Discovery of wurtzite solid solutions with enhanced piezoelectric response using machine learning

Drew Behrendt ©, Sayan Banerjee ©, Jiahao Zhang ©, and Andrew M. Rappe ©*

Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania, 19104-6323, USA

While many piezoelectric materials are known, there is still great potential to improve on the figures of merit of existing materials through compositional doping and forming solid solutions. Specifically, it has been shown that doping and alloying wurtzite-structured materials can improve the piezoelectric response; however, a vast compositional space has remained unexplored. In this work, we apply a multilevel screening protocol combining machine learning, chemical intuition, and thermodynamics to systematically discover dopant combinations in the wurtzite material space that improve the desired piezoelectric response. Through our protocol, we use computationally inexpensive screening calculations to consider more than 3000 possible ternary wurtzite solid solutions from nine different wurtzite base systems: AlN, BeO, CdS, CdSe, GaN, ZnO, ZnS, ZnSe, and AgI. Finally, based on thermodynamic analysis and explicit piezoelectric response calculations, we predict 11 materials with improved piezoelectric response, due to the incorporation of electropositive dopants.

I. INTRODUCTION

Piezoelectrics are noncentrosymmetric materials that are capable of interconverting mechanical and electrical energy for a variety of applications [1]. Piezoelectrics provide the basis for microelectronic energy harvesting, acoustic wave devices, actuators, and other devices that are widely used in research, industry, and military applications [2–4]. The energy efficiency and power output of these materials increases with increasing relative piezoelectric response, but decreases with increasing dielectric constant. Perovskite ferroelectrics are prototypical piezoelectrics; however, while perovskites have a very high piezoelectric response, they also possess high dielectric constants [4]. Furthermore, these materials often lose their beneficial properties at high temperatures. An alternate material class is wurtzite; though wurtzites have smaller piezoelectric constants than their perovskite counterparts, they are known for their low dielectric constants and high-temperature performance [5–7]. Furthermore, these materials are highly compatible with complementary metal-oxide-semiconductor (CMOS) technology, which makes them highly attractive for use in piezoelectric applications [2,3]. Doping has long been used to improve the electronic properties of host materials; for example, perovskite lead magnesium niobate (PMN) is alloyed with lead titanate (PT) to enhance the temperature stability and magnitude of piezoelectricity. Adding elements into common wurtzite materials, such as doping aluminum nitride with scandium or zinc oxide with magnesium, has been shown to appreciably increase the piezoelectric response and even induce ferroelectricity, which is of interest for many additional applications [6,8–14].

With the recent growth of machine learning (ML) applications in materials science, many methods have been proposed to speed up materials discovery using ML-based methods [15–23]. Feature selection, where one finds which input features are most correlated to a target output, is a primary challenge of current ML applications in big data. Choosing features is at the heart of this challenge, since it can often be subjective, and the only way to find meaningful correlations is to have representative and meaningful features. Once chosen and selected, however, features can be particularly useful in rational materials design because of their role in providing an interpretative picture of the underlying physics. These material features can be applied in active learning and Bayesian based approaches for accelerated discovery of materials, and new approaches are actively being developed for both computation and experiment [24–28]. New approaches have also been implemented to go beyond just correlation to find causation in structure property relationships and develop explainable ML models to be more transferable and interpretable [29,30].

Feature selection approaches have been shown to be successful for fields spanning from materials discovery [19,31–34] to chemical reaction development using homogeneous [35–38] and heterogeneous [23,39–42] catalysts. Many recent works have identified a "material gene" to perform advanced analysis of material structure-property relations, where the material gene corresponds to the most important feature(s) that correlate with and therefore influence a target property [40,43–50]. The features can be broadly classified into two different types: primitive and calculated. Primitive features, or primitives, are purely based on summary statistics of atomic data such as mass, location on the periodic table, electronegativity, and charges [48]. Calculated features, referred to as proxies, involve calculations such as *ab initio* density functional theory calculations, and these include ionic charges, bond lengths, and bond angles. It is worth noting that, to be most useful, calculated features of a material should be simpler and less time consuming to calculate than the target property of interest. For example, calculation of the lattice parameters requires much less computational time than piezoelectric

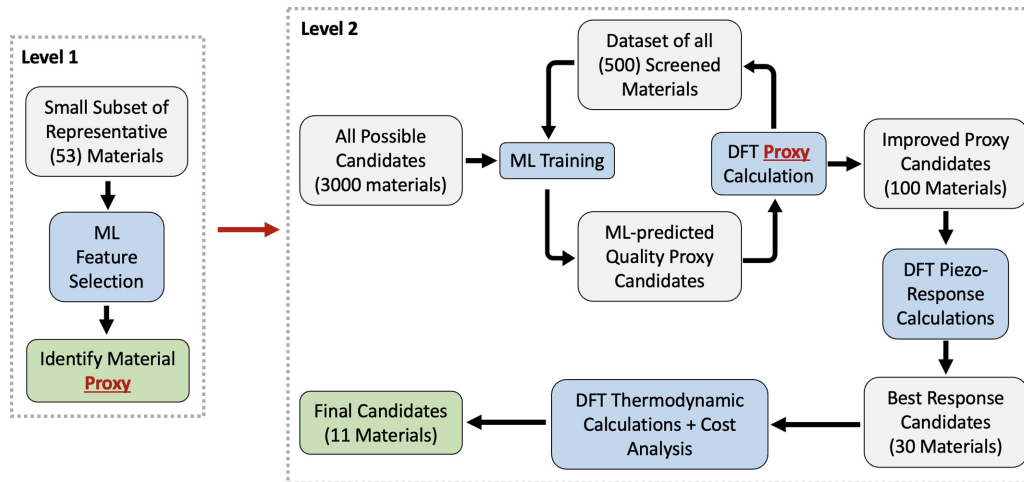*Corresponding author: rappe@sas.upenn.edu

