much less. The reported cost for a particular low or high fidelity DFT computation was estimated by averaging over the respective cost for the entire data set of 248 points.

An important aspect of any ML model is the time spent to generate the training data itself. For the best models, *i.e.*, with N_{hi} = 200 points, the traditional SF(GPR) model incur DFT cost only for the high-fidelity y_{hi} data, while the Δ(GPR) and the LFAF(GPR) models incur an additional DFT cost for N_{lo} = 200 low-fidelity y_{lo} computations. Thus, the Δ-learning and LFAF approaches, although more accurate, have high training data generation cost. CK models, on other hand, provide a lot more flexibility as one can control the data generation cost by varying the value of N_{lo} and N_{hi}. Although the best model for CK approach has similar data generation cost (and accuracy) to other MFIF approaches with N_{lo} = N_{hi} = 200, this cost can be significantly reduced by choosing large N_{lo} and small N_{hi} values, while only marginally compromising on the model accuracy, as evident from Figure 8a (*e.g.* see performance at N_{hi} = 100 or 150 and varying N_{lo}). Further, it should be noted that the larger the cost difference between the low- and the high-fidelity measurement, the greater the benefit of using the CK approach will be. In rare cases, wherein even the low fidelity measurement is unfeasible (say, a low-fidelity DFT computation on a million atom system) CK is the only practical MFIF approach.

In Figure 8b, prediction costs of the different ML approaches are contrasted with direct computation of the high-fidelity DFT values, which can be seen to be orders of magnitude more expensive than all the ML methods. Ideally, an error (or noise) associated with the high-fidelity DFT computation should also be incorporated here; however, its ignored owing to its relatively smaller (and unknown) magnitude. The most interesting aspect of this figure is the zero cost associated with the CK model and the SF(GPR) models when making a new prediction. This is because these models use only the readily available chemical features $\boldsymbol{x}$ to make a prediction. Δ(GPR) and LFAF(GPR), on the other hand, incur substantial data cost as they require additional low-fidelity DFT computations when making a new prediction.

Thus, the CK approach not only outperforms traditional SF based ML methods in terms of accuracy but also performs comparably to both Δ-learning and LFAF approaches at much lower prediction cost, providing an efficient pathway to utilize data available at multiple levels of fidelity, while not requiring additional measurements for new cases. The implication of these results can be better appreciated in two scenarios: (1) when the exploration space is too big to be able to perform many high-fidelity measurements and (2) when some of the high-fidelity measurements are too expensive or practically infeasible. Search for $A_2BB'X_6$-type double perovskite halides with optimal band gap (or any other property) is an example of the former case in which the chemical space of different combinations of A, B, B′, and X species can be first explored using many (a modest % of the entire chemical space) low-fidelity computations to build an initial CK model and then using it to predict properties for the remaining cases. As highlighted in the first example of Figure 1, the use of expensive DFT functionals to compute properties of large molecules/polymers could be an example of the latter case, in which again estimation from many cheap low-fidelity functionals can be combined with some high-fidelity measurements on small molecules to make predictions for new cases using the CK model—in-line with the work performed in this study. Thus CK is a flexible and valuable MFIF tool for accurate property prediction and discovery of superior materials.

4. CONCLUSIONS

Using the example of dopant formation energies in hafnia, we investigated how different multifidelity information fusion (MFIF) schemes can be utilized to solve common materials science problems that involve availability of data at multiple levels of fidelity (or accuracy). In particular, we studied how low- and high-fidelity hafnia dopant energy data can be combined to create a machine learning model that makes inexpensive but accurate predictions at the high-fidelity level. We compared the performance (both accuracy and data cost) of three MFIF approaches, namely, Δ-learning, low fidelity as feature (LFAF) and cokriging model, along with a commonly used single fidelity approach. A hafnia dopant formation energy data set, consisting of 42 substitutional dopants in 6 different hafnia phases, was chosen to perform this comparative study

