Development of Aluminum Scandium Nitride Molecular Dynamics Force Fields with Scalable Multi-Objective Bayesian Optimization

JESSE M. SESTITO [1,2,3,4] MICHAELA KEMPNER,[1,2]
TEQUILA A.L. HARRIS,[1] EVA ZARKADOULA,[2] and YAN WANG[1]

1.—George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. 2.—Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA. 3.—College of Engineering, Valparaiso University, Valparaiso, IN 46383, USA. 4.—e-mail: jesse.sestito@valpo.edu

Scandium (Sc)-doped aluminum nitride (AlN) exhibits improved piezoelectric properties, which is favorable for sensor applications. Although many experimental studies exist to fine tune the material properties for design purposes, an atomistic level understanding of the structure-property (S-P) relationships is needed, which is the aim of this work. Molecular dynamics can be used to understand the S-P relationships. However, the limited availability of suitable force fields has been a major challenge for accurate property predictions. In this article, a robust force field calibration method using a scalable multi-objective Bayesian optimization approach is presented. Optimizations with three, six, and eight objectives are applied to calibrate aluminum scandium nitride force fields based on the piezoelectric characteristics, modulus of elasticity, and lattice parameters at different Sc-doped levels. The performances of the different force fields are compared, and the performance of the higher dimensional objective problems is discussed. The highly scalable molecular dynamics force field development method is successfully implemented, resulting in the creation of several aluminum scandium nitride molecular dynamics force fields for piezoelectric applications at varying Sc dope levels.

INTRODUCTION

Aluminum nitride (AlN) is a wide-band-gap material that exhibits high thermal conductivity, low thermal expansion, high stability, and corrosion resistance.[1–3] These properties make AlN an ideal candidate for harsh irradiation environments,[4,5] such as nuclear reactors. Because of its good radiation response, AlN is a good candidate material for piezoelectric sensor applications.[6–10] Recent studies have shown that doping AlN with scandium (Sc) can improve the piezoelectric response up to 40%.[11,12] The improved piezoelectric properties of

aluminum scandium nitride (AlScN) make it an ideal candidate for piezoelectric sensor applications.[13–15] The design of new materials with improved properties requires an experimental investigation of their behavior and a fundamental understanding of their properties at the atomistic level.

Molecular dynamics (MD) simulation is a useful tool for obtaining fundamental insights into the material interactions occurring at the atomic level and very short time scales, usually not accessible by experiments.[16–26] One of the limitations of MD simulations is the lack of accurate interatomic force fields suitable for describing the properties of interest. While some MD force fields (MDFF) for AlN are available,[27–29] there is no MDFF for AlScN, limiting the ability to create MD simulations to examine the piezoelectric properties of AlScN.

Due to the lack of experimental data for AlScN, generating new MDFFs can be challenging.[30,31] Furthermore, many parameters or degrees of freedom in multi-body interactions and multiple known properties are required for a MDFF to be calibrated to predict properties of interest. Global optimization methods, such as Bayesian optimization and other machine learning methods, have been used for calibrating large high-dimensional parameter space MDFFs for a specific property.[32–37] Limited work has been conducted to improve the robustness of parameterization with the consideration of multiple properties. Recently, multi-objective Bayesian optimization (MOBO) was used to calibrate a polycaprolactone MDFF for two properties.[38] Bayesian optimization is a surrogate-based global optimization scheme where a surrogate is constructed to approximate the objective function and guide the searching process. In addition, an acquisition function is constructed based on the surrogate in the same search space to assist the sequential sampling. The next sample is determined by maximizing the acquisition function. With the new sample evaluated, the surrogate model is also updated in preparation for the next iteration. In MOBO, the goal of searching is to construct the Pareto front associated with multiple objectives. The acquisition function needs to be constructed to serve this purpose. However, to simultaneously consider multiple properties, a more scalable multi-objective optimization approach is needed. The development of a robust MDFF that can predict properties accurately at different Sc-doped levels in AlScN also requires multi-objective optimization approaches.

In this work, MDFFs to examine the piezoelectric response of AlScN at the atomistic level are created using machine learning methods. The two-objective MDFF development method[38] is extended to a highly scalable method to predict multiple material properties for AlScN. Here, the MDFF calibration with the simultaneous considerations of multiple properties is accomplished by using a scalable multi-objective optimization scheme for high-objective space. An existing AlN MDFF[27] is used as the baseline setting for the development of the AlScN MDFFs. The highly scalable method is applied to simultaneously calibrate the AlScN MDFF for up to eight properties, including the lattice parameters, a and c, at 23.5% Sc-doped AlScN as well as for the piezoelectric coefficient along the c lattice and modulus of elasticity at 23.5%, 33.2%, and 42.7% Sc-doped AlScN. These compositions were chosen based on their measured d_{33} piezoelectric coefficient,[12] which is up to four times the d_{33} coefficient of AlN. The scalability of the MDFF development method is also evaluated by increasing the number of objectives.

METHODS

Molecular Dynamics Simulation Set-up

To calibrate a MDFF, MD simulations are run to predict the target properties, which are also measured experimentally. The differences between the simulation predictions and the experimental observations are minimized by adjusting the MDFF parameters. In this section, the MD process for calculating the d_{33} piezoelectric coefficient, modulus of elasticity, lattice parameters, and melting temperature of AlScN is outlined.

The MD simulations are performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).[39] All MD simulations are performed with a damping time of 1 ps for the selected temperature and pressure. Simulations are run using a Nose-Hoover thermostat at 300 K and 0 atm, unless otherwise stated. A time step of 1 fs is used in all simulations.