FIG. 1. Multilevel screening workflow to explore the compositional phase space of the nine different wurtzite base materials to maximize the piezoelectric tensor component e_{33}. The gray, blue, and green boxes represent data sets, methods, and deliverables, respectively.

tensor calculations. Our multilevel screening protocol is based on the combination of these two types of featurization.

In the current work, we have investigated nine different base wurtzite materials and their solid solutions with every relevant metal and metalloid element, over 3000 possible candidate materials. We narrow down this space to 30 solid solutions which are predicted to improve the piezoelectric response. Finally, we perform thermodynamic and cost analysis to predict the 11 best candidates for future experimental verification. An overview of this workflow is shown in Fig. 1. The first level of screening is designed to find one screening proxy to use in place of expensive piezoelectric calculations. The second level of the screening serves to reduce the space of all examined ternary solid solutions to just a few suggested high-value materials using automated machine learning candidate selection.

II. COMPUTATIONAL DETAILS

We consider a $2 \times 2 \times 1$ wurtzite supercell consisting of eight metal and eight nonmetal sites. For doped materials, two of the eight cation sites have been replaced [Fig. 2(a)]. The relative positions of the dopants within the unit cell has been shown to affect the piezoelectric response [6], so it is kept constant throughout this entire study to compare elemental doping effects on an equal footing. The calculations for the e_{33} component of the piezoelectric tensor are performed using the Quantum Espresso [51] software package with optimized, norm-conserving pseudopotentials generated by OPIUM [52,53]. To do this, the system is strained along the c axis for five different values (-1.0 to $+1.0$ percent strain), and then atomic positions are allowed to relax. The polarization is then calculated using the Berry's phase method [54,55]. The slope of polarization vs strain was calculated and taken to be the e_{33} component [56] of the piezoelectric tensor for the calculated material. For doping, all combinations of metallic elements that would preserve charge neutrality are considered for each system. For example, combinations of ($+3, +3$) or ($+2, +4$) dopants are considered for AlN, whereas for ZnO, ($+1, +3$) or ($+2, +2$) are screened. Five

different machine learning (ML) methods, linear regression, least absolute shrinkage and selection operator (LASSO), ridge, recursive feature elimination (RFE), and random forest (RF), from python's sklearn are employed [57]. By training multiple machine learning algorithms on each round of the guided sampling, we can ensure faster sampling of a larger candidate space since the algorithms will select different types of materials from the limited initial database, and ultimately we can ensure that if any algorithm finds a candidate to be viable then it is screened. Additionally, after training these different methods we compare the relative effectiveness of each in predicting the material proxy from the primary features.

III. PROXY SELECTION

To identify the best screening proxy, we started with a moderate dataset of additive elements in the most commonly used wurtzite, aluminium nitride (AlN), to improve the piezoelectric response (level 1 in Fig. 1). There have been multiple studies showing the effects of codoping into AlN. Various material descriptors, especially the lattice $\frac{c}{a}$ ratio, have been identified as key properties that correlate with the piezoelectric response [6,7,9,58–60]. In total, 53 materials [61] are included in our initial dataset, and the results for the piezoelectric response are summarized in Table I. Notably, AlN doped with boron and scandium together shows a slight increase in the piezoelectric constant relative to AlN with scandium alone, which is currently attracting great interest [2,3].

TABLE I. Notable piezoelectric response for dopants in AlN, from among the initial dataset of 53 materials.