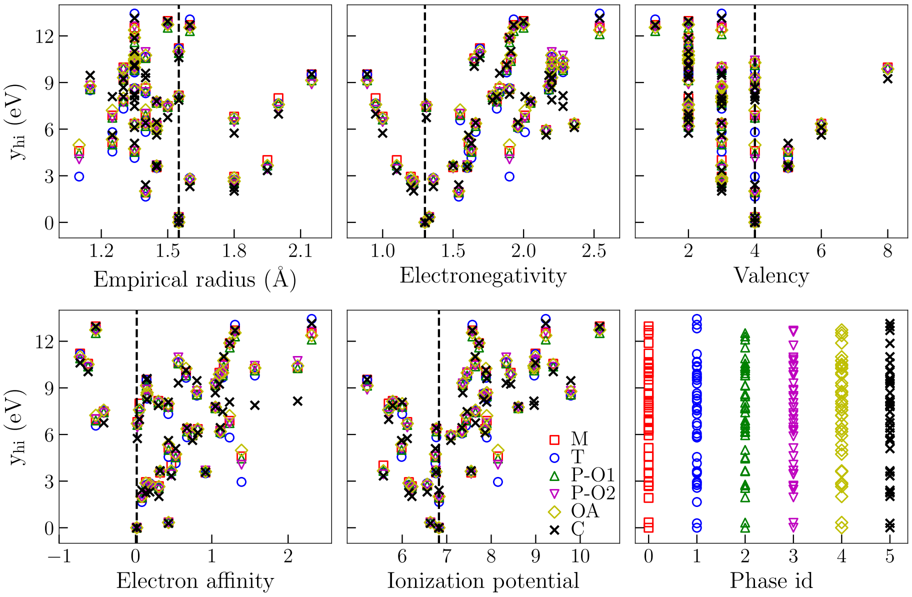


Figure A.1. Correlation between the high-fidelity formation energy (y_{hi}) and the six different features employed in this study. The vertical dashed lines in the first five panels indicate the respective value for Hf.

and illustrate merits and limitations of each approach. The data set was constructed at two levels of fidelity by modifying DFT computational parameters, such as *k*-point sampling and plane wave cutoff energy. Further, the choice of this data set was made because of its modest size (~250 cases) to allow conclusive comparisons of the different ML approaches and the timeliness of the study of dopants in hafnia due to their potential use in ferroelectric and other electronic applications.

Our results clearly suggest that all three MFIF approaches are more accurate than traditional single fidelity based ML methods, such as Gaussian process regression or kernel ridge regression, owing to utilization of additional relevant information contained in the low-fidelity estimates. Further, among the three MFIF approaches, cokriging was found to be most efficient and flexible. While the Δ-learning and LFAF models displayed similar accuracy to that of cokriging models, the former two approaches suffer from a major limitation in the requirement of availability of low-fidelity data when a prediction for a new case has to be made. This increases the prediction cost associated with Δ-learning and LFAF models as time/resources have to be spent to measure the low-fidelity data. The cokriging approach, on the other hand, makes a prediction for a new case at the high-fidelity level using just the features and independent of the knowledge of the corresponding low-fidelity estimate. Moreover, it provides a pathway to combine different numbers of low- and high-fidelity data points during model training, which cannot be achieved using Δ-learning and LFAF approaches. This is particularly useful in problems involving exploration of large chemical spaces or in which its impractical to obtain high-fidelity measurement for a few special cases, as in such scenarios learning models can be built using large number of available low-fidelity and a few high-fidelity data points. As validated in our study, the accuracy for the cokriging model can be significantly better than other MFIF approaches in such cases.

Thus, our work suggests that the multifidelity cokriging approach is particularly useful for searching new materials with superior properties, given that materials science is filled with examples of hierarchical data from high-throughput DFT computations and/or experiments with varying levels of accuracy. Further, instead of just one property, as presented in this work, this approach can easily be extended to simultaneously account for multiple properties, allowing for the search of materials with multiple enhanced functionalities using multiobjective optimization.

5. APPENDIX

5.1. Feature Set Correlations. Different panels in Figure A.1 present a pairwise correlation between each of the aforementioned six features and the high-fidelity dopant formation energies. Although a clear feature−property correlation is missing, a weak trend of lower formation energies for dopants with chemical attributes similar to that of hafnia (marked by vertical dashed lines) is evident from the figure. This is expected as the substitutional dopants with chemical attributes similar to that of Hf should be easily accommodated in the host lattice of hafnia, in-line with the essence of Hume−Rothery rules.[72] At the same time, a strong correlation is unlikely as no single dopant attribute is expected to clearly explain the resulting formation energies.

5.2. Details on Gaussian Process Regression and Multifidelity Cokriging Model. GPR is a Bayesian analogue for the frequentist-based kernel ridge regression.[73] In this, one is interested in finding a probabilistic representation y of a true function $f(x)$, generally accompanied with a measurement-caused white noise ϵ (*i.e.*, $y = f(x) + \epsilon$, with $\mathbb{E}[\epsilon(x)\epsilon(x') = \sigma_n^2\delta(x - x')]$), using the available data and a Bayesian prior expressing beliefs about the target function. Given N training data points, with input and output pairs ($\mathbf{X}$, $\boldsymbol{y}$), and a test input $\boldsymbol{x}_*$, the joint training and test marginal likelihood can be represented as $p(\boldsymbol{y}, y_*) = \mathcal{N}(\boldsymbol{y}, \mathbf{K})$. Here, the covariance matrix $\mathbf{K}$ can be expanded as follows, explicitly distinguishing between the training and test data blocks:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_N + \sigma_n^2\mathbf{I} & \mathbf{K}_{N*} \\ \mathbf{K}_{*N} & \mathbf{K}_{**} \end{bmatrix} \quad (3)$$