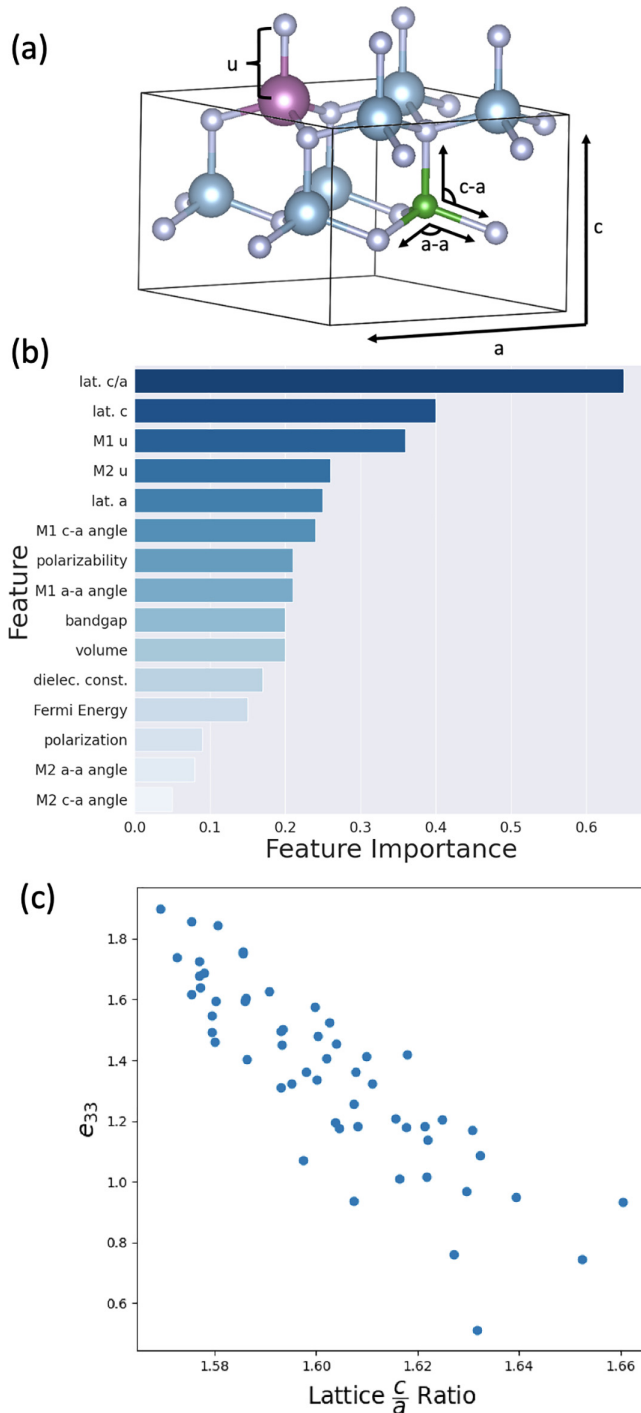| Dopants	e_{33}
None	1.46
Sc	1.86
Sc,B	1.90
Mg,Hf	1.84
Mg,Ti	1.60

FIG. 2. Feature selection in the initial dataset of 53 codoped AlN materials: (a) Example wurtzite structure with dopants. The blue atoms are the cations, replaced with the green and purple atoms as dopants, respectively. (b) ML feature selection to identify the most important material proxy. (c) Scatter plot showing the correlation between lattice $\frac{c}{a}$ and e_{33} response.

For each material in the dataset, we used 15 possible candidate proxy properties, based on previous work on material genes and wurtzite characterization [61]. We then used machine learning feature selection to find the material proxy that is, on average, most correlated to e_{33}, and the results are shown in Fig. 2. Note that the feature importance for each feature is

averaged across all of the ML methods for proxy selection. We find the lattice $\frac{c}{a}$ ratio to be the most important feature [Fig. 2(b)], which has also been reported in previous literature for host materials alone and based on manual manipulation of the lattice parameters for sample systems [7–9]. Therefore, as a lower lattice $\frac{c}{a}$ ratio for wurtzite solid solutions corresponds to a higher value of e_{33} [Fig. 2(c)], in accordance with previous works [7–9], we use the $\frac{c}{a}$ ratio as a screening proxy for our automated workflow (Fig. 1).

IV. GUIDED SAMPLING

Based on the lattice $\frac{c}{a}$ ratio proxy, we generated a set of over 3000 possible candidate solid solutions from nine different wurtzite base systems: AlN, BeO, CdS, CdSe, GaN, ZnO, ZnS, ZnSe, and AgI. Starting from the initial 53-material dataset and base materials, we then used machine learning predictions to iteratively select candidates from the 3000 possible combinations to screen. For each atom, uncalculated descriptors were chosen to reflect the fundamental chemical characteristics, including atomic number, atomic mass, elemental melting temperature, charge, column and row on periodic table, atomic radius [62], electronegativity [63], and preferred valence. Primitive features were collected by taking the average, standard deviation, minimum, maximum, and range of the atomic descriptors for each material (a total of 50 primitives for each material). The key screening steps are (i) training a ML model using the primitives of the materials in the database in a specific iteration, (ii) using the trained ML model to predict which of the unstudied candidates will have the lowest $\frac{c}{a}$ ratio to screen next, and (iii) adding the newly screened materials to the database (the iterative ML screening in Fig. 1). This process was repeated until the best 500 candidates were screened.

During the protocol, it became clear that doped versions of certain base materials are much more likely to have a lower lattice ratio than others (specifically for BeO, AlN, ZnO, and GaN). Therefore, in order to screen the best candidates for each base material, once solid solutions of a given base material are no longer predicted to give significant improvement to the proxy of the host, the rest of the elemental combinations of that base material were removed from the remaining guided sampling candidates. This ensures that we suggest new variants for all nine different base materials, even if some have a higher propensity toward high $\frac{c}{a}$ ratio than others. Additionally, some elements are inherently unstable in the host material and the relaxation calculations fail for a variety of reasons. In this circumstance, a high value of $\frac{c}{a}$ ratio of 2.0 was assigned to these codopants so that the ML predictions would avoid similar materials.

V. INSIGHTS AND DISCUSSION

After screening, we performed fivefold cross validation for each ML model. This is done to ensure that the ML models are accurately predicting the lattice ratios from the primitive features so that we can be confident that all viable candidates are screened correctly; the results are shown in Fig. 3. We find that the random forest algorithm worked best to predict the lattice ratios. However, all of the methods used are relatively
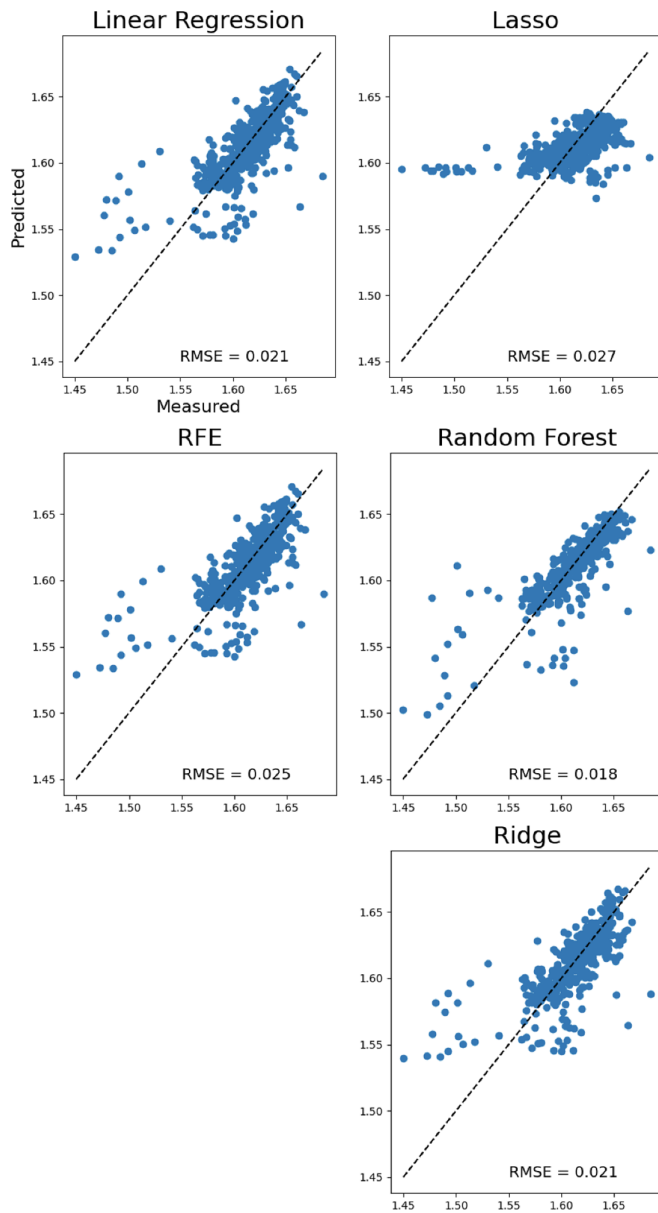
FIG. 3. Validation of the predictive power of ML models. The low root mean square error (RMSE) for multiple models indicates that all the viable candidates are screened.

FIG. 4. Importance of primary features to predicting lattice $\frac{c}{a}$ ratio. "At" stands for atomic, "std" for standard deviation, and the melting temperature is of the material for each element in solid form. All features involving no calculations are summary statistics for the constituent elements in each candidate material.

accurate at predicting the proxy from primitives, which is evident from the low RMSE given in Fig. 3. All of the methods have particular difficulty with materials that are predicted to have $\frac{c}{a}$ ratios near the average, yet are calculated by DFT to possess extremely low lattice ratios; this is in part due to the fact that most of these extreme values represent materials that become unstable and deviate from the original wurtzite structure. Furthermore, because the only poor predictions are in the low $\frac{c}{a}$ regime, we conclude that the unscreened materials predicted to have high $\frac{c}{a}$ are unlikely to be good candidate piezoelectrics. At the end of the screening, only about 1/5 of all the screened solid solutions are found to be insulating and actually effective at reducing the lattice ratio for the parent material; these materials are then verified by subsequent e_{33} calculations in accordance with the workflow in Fig. 1.
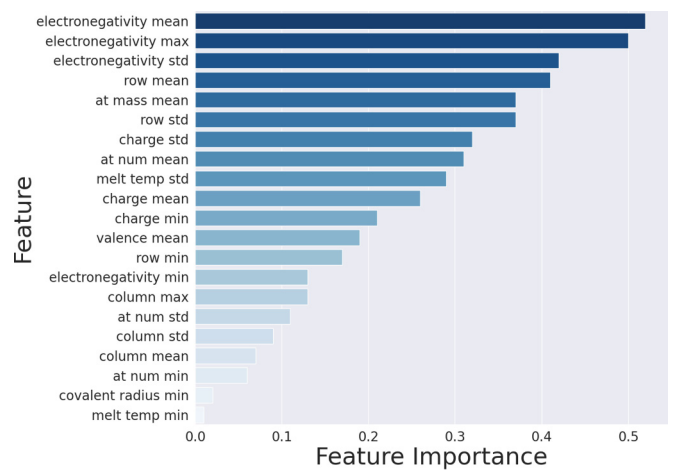
As shown in Fig. 3, the algorithms that reduced the number of features, namely RFE and LASSO, had slightly worse predictive power from the primitive features. This indicates that using many primitives together provides more predictive power of the ultimate lattice ratios in the material. However, determination of the most important features still can provide insight to the structure-property relations, guiding which elements are best to add to each wurtzite system. Since most of the ML algorithms trained on the primitive material features provide reliable prediction of the resulting $\frac{c}{a}$ ratio and insight into the possible piezoelectric response, we performed feature selection to see which primaries were most important (Fig. 4). Two major atomic characteristics stand out as essential to a low $\frac{c}{a}$ ratio and consequently improved piezoelectric response: electronegativity and mass. Materials containing extremely highly electronegative elements as the anion, oxygen, and nitrogen in particular, had characteristically lower $\frac{c}{a}$ ratios. Additionally, materials with a high standard deviation of electronegativity in constituent elements, containing extremely electropositive elements as well, tended to have lower $\frac{c}{a}$ ratios. With the high importance of mean row and atomic mass, materials with a low average row and atomic mass also tended to have low $\frac{c}{a}$ ratios. This means that the addition of small, electropositive elements to AlN, ZnO, and BeO will reliably lead to a heightened electronic response in wurtzites. This finding is aligned with recent evidence of Sc and B enhancing piezoelectricity in AlN, Mg in ZnO, and even the classic example of adding Mg and Nb to lead titanate. However, since the LASSO algorithm had relatively poor predictive power, we posit that these primitives alone are not enough to predict the $\frac{c}{a}$ ratio, and that all (or almost all) of the primitives provide enough information for accurate prediction of the lattice ratio and piezoelectric properties from elemental descriptors.