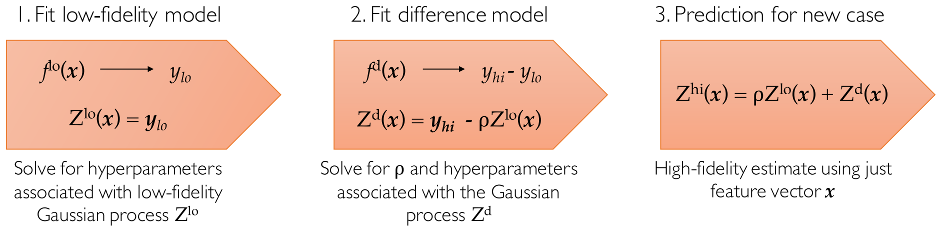

1. Fit low-fidelity model
$f^{lo}(x) \longrightarrow y_{lo}$
$Z^{lo}(x) = y_{lo}$
Solve for hyperparameters associated with low-fidelity Gaussian process Z^{lo}
2. Fit difference model
$f^{d}(x) \longrightarrow y_{hi} - y_{lo}$
$Z^{d}(x) = y_{hi} - \rho Z^{lo}(x)$
Solve for ρ and hyperparameters associated with the Gaussian process Z^{d}
3. Prediction for new case
$Z^{hi}(x) = \rho Z^{lo}(x) + Z^{d}(x)$
High-fidelity estimate using just feature vector x

Figure A.2. Different steps involved in the construction of the CK model. First, a low-fidelity model is built by fitting Z^{lo} Gaussian process on the low-fidelity data. Second, for the available high-fidelity data, a difference model is built after applying a scaling factor (ρ) to the low-fidelity data. Lastly, for a new case, prediction is made using the fitted low-fidelity model, the scaling coefficient, and the difference model.

Note that the individual elements of the covariance matrix are given by

$$\mathbf{K}^{ij} = k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f \exp\left(-\frac{1}{2\sigma_l^2}\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right) + \delta_{ij}\sigma_n^2 \tag{4}$$

where σ_f, σ_l, and σ_n are the hyper-parameters controlling the characteristics of the covariance function (thus the prior information) and are determined by maximizing their log marginal likelihood. After the model hyperparameters have been estimated, the predictive mean (μ_*) and variance (Σ_*) for a test point with feature $\boldsymbol{x}_*$ is obtained by maximizing the conditional likelihood, leading to the expression:

$$\mu_* = \mathbf{K}_{*N}[\mathbf{K}_N + \sigma_n^2\mathbf{I}]^{-1}\boldsymbol{y} \tag{5}$$

$$\Sigma_* = \mathbf{K}_{**} - \mathbf{K}_{*N}[\mathbf{K}_N + \sigma_n^2\mathbf{I}]^{-1}\mathbf{K}_{N*} + \sigma_n^2\mathbf{I} \tag{6}$$

The mathematical framework for the CK approach was first established by Kennedy and O'Hagan,[39] and it has been actively used in the domain of computer experiments and engineering design.[74] Below we briefly describe a few details on formulation of the two-level CK approach, while the details of a general n-level CK scheme can be found elsewhere.[75,76] Say we have N_{lo} and N_{hi} number of low-fidelity and high-fidelity points with feature vectors $\boldsymbol{x}_{\text{lo}}$ and $\boldsymbol{x}_{\text{hi}}$ and property values y_{lo} and y_{hi}, respectively. The different data sets can be cumulatively written as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{\text{lo}} \\ \mathbf{X}_{\text{hi}} \end{pmatrix} = (\boldsymbol{x}_{\text{lo}}^{(1)}, ..., \boldsymbol{x}_{\text{lo}}^{(N_{\text{lo}})}, \boldsymbol{x}_{\text{hi}}^{(1)}, ..., \boldsymbol{x}_{\text{hi}}^{(N_{\text{hi}})})^{\text{T}} \tag{7a}$$

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_{\text{lo}} \\ \boldsymbol{y}_{\text{hi}} \end{pmatrix} = (y_{\text{lo}}^{(1)}, ..., y_{\text{lo}}^{(N_{\text{lo}})}, y_{\text{hi}}^{(1)}, ..., y_{\text{hi}}^{(N_{\text{hi}})})^{\text{T}} \tag{7b}$$

where $\boldsymbol{x}_{\text{lo}}^{(i)}$ and $\boldsymbol{x}_{\text{hi}}^{(i)}$ are d-dimensional feature vector for the ith data point. As stated in the main article, we will consider the nature of our problem such that $\mathbf{X}_{\text{hi}} \subseteq \mathbf{X}_{\text{lo}}$. Now, in analogy to GPR, the property value at $\mathbf{X}$ can be assumed to be a realization of a Gaussian random variable, such that

$$\begin{aligned} Z &= \begin{bmatrix} Z_{\text{lo}}(\mathbf{X}_{\text{lo}}) \\ Z_{\text{hi}}(\mathbf{X}_{\text{hi}}) \end{bmatrix} \\ &= [Z_{\text{lo}}(\boldsymbol{x}_{\text{lo}}^{(1)}), ..., Z_{\text{lo}}(\boldsymbol{x}_{\text{lo}}^{(N_{\text{lo}})}), Z_{\text{hi}}(\boldsymbol{x}_{\text{hi}}^{(1)}), ..., Z_{\text{hi}}(\boldsymbol{x}_{\text{hi}}^{(N_{\text{hi}})})]^{\text{T}} \end{aligned} \tag{8}$$

where $Z_{\text{lo}}(.)$ and $Z_{\text{hi}}(.)$ are Gaussian processes resembling the low- and the high-fidelity data, respectively. Using the autoregressive model introduced in ref 39, the high-fidelity process $Z_{\text{hi}}(.)$ can be assumed to be an outcome of a low-fidelity process $Z_{\text{lo}}(.)$ and an independent Gaussian process $Z_d(.)$ (*i.e.*, $Z_{\text{lo}}(\boldsymbol{x}) \perp Z_d(\boldsymbol{x})$), such that

$$Z_{\text{hi}}(\boldsymbol{x}) = \rho Z_{\text{lo}}(\boldsymbol{x}) + Z_d(\boldsymbol{x}) \tag{9}$$

Further, this model is based on the Markov property that $\text{cov}\{Z_{\text{hi}}(\boldsymbol{x}), Z_{\text{lo}}(\boldsymbol{x}')|Z_{\text{lo}}(\boldsymbol{x})\} = 0, \forall\ \boldsymbol{x} \neq \boldsymbol{x}'$, meaning that nothing more can be learned about the high-fidelity data from the low-fidelity data, if the high-fidelity function at $\boldsymbol{x}$ is known.

Now to describe the Gaussian processes, we consider a squared exponential kernel of the form:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2\Gamma(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{x}') = \sigma^2\left\{\exp\left(-\sum_{j=1}^{d}\theta_j\|\chi_j - \chi'_j\|^2\right)\right\} \tag{10}$$

where χ_j is the jth component of feature vector $\boldsymbol{x}$. Using this kernel definition, the complete covariance matrix for the 2-level CK model reduces to

$$\begin{aligned} \mathbf{K} &= \begin{bmatrix} \text{cov}\{Z_{\text{lo}}(\mathbf{X}_{\text{lo}}), Z_{\text{lo}}(\mathbf{X}_{\text{lo}})\} & \text{cov}\{Z_{\text{hi}}(\mathbf{X}_{\text{hi}}), Z_{\text{lo}}(\mathbf{X}_{\text{lo}})\} \\ \text{cov}\{Z_{\text{hi}}(\mathbf{X}_{\text{hi}}), Z_{\text{lo}}(\mathbf{X}_{\text{lo}})\} & \text{cov}\{Z_{\text{hi}}(\mathbf{X}_{\text{hi}}), Z_{\text{hi}}(\mathbf{X}_{\text{hi}})\} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{lo}}, \mathbf{X}_{\text{lo}}) & \rho\sigma_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{lo}}, \mathbf{X}_{\text{hi}}) \\ \rho\sigma_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{lo}}, \mathbf{X}_{\text{hi}}) & \rho\sigma_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{lo}}, \mathbf{X}_{\text{hi}}) + \rho\sigma_d^2\Gamma_d(\mathbf{X}_{\text{lo}}, \mathbf{X}_{\text{hi}}) \end{bmatrix} \end{aligned} \tag{11}$$