During the verification of e_{33} using Berry's phase polarization calculations (DFT response calculations in Fig. 1), we find that while all good piezoelectrics have low $\frac{c}{a}$, not all wurtzite solid solutions with low $\frac{c}{a}$ are good piezoelectrics.
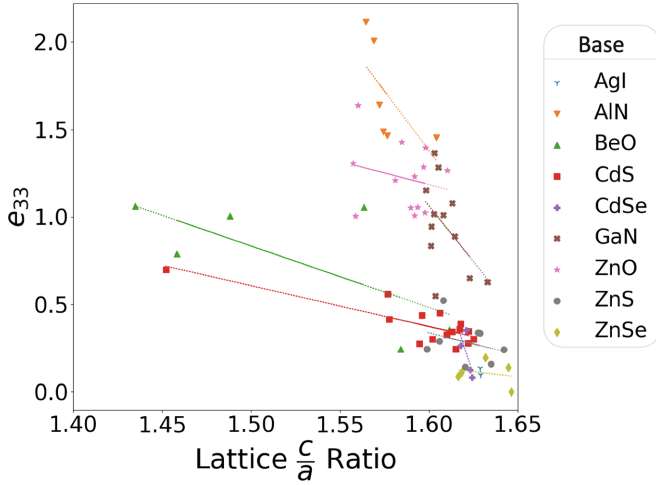
FIG. 5. Lattice ratios and piezoelectric responses for the final screened materials. Each color corresponds to a base material and its respective codopants.

For example, some of the the greatly reduced $\frac{c}{a}$ ratio systems unphysically distort the base material out of the wurtzite phase, leading to an unstable system. We observe that BeO in particular often becomes unstable with the addition of other elements. Furthermore, we find that the trend of lower $\frac{c}{a}$ ratio correlating with higher piezoelectric constant is not a single linear trend for all different wurtzite materials, as suggested in previous work [7], but instead we find a linear trend for dopants within each given base material, as shown in Fig. 5. The resulting 30 materials from this final screening that notably increased the piezoelectric coefficient of their respective base material are listed in Table II. All of these materials are predicted to increase the e_{33} of the parent material. Thermodynamics of the proposed solid solutions were evaluated by comparing the energy of the proposed system with that of the component binary wurtzites with the chosen anion. For reported segregation energies in Table S1, negative values indicate preference for phase segregation and positive values that the segregation is unstable. As these solid solutions are often made by sputtering and ionic migration in the materials is slow and unfavorable, we believe that all proposed solid solutions would be stable at low concentrations. However, combinations with a higher segregation energy are more likely to be stable at higher concentrations. The price and toxicity of proposed dopants was also considered in the practical analysis (Fig. 1). For each material, the best candidate doping combinations after considering such factors are highlighted in bold in Table II. Notably, our set of suggested materials is void of transition metal solid solutions. We found that transition metals with unfilled d orbitals led to complications in the band structure that led to metallic character, states in the band gap that can lead to leakage in real materials, and overall reduced piezoelectric response.

While the proxy $\frac{c}{a}$ ratio helps to drastically reduce the computational expense for the materials screening, it is important to further validate ML-based predictions. The low R^2 values in Fig. 5 indicate that there are complicating factors beyond the $\frac{c}{a}$ ratio that govern the piezoelectric properties, particularly for zinc-containing materials (Fig. 5). Additional

TABLE II. Notable piezoelectric response for codoped systems. Base materials are in parentheses and materials written in bold are of particular interest for increasing the piezoelectric constant of the base system.