Next, the involved hyperparameters, *i.e.*, ρ, σ_{lo}, σ_d, $\boldsymbol{\theta}_{\text{lo}}$, and $\boldsymbol{\theta}_d$ are solved using maximum likelihood estimate (MLE). Since the low-fidelity data is assumed to be independent of the high-fidelity data, first μ_{lo} (mean), σ_{lo} (variance), and $\boldsymbol{\theta}_{\text{lo}}$ are estimated by maximizing the log-likelihood function of just the low-fidelity data. Then, the parameters (ρ, μ_d, σ_d, and $\boldsymbol{\theta}_d$) associated with the difference model are estimated, again using the MLE. Finally, the CK predictions for a new data $\boldsymbol{x}_*$ is given by

$$\mu_{\text{hi}}(\boldsymbol{x}_*) = \hat{\mu} + \boldsymbol{k}^{\text{T}}\mathbf{K}^{-1}(\boldsymbol{y} - \mathbf{I}\hat{\mu}) \tag{12a}$$

$$\sigma_{\text{hi}}^2(\boldsymbol{x}_*) = \hat{\rho}^2\hat{\sigma}_{\text{lo}} + \hat{\sigma}_d^2 - \boldsymbol{k}^{\text{T}}\mathbf{K}^{-1}\boldsymbol{k} \tag{12b}$$

where $\hat{\mu} = \mathbf{I}\mathbf{K}^{-1}\boldsymbol{y}/\mathbf{I}^{\text{T}}\mathbf{K}^{-1}\mathbf{I}$ and $\mathbf{I}$ is the identity vector. Also, the column vector $\boldsymbol{k}$ is given by

$$\boldsymbol{k} = \begin{bmatrix} \hat{\rho}\hat{\sigma}_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{lo}}, \boldsymbol{x}_*) \\ \hat{\rho}^2\hat{\sigma}_{\text{lo}}^2\Gamma_{\text{lo}}(\mathbf{X}_{\text{hi}}, \boldsymbol{x}_*) + \hat{\sigma}_d^2\Gamma_d(\mathbf{X}_{\text{hi}}, \boldsymbol{x}_*) \end{bmatrix} \tag{13}$$

The overall workflow of the CK model is illustrated in Figure A.2. We make a note that the first step in the CK model is dependent on learning just the low-fidelity data. Thus, if N_{lo} is a large number, it allows the overall CK model to gain some idea (at the level of accuracy of the low-fidelity data) about the different regions of the feature space spanned by the low-fidelity data. In the regions where the high-fidelity data is

DOI: 10.1021/acsami.9b02174

available, a correction in form of the difference model is introduced as part of the second step in Figure A.2. Finally, for a new case prediction is made as sum of the low-fidelity estimate and the corrections terms (*i.e.*, scaling and difference).

5.3. Generalization to Other Regression Models.

Although the learning algorithm in the SF, Δ-learning, and LFAF approaches involved the GPR scheme, their comparative performance is expected to be independent of the learning (or regression) framework employed. Thus, in Figure A.3 we show the relative performance of these approaches using KRR scheme. Clearly, the results for both KRR and GPR schemes are similar.

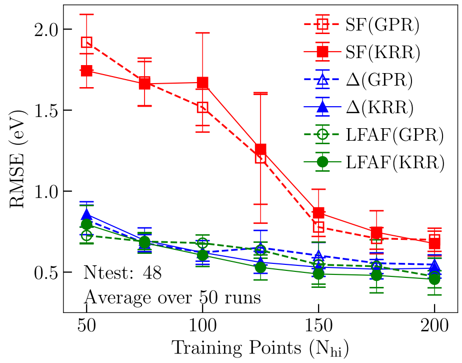


Figure A.3. Prediction accuracy for the KRR and GPR models trained using SF, Δ-learning, and LFAF methods, showcasing the independence of the main conclusions made in this work with respect to the regression scheme employed. All results are for test set consisting of 48 randomly selected points and are averaged over 50 runs, with error bars illustrating 1σ deviation.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rampi.ramprasad@mse.gatech.edu.

ORCID

Rohit Batra: 0000-0002-1098-7035
Ghanshyam Pilania: 0000-0003-4460-1572
Blas P. Uberuaga: 0000-0001-6934-6219
Rampi Ramprasad: 0000-0003-4630-1565

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.B. and R.R. acknowledge financial support for this work through Grant No. W911NF-15-1-0593 from the Army Research Office (ARO). B.P.U. and G.P. acknowledge support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. Los Alamos National Laboratory, an affirmative action equal opportunity employer, is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218NCA000001). Computational support for this work was provided jointly through LANL's high performance computing clusters and an Extreme Science and Engineering Discovery Environment (XSEDE) allocation number TG-DMR080058N.

REFERENCES

(1) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255−260.

(2) Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.

(3) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016.

(4) Janai, J.; Güney, F.; Behl, A.; Geiger, A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State-of-the-art. *arXiv.org* **2017**, 1704.05519.

(5) Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* **2016**, *529*, 484−489.

(6) Kumar, A.; Irsoy, O.; Ondruska, P.; Iyyer, M.; Bradbury, J.; Gulrajani, I.; Zhong, V.; Paulus, R.; Socher, R. Ask me Anything: Dynamic Memory Networks for Natural Language Processing. *International Conference on Machine Learning*; 2016; pp 1378−1387.

(7) Mueller, T.; Kusne, A. G.; Ramprasad, R. *Rev. Comp. Ch.*; John Wiley & Sons, Inc, 2016; pp 186−273.

(8) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547−555.

(9) Lookman, T.; Balachandran, P. V.; Xue, D.; Hogden, J.; Theiler, J. Statistical Inference and Adaptive Design for Materials Discovery. *Curr. Opin. Solid State Mater. Sci.* **2017**, *21*, 121.

(10) Agrawal, A.; Choudhary, A. Perspective: Materials Informatics and Big Data: Realization of the Fourth Paradigm of Science in Materials Science. *APL Mater.* **2016**, *4*, 053208.

(11) Draxl, C.; Scheffler, M. NOMAD: The FAIR Concept for Big-Data-Driven Materials Science. *MRS Bull.* **2018**, *43*, 676−682.

(12) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science: Recent Progress and Emerging Applications. *Rev. Comput. Ch.* **2016**, *29*, 186−273.

(13) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning and Materials Informatics: Recent Applications and Prospects. *npj Comput. Mater.* **2017**, *3*, 54.

(14) Ward, L.; Wolverton, C. Atomistic Calculations and Materials Informatics: A Review. *Curr. Opin. Solid State Mater. Sci.* **2017**, *21*, 167−176.

(15) Alberi, K.; Nardelli, M. B.; Zakutayev, A.; Mitas, L.; Curtarolo, S.; Jain, A.; Fornari, M.; Marzari, N.; Takeuchi, I.; Green, M. L. The 2019 Materials by Design Roadmap. *J. Phys. D Appl. Phys.* **2018**, *52*, 013001.

(16) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 094104.

(17) Seko, A.; Hayashi, H.; Tanaka, I. Compositional Descriptor-based Recommender System for the Materials Discovery. *J. Chem. Phys.* **2018**, *148*, 241719.

(18) Kanamori, K.; Toyoura, K.; Honda, J.; Hattori, K.; Seko, A.; Karasuyama, M.; Shitara, K.; Shiga, M.; Kuwabara, A.; Takeuchi, I. Exploring a Potential Energy Surface by Machine Learning for Characterizing Atomic Transport. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2018**, *97*, 125124.

(19) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat. Commun.* **2016**, *7*, 11241.

(20) Balachandran, P. V.; Young, J.; Lookman, T.; Rondinelli, J. M. Learning from data to design functional materials without inversion symmetry. *Nat. Commun.* **2017**, *8*, 14282.

(21) Pilania, G.; Whittle, K. R.; Jiang, C.; Grimes, R. W.; Stanek, C. R.; Sickafus, K. E.; Uberuaga, B. P. Using Machine Learning to Identify Factors that Govern Amorphization of Irradiated Pyrochlores. *Chem. Mater.* **2017**, *29*, 2574−2583.

DOI: 10.1021/acsami.9b02174

(22) Ziletti, A.; Kumar, D.; Scheffler, M.; Ghiringhelli, L. M. Insightful Classification of Crystal Structures using Deep Learning. *Nat. Commun.* **2018**, *9*, 2775.

(23) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311−2323.

(24) Ramakrishna, S.; Zhang, T.-Y.; Lu, W.-C.; Qian, Q.; Low, J. S. C.; Yune, J. H. R.; Tan, D. Z. L.; Bressan, S.; Sanvito, S.; Kalidindi, S. R. Materials Informatics. *J. Intell. Manuf.* **2018**, 1−20.

(25) Pilania, G.; McClellan, K. J.; Stanek, C. R.; Uberuaga, B. P. Physics-informed Machine Learning for Inorganic Scintillator Discovery. *J. Chem. Phys.* **2018**, *148*, 241729.

(26) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions using Machine Learning. *Sci. Rep.* **2013**, *3*, 2810.

(27) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511−522.

(28) Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A Universal Strategy for the Creation of Machine Learning-based Atomistic Force Fields. *npj Comput. Mater.* **2017**, *3*, 37.

(29) Kim, C.; Pilania, G.; Ramprasad, R. From Organized High-throughput Data to Phenomenological Theory using Machine Learning: The Example of Dielectric Breakdown. *Chem. Mater.* **2016**, *28*, 1304−1311.