Dopants (Base)	c/a	e_{33} (C/m^2)	Improvement	Avg. Dopant cost (USD/kg)
(AlN)	1.604	1.46		
B,Y	1.572	1.64	1.12x	1200
Mg,Hf	1.581	1.84	1.26x	700
Sc	1.575	1.86	1.27x	15 000
Sc,B	**1.569**	**1.90**	**1.30x**	**8700**
Be,Zr	1.569	2.01	1.37x	860
Be,Hf	**1.565**	**2.12**	**1.45x**	**1100**
(BeO)	1.612	0.356		
Li,Y	1.458	0.791	2.22x	75
Li,Sc	1.376	0.954	2.68x	7560
Ga,Na	1.488	1.01	2.84x	140
Mg	**1.435**	**1.06**	**2.98x**	**2**
Al,Li	**1.563**	**1.06**	**2.98x**	**60**
(CdS)	1.625	0.302		
B,K	**1.452**	**0.699**	**2.22x**	**1200**
(CdSe)	1.624	0.082		
Na,Sc	1.618	0.264	3.22x	7500
Ca,Mg	**1.621**	**0.353**	**4.30x**	**5**
(GaN)	1.633	0.629		
Mg,Ti	1.614	0.889	1.41x	3
Be,Zr	1.601	0.946	1.50x	850
B,Sc	1.608	1.01	1.61x	8700
Sc	1.613	1.08	1.72x	15 000
B,Y	**1.598**	**1.15**	**1.83x**	**1200**
(ZnO)	1.610	1.27		
B,Na	1.557	1.31	1.03x	1200
Mg	1.598	1.40	1.10x	3
Be,Mg	1.585	1.43	1.13x	420
Be,Ca	**1.560**	**1.64**	**1.29x**	**420**
(ZnS)	1.642	0.243		
Be,Cd	1.629	0.335	1.38x	420
Be,Mg	1.627	0.340	1.40x	420
Be,Ca	**1.608**	**0.522**	**2.15x**	**420**
(ZnSe)	1.647	0.001		
Na,Sc	**1.644**	**0.140**	**140x**	**7500**
Be,Mg	1.632	0.198	198x	420
(AgI)	1.629	0.098		
Li	**1.628**	**0.138**	**1.41x**	**120**

factors, such as system Born effective charges, are discussed in the Supplemental Material and will be investigated in the future [61]. However, we report that as a screening parameter, the $\frac{c}{a}$ proxy has been effective to find valuable candidate wurtzite solid solutions which can improve the piezoelectric response for all nine base materials.

VI. CONCLUSION

Overall, through a multistep screening protocol and high-throughput study, we identify previously unstudied solid solution candidates to improve the piezoelectric response of nine chosen wurtzite base materials. During this process, we emphasize the fundamental relation of a $\frac{c}{a}$ lattice ratio in wurtzites to its respective piezoelectric response, and use it

as a proxy to develop a computationally inexpensive protocol for piezoelectric material discovery. Furthermore, we are able to support the idea that primitives can be effectively used in machine learning methods to predict basic material properties such as lattice parameters, given a sizable dataset. Finally, we propose the best set of candidates for further experimental verification and discuss the qualitative reasons why certain elements are better for improved response. We hope that the present work serves as a practical example in the design of materials with improved material properties through computationally efficient multistep high-throughput studies.

[1] N. Sezer and M. Koc, A comprehensive review on the state-of-the-art of piezoelectric energy harvesting, Nano Energy **80**, 105567 (2021).

[2] Y. Song, C. Perez, G. Esteves, J. S. Lundh, C. B. Saltonstall, T. E. Beechem, J. I. Yang, K. Ferri, J. E. Brown, Z. Tang, J. P. Maria, D. W. Snyder, R. H. Olsson, B. A. Griffin, S. E. Trolier-Mckinstry, B. M. Foley, and S. Choi, Thermal conductivity of aluminum scandium nitride for 5G mobile applications and beyond, ACS Appl. Mater. Interfaces **13**, 19031 (2021).

[3] V. Yoshioka, J. Lu, Z. Tang, J. Jin, R. H. Olsson, and B. Zhen, Strongly enhanced second-order optical nonlinearity in cmos-compatible $Al_{1-x}Sc_xN$ thin films, APL Mater. **9**, 101104 (2021).

[4] G. Piazza, V. Felmetsger, P. Muralt, R. H. Olsson, and R. Ruby, Piezoelectric aluminum nitride thin films for microelectromechanical systems, MRS Bull. **37**, 1051 (2012).

[5] O. Ambacher, B. Christian, N. Feil, D. F. Urban, C. Elsässer, M. Prescher, and L. Kirste, Wurtzite scaln, inaln, and gaaln crystals, a comparison of structural, elastic, dielectric, and piezoelectric properties, J. Appl. Phys. **130**, 045102 (2021).

[6] H. Momida, A. Teshigahara, and T. Oguchi, Strong enhancement of piezoelectric constants in $Sc_xAl_{1-x}N$: First-principles calculations, AIP Adv. **6**, 065006 (2016).

[7] H. Momida and T. Oguchi, Effects of lattice parameters on piezoelectric constants in wurtzite materials: A theoretical study using first-principles and statistical-learning methods, Appl. Phys. Express **11**, 041201 (2018).

[8] M. Noor-A-Alam, O. Z. Olszewski, and M. Nolan, Ferroelectricity and large piezoelectric response of AlN/ScN superlattice, ACS Appl. Mater. Interfaces **11**, 20482 (2019).

[9] A. Py-Renaudie, P. Daoust, M. Cote, P. Desjardins, and R. A. Masut, *Ab initio* piezoelectric properties of wurtzite ZNO-based alloys: Impact of the c/a cell ratio, Phys. Rev. Mater. **4**, 053601 (2020).

[10] O. Zywitzki, T. Modes, S. Barth, H. Bartzsch, and P. Frach, Effect of scandium content on structure and piezoelectric properties of alscn films deposited by reactive pulse magnetron sputtering, Surf. Coat. Technol. **309**, 417 (2017).