(30) Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Ramprasad, R. Solving the Electronic Structure Problem with Machine Learning. *npj Comput. Mater.* **2019**, *5*, 22.

(31) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122*, 17575−17585.

(32) Jha, A.; Chandrasekaran, A.; Kim, C.; Ramprasad, R. Impact of Dataset Uncertainties on Machine Learning Model Predictions: The Example of Polymer Glass Transition Temperatures. *Modell. Simul. Mater. Sci. Eng.* **2019**, *27*, 024002.

(33) Mannodi-Kanakkithodi, A.; Chandrasekaran, A.; Kim, C.; Huan, T. D.; Pilania, G.; Botu, V.; Ramprasad, R. Scoping the Polymer Genome: A Roadmap for Rational Polymer Dielectrics Design and Beyond. *Mater. Today* **2018**, *21*, 785−796.

(34) Huang, B.; Symonds, N. O.; Lilienfeld, O. A. v. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2018; pp 1−27.

(35) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Big Data meets Quantum Chemistry Approximations: The Δ-machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.

(36) Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction Model of Band Gap for Inorganic Compounds by Combination of Density Functional Theory Calculations and Machine Learning Techniques. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 115104.

(37) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *npj Comput. Mater.* **2018**, *4*, 25.

(38) Forrester, A. I.; Sóbester, A.; Keane, A. J. Multi-fidelity Optimization via Surrogate Modelling. *Proc. R. Soc. London, Ser. A* **2007**, *463*, 3251−3269.

(39) Kennedy, M. C.; O'Hagan, A. Predicting the Output from a Complex Computer Code when Fast Approximations are Available. *Biometrika* **2000**, *87*, 1−13.

(40) Pilania, G.; Gubernatis, J. E.; Lookman, T. Multi-fidelity Machine Learning Models for Accurate Bandgap Predictions of Solids. *Comput. Mater. Sci.* **2017**, *129*, 156−163.

(41) Zaspel, P.; Huang, B.; Harbrecht, H.; von Lilienfeld, O. A. Boosting Quantum Machine Learning Models with Multi-level Combination Technique: Pople Diagrams Revisited. *J. Chem. Theory Comput.* **2019**, *15*, 1546.

(42) Robertson, J. High Dielectric Constant Gate Oxides for Metal Oxide Si Transistors. *Rep. Prog. Phys.* **2006**, *69*, 327.

(43) Wilk, G. D.; Wallace, R. M.; Anthony, J. M. High-k Gate Dielectrics: Current Status and Materials Properties Considerations. *J. Appl. Phys.* **2001**, *89*, 5243−5275.

(44) Zhu, H.; Tang, C.; Fonseca, L. R. C.; Ramprasad, R. Recent Progress in Ab Initio Simulations of Hafnia-based Gate Stacks. *J. Mater. Sci.* **2012**, *47*, 7399−7416.

(45) Zhu, H.; Ramanath, G.; Ramprasad, R. Interface Engineering through Atomic Dopants in HfO2-based Gate Stacks. *J. Appl. Phys.* **2013**, *114*, 114310.

(46) Shi, N.; Ramprasad, R. Local Dielectric Permittivity of HfO2 based Slabs and Stacks: A First Principles Study. *Appl. Phys. Lett.* **2007**, *91*, 242906.

(47) Tang, C.; Ramprasad, R. Oxygen Defect Accumulation at Si:HfO2 Interfaces. *Appl. Phys. Lett.* **2008**, *92*, 182908.

(48) Ramprasad, R.; Shi, N. Dielectric Properties of Nanoscale HfO2 Slabs. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *72*, 052107.

(49) Böscke, T. S.; Müller, J.; Bräuhaus, D.; Schröder, U.; Böttger, U. Ferroelectricity in Hafnium Oxide Thin Films. *Appl. Phys. Lett.* **2011**, *99*, 102903.

(50) Schroeder, U.; Yurchuk, E.; Müller, J.; Martin, D.; Schenk, T.; Polakowski, P.; Adelmann, C.; Popovici, M. I.; Kalinin, S. V.; Mikolajick, T. Impact of Different Dopants on the Switching Properties of Ferroelectric Hafnium Oxide. *Jpn. J. Appl. Phys.* **2014**, *53*, 08LE02.

(51) Park, M. H.; Lee, Y. H.; Kim, H. J.; Kim, Y. J.; Moon, T.; Kim, K. D.; Muller, J.; Kersch, A.; Schroeder, U.; Mikolajick, T.; Hwang, C. S. Ferroelectricity and Antiferroelectricity of Doped Thin HfO2-based Films. *Adv. Mater.* **2015**, *27*, 1811−1831.