[11] A. Krishnamoorthy, S. C. Tiwari, A. Nakano, R. K. Kalia, and P. Vashishta, Electric-field-induced crossover of polarization reversal mechanisms in $Al_{1-x}Sc_xN$ ferroelectrics, Nanotechnology **32**, 49LT02 (2021).

[12] H. Moriwake, R. Yokoi, A. Taguchi, T. Ogawa, C. A. Fisher, A. Kuwabara, Y. Sato, T. Shimizu, Y. Hamasaki, H. Takashima, and M. Itoh, A computational search for wurtzite-structured ferroelectrics with low coercive voltages, APL Mater. **8**, 121102 (2020).

[13] X. Yu, L. Zhu, X. Li, J. Zhao, T. Wu, W. Yu, and W. Li, Doping engineering for optimizing piezoelectric and elastic performance of AlN, Materials **16**, 1778 (2023).

[14] J. Startt, M. Quazi, P. Sharma, I. Vazquez, A. Poudyal, N. Jackson, and R. Dingreville, Unlocking AlN piezoelectric performance with earth-abundant dopants, Adv. Electron. Mater. **9**, 2201187 (2023).

[15] B. Sanchez-Lengeling and A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science **361**, 360 (2018).

[16] H. Tao, T. Wu, M. Aldeghi, T. C. Wu, A. Aspuru-Guzik, and E. Kumacheva, Nanoparticle synthesis assisted by machine learning, Nat. Rev. Mater. **6**, 701 (2021).

[17] Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, To address surface reaction network complexity using scaling relations machine learning and DFT calculations, Nat. Commun. **8**, 14621 (2017).

[18] M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen *et al.*, Accelerated discovery of CO_2 electrocatalysts using active machine learning, Nature (London) **581**, 178 (2020).

[19] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, Nature (London) **533**, 73 (2016).

[20] A. Zunger, Inverse design in search of materials with target functionalities, Nat. Rev. Chem. **2**, 0121 (2018).

[21] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science, Nature (London) **559**, 547 (2018).

[22] Z. Ren, S. I. P. Tian, J. Noh, F. Oviedo, G. Xing, J. Li, Q. Liang, R. Zhu, A. G. Aberle, S. Sun *et al.*, An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties, Matter **5**, 314 (2022).

[23] H. Xin, Catalyst design with machine learning, Nat. Energy **7**, 790 (2022).

[24] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Comput. Mater. **5**, 21 (2019).

[25] Q. Liang, A. E. Gongora, Z. Ren, A. Tiihonen, Z. Liu, S. Sun, J. R. Deneault, D. Bash, F. Mekki-Berrada, S. A. Khan, K. Hippalgaonkar, B. Maruyama, K. A. Brown, J. Fisher III, and T. Buonassisi, Benchmarking the performance of bayesian optimization across multiple experimental materials science domains, npj Comput. Mater. **7**, 188 (2021).

[26] 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).

[27] Y. Liu, K. P. Kelley, R. K. Vasudevan, H. Funakubo, M. A. Ziatdinov, and S. V. Kalinin, Experimental discovery of structure–property relationships in ferroelectric materials via active learning, Nat. Mach. Intell. **4**, 341 (2022).

[28] M. Ziatdinov, Y. Liu, K. Kelley, R. Vasudevan, and S. V. Kalinin, Bayesian active learning for scanning probe microscopy: From gaussian processes to hypothesis learning, ACS Nano **16**, 13492 (2022).

[29] B. Kailkhura, B. Gallagher, S. Kim, A. Hiszpanski, and T. Y.-J. Han, Reliable and explainable machine-learning methods for accelerated material discovery, npj Comput. Mater. **5**, 108 (2019).

[30] F. Oviedo, J. L. Ferres, T. Buonassisi, and K. T. Butler, Interpretable and explainable machine learning for materials science and chemistry, Acc. Mater. Res. **3**, 597 (2022).

[31] P. V. Balachandran, D. Xue, J. Theiler, J. Hogden, J. E. Gubernatis, and T. Lookman, Importance of feature selection in machine learning and adaptive design for materials, *Materials Discovery and Design* (Springer, Cham, Switzerland, 2018), pp. 59–79.

[32] L. Bassman Oftelie, P. Rajak, R. K. Kalia, A. Nakano, F. Sha, J. Sun, D. J. Singh, M. Aykol, P. Huck, K. Persson, and P. Vashishta, Active learning for accelerated design of layered materials, npj Comput. Mater. **4**, 74 (2018).

[33] B. Cao, L. A. Adutwum, A. O. Oliynyk, E. J. Luber, B. C. Olsen, A. Mar, and J. M. Buriak, How to optimize materials and devices via design of experiments and machine learning: Demonstration using organic photovoltaics, ACS Nano **12**, 7434 (2018).

[34] A. Talapatra, B. P. Uberuaga, C. R. Stanek, and G. Pilania, A machine learning approach for the prediction of formability and thermodynamic stability of single and double perovskite oxides, Chem. Mater. **33**, 845 (2021).

[35] A. Milo, E. N. Bess, and M. S. Sigman, Interrogating selectivity in catalysis using molecular vibrations, Nature (London) **507**, 210 (2014).