(52) Starschich, S. Ferroelectric, Pyroelectric and Piezoelectric Effects of Hafnia and Zirconia based Thin Films. Ph.D. thesis, RWTH Aachen University, 2017.

(53) Park, M. H.; Kim, H. J.; Kim, Y. J.; Moon, T.; Do Kim, K.; Hwang, C. S. Toward a Multifunctional Monolithic Device based on Pyroelectricity and the Electrocaloric Effect of Thin Antiferroelectric HfxZr1-xO2 Films. *Nano Energy* **2015**, *12*, 131−140.

(54) Ohtaka, O.; Fukui, H.; Kunisada, T.; Fujisawa, T.; Funakoshi, K.; Utsumi, W.; Irifune, T.; Kuroda, K.; Kikegawa, T. Phase Relations and Volume Changes of Hafnia under High Pressure and High Temperature. *J. Am. Ceram. Soc.* **2001**, *84*, 1369−1373.

(55) Sang, X.; Grimley, E. D.; Schenk, T.; Schroeder, U.; Lebeau, J. M. On the structural origins of ferroelectricity in HfO2 thin films. *Appl. Phys. Lett.* **2015**, *106*, 162905.

(56) Huan, T. D.; Sharma, V.; Rossetti, G. A.; Ramprasad, R. Pathways towards ferroelectricity in hafnia. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 1−5.

(57) Batra, R.; Huan, T. D.; Jones, J. L.; Rossetti, G.; Ramprasad, R. Factors Favoring Ferroelectricity in Hafnia: A First-Principles Computational Study. *J. Phys. Chem. C* **2017**, *121*, 4139−4145.

(58) Barabash, S. V. Prediction of New Metastable HfO2 phases: Toward Understanding Ferro- and Antiferroelectric Films. *J. Comput. Electron.* **2017**, *16*, 1227.

(59) Starschich, S.; Boettger, U. An Extensive Study of the Influence of Dopants on the Ferroelectric Properties of HfO2. *J. Mater. Chem. C* **2017**, *5*, 333−338.

(60) Schroeder, U.; Yurchuk, E.; Müller, J.; Martin, D.; Schenk, T.; Polakowski, P.; Adelmann, C.; Popovici, M. I.; Kalinin, S. V.; Mikolajick, T. Impact of Different Dopants on the Switching Properties of Ferroelectric Hafnium Oxide. *Jpn. J. Appl. Phys.* **2014**, *53*, 08LE02.

(61) Batra, R.; Huan, T. D.; Rossetti, G. A.; Ramprasad, R. Dopants Promoting Ferroelectricity in Hafnia: Insights from a Comprehensive Chemical Space Exploration. *Chem. Mater.* **2017**, *29*, 9102−9109.

(62) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-energy Calculations using a Plane-wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.

(63) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

DOI: 10.1021/acsami.9b02174

(64) Blöchl, P. E. Projector Augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953−17979.

(65) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-zone Integrations. *Phys. Rev. B* **1976**, *13*, 5188−5192.

(66) Falkowski, M.; Künneth, C.; Materlik, R.; Kersch, A. Unexpectedly Large Energy Variations from Dopant Interactions in Ferroelectric HfO2 from High-throughput Ab Initio Calculations. *npj Comput. Mater.* **2018**, *4*, 73.

(67) Lee, C.-K.; Cho, E.; Lee, H.-S.; Hwang, C. S.; Han, S. First-principles Study on Doping and Phase Stability of HfO2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 012102.

(68) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 751−767.

(69) Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.

(70) Haynes, W. M. *CRC handbook of chemistry and physics*; CRC press, 2014.

(71) Mannodi-Kanakkithodi, A.; Pilania, G.; Ramprasad, R. Critical Assessment of Regression-based Machine Learning Methods for Polymer Dielectrics. *Comput. Mater. Sci.* **2016**, *125*, 123−135.

(72) Rothery, W. H. *Atomic Theory: For Students of Metallurgy*; Institute of Metals, 1946.

(73) Rasmussen, C. E.; Williams, C. K. *Gaussian Process for Machine Learning*; MIT press, 2006.

(74) Forrester, A.; Sobester, A.; Keane, A. *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons, 2008.

(75) Le Gratiet, L. Bayesian Analysis of Hierarchical Multifidelity Codes. *SIAM/ASA J. Uncertain. Quant.* **2013**, *1*, 244−269.

(76) Le Gratiet, L.; Garnier, J. Recursive Co-kriging Model for Design of Computer Experiments with Multiple Levels of Fidelity. *Int. J. Uncertain. Quant.* **2014**, *4*, 365.