[36] S. Singh, M. Pareek, A. Changotra, S. Banerjee, B. Bhaskararao, P. Balamurugan, and R. B. Sunoj, A unified machine-learning protocol for asymmetric catalysis as a proof of concept demonstration using asymmetric hydrogenation, Proc. Natl. Acad. Sci. USA **117**, 1339 (2020).

[37] S. Banerjee, A. Sreenithya, and R. B. Sunoj, Machine learning for predicting product distributions in catalytic regioselective reactions, Phys. Chem. Chem. Phys. **20**, 18311 (2018).

[38] G. dos Passos Gomes, R. Pollice, and A. Aspuru-Guzik, Navigating through the maze of homogeneous catalyst design with machine learning, Trends in Chemistry **3**, 96 (2021).

[39] J. A. Esterhuizen, B. R. Goldsmith, and S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, Nat. Catal. **5**, 175 (2022).

[40] R. B. Wexler, J. M. P. Martirez, and A. M. Rappe, Chemical pressure-driven enhancement of the hydrogen evolving activity of Ni_2P from nonmetal surface doping interpreted via machine learning, J. Am. Chem. Soc. **140**, 4678 (2018).

[41] J. A. Esterhuizen, B. R. Goldsmith, and S. Linic, Theory-guided machine learning finds geometric structure-property relationships for chemisorption on subsurface alloys, Chem **6**, 3100 (2020).

[42] Q. Gao, H. S. Pillai, Y. Huang, S. Liu, Q. Mu, X. Han, Z. Yan, H. Zhou, Q. He, H. Xin, and H. Zhu, Breaking adsorption-energy scaling limitations of electrocatalytic nitrate reduction on intermetallic CuPd nanocubes by machine-learned insights, Nat. Commun. **13**, 2338 (2022).

[43] A. Mazheika, Y. G. Wang, R. Valero, F. Vines, F. Illas, L. M. Ghiringhelli, S. V. Levchenko, and M. Scheffler, Artificial-intelligence-driven discovery of catalyst genes with application to co2 activation on semiconductor oxides, Nat. Commun. **13**, 419 (2022).

[44] P. V. Balachandran, S. R. Broderick, and K. Rajan, Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning, Proc. R. Soc. A: Math. Phys. Eng. Sci. **467**, 2271 (2011).

[45] J. He, J. Li, C. Liu, C. Wang, Y. Zhang, C. Wen, D. Xue, J. Cao, Y. Su, L. Qiao, and Y. Bai, Machine learning identified materials descriptors for ferroelectricity, Acta Mater. **209**, 116815 (2021).

[46] R. Lyu, C. E. Moore, T. Liu, Y. Yu, and Y. Wu, Predictive design model for low-dimensional organic-inorganic halide perovskites assisted by machine learning, J. Am. Chem. Soc. **143**, 12766 (2021).

[47] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Phys. Rev. B **89**, 094104 (2014).

[48] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, npj Comput. Mater. **2**, 16028 (2016).

[49] Y. Xu, L. Jiang, and X. Qi, Machine learning in thermoelectric materials identification: Feature selection and analysis, Comput. Mater. Sci. **197**, 110625 (2021).

[50] D. Behrendt, S. Banerjee, C. Clark, and A. M. Rappe, High-throughput computational screening of bioinspired dual-atom alloys for CO_2 activation, J. Am. Chem. Soc. **145**, 4730 (2023).

[51] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, Quantum espresso: a modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter **21**, 395502 (2009).

[52] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Optimized pseudopotentials, Phys. Rev. B **41**, 1227(R) (1990).

[53] N. J. Ramer and A. M. Rappe, Designed nonlocal pseudopotentials for enhanced transferability, Phys. Rev. B **59**, 12471 (1999).

[54] D. Vanderbilt, Berry-phase theory of proper piezoelectric response, J. Phys. Chem. Solids **61**, 147 (2000).

[55] D. Vanderbilt, *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators* (Cambridge University Press, Cambridge, UK, 2018).

[56] G. Saghi-Szabo, R. E. Cohen, and H. Krakauer, First-principles study of piezoelectricity in $PbTiO_3$, Phys. Rev. Lett. **80**, 4321 (1998).

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. Machine Learning Res. **12**, 2825 (2011).

[58] M. Noor-A-Alam, O. Z. Olszewski, H. Campanella, and M. Nolan, Large piezoelectric response and ferroelectricity in Li and V/NB/Ta Co-doped w-AlN, ACS Appl. Mater. Interfaces **13**, 944 (2021).

[59] Y. Iwazaki, T. Yokoyama, T. Nishihara, and M. Ueda, Highly enhanced piezoelectric property of Co-doped AlN, Appl. Phys. Express **8**, 061501 (2015).

[60] C. Tholander, F. Tasnadi, I. A. Abrikosov, L. Hultman, J. Birch, and B. Alling, Large piezoelectric response of quarternary wurtzite nitride alloys and its physical origin from first principles, Phys. Rev. B **92**, 174119 (2015).

[61] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.8.055406 for all the DFT parameters and methods, iteration of all features and hyperparameters used in ML models, method validation, consideration of other screening proxies, and thermodynamic information on candidate materials.

[62] R. D. Shannon, Revised effective ionic radii and systematic studies of interatomie distances in halides and chaleogenides, Acta Cryst **32**, 751 (1976).

[63] C. Tantardini and A. R. Oganov, Thermochemical electronegativities of the elements, Nat. Commun. **12**, 2087 (2021).