The AlN Vashishta MDFF[27] is used for AlN simulations. The AlN material structure is generated by creating a unit cell of the AlN wurtzite crystal. The unit cell is composed of 12 atoms such that the z-dimension is the same length as the lattice parameter c and the x-dimension is twice the length of the lattice parameter a for AlN. The unit cell is then replicated four times in the x-, y-, and z-directions.

For AlScN simulations, aluminum atoms are randomly changed to Sc atoms to generate the desired percentage of doped material such that the Sc dope level is an atomic percentage of Sc content without consideration of nitrogen. This procedure is an approximation, which does not guarantee that the short-range order of the material is accurate. A structure of AlN and the structures of AlScN at 23.5%, 33.2%, and 42.7% Sc dope are generated. These three chosen dope levels are the extents and midpoint of the piezoelectric region for AlScN found experimentally in Ref. 12. Each structure is equilibrated first using an NPT simulation for 20 ps to ensure energy stabilization. Using the equilibrated system as a starting point, the MD simulations, as described below, are run to predict the properties of interest.

To measure the piezoelectric coefficient, an electric field (Q_z) of 250 MV/m is applied in the z-direction. The x- and y-directions are held at a constant pressure of 0 atm; 250 MV/m is selected to simulate the approximate electric field from the piezoelectric effect in AlN.[40,41] The electric field induces a strain in the z-direction for 1 ps to reach equilibrium. With the electric field still applied, the simulation is run for 20 ps, and the piezoelectric coefficient along the c lattice (d_{33}) is calculated as

$$d_{33} = \frac{\in_z}{Q_z} \qquad (1)$$

Development of Aluminum Scandium Nitride Molecular Dynamics Force Fields with Scalable Multi-Objective Bayesian Optimization

3489

where $\in_z$ is the average strain induced by the electric field.

To measure the modulus of elasticity (E), the equilibrated system is stretched in the z-direction at a strain rate of 0.001 ps^{-1}. The pressure in the x- and y- directions is kept at 0 atm, allowing the system to relax in those directions, while the stress in the z-direction is measured. The stress (σ_{zz}) and strain in the z-direction ($\in_z$) follow a linear relationship

$$\sigma_{zz} = E \in_z . \tag{2}$$

The modulus of elasticity is then extracted from the average σ_{zz} and $\in_z$ values from a 50 ps simulation.

To measure the average lattice parameters of the material, the equilibrated system undergoes an NPT simulation for 50 ps. The lattice parameters can be calculated as

$$c = \frac{\bar{l}_z}{n_z} \tag{3}$$

and

$$a = 0.5 \times \frac{\bar{l}_x}{n_x}, \tag{4}$$

where a and c are the lattice parameters for a wurtzite crystal, $\bar{l}_x$ and $\bar{l}_z$ are the average lengths of the simulation cell along the x- and z-axis over the 50 ps simulation, and n_x and n_z are the numbers of unit cell duplications along the two directions.

To estimate the melting temperature, the equilibrated system of size five unit cells cubed is heated from 300 K at a constant rate to 4000 K over 200 ps while holding the pressure at 0 atm. The properties of interest, i.e., density, volume, potential energy, kinetic energy, and mean squared displacement, are calculated at each time step. These properties are expected to follow a linear trend before and after the melting temperature of the material. The melting temperature can be approximated where the linear trend terminates. This follows the methodology outlined in Refs. [42–46]. Calculating the melting temperature for the Vashishta MDFF using a single-phase method, as done in this work, produces a large range for the potential melting temperature, resulting in a low accuracy measurement. The estimated melting temperature is below the true melting temperature with the MDFF. The interface method to calculate the melting temperature[45–47] was attempted, but results were inconsistent and less accurate than the single-phase method for the Vashishta force field.

Formulation of Aluminum Scandium Nitride Molecular Dynamics Force Field

The Vashishta MDFF for AlN can accurately predict the structural properties and dislocation

formation.[48] It is used as the initial baseline setting for the AlScN MDFF. This MDFF already predicts two- and three-body interactions between Al and N. The Vashishta style energy (U) from the two-body interactions is calculated as

$$U_{ij}(r) = \frac{H_{ij}}{r^{\eta_{ij}}} + \frac{Z_i Z_j}{r} \exp\left(-\frac{r}{\lambda_{ij}}\right) - \frac{D_{ij}}{r^4} \exp\left(-\frac{r}{\in_{ij}}\right) - \frac{W_{ij}}{r^6}, r < r_{c,i} \tag{5}$$

where H_{ij} is the strength of the short-range steric repulsion, r is the distance between atom i and j, η_{ij} is the exponent of the steric repulsion term, Z_i is the effective charge, D_{ij} is the strength of the charge-dipole attraction, W_{ij} is the van der Waals interaction strength, and λ_{ij} and $\in_{ij}$ are the screening lengths for Coulomb and charge-dipole terms, respectively.[49] The values of the two-body interaction parameters are listed in Table I.

The potential energy for three-body interactions is calculated as

$$U_{ijk}\left(r_{ij}, r_{ik}, \theta_{ijk}\right)$$
$$= B_{ijk} \frac{\left[\cos\left(\theta_{ijk}\right) - \cos\left(\theta_{0,ijk}\right)\right]^2}{1 + C_{ijk}\left[\cos\left(\theta_{ijk}\right) - \cos\left(\theta_{0,ijk}\right)\right]^2} \exp\left(\frac{\gamma_{ij}}{r_{ij} - r_{0,ij}}\right),$$
$$\left(r_{ij} < r_{0,ij}, r_{ik} < r_{0,ik}\right) \tag{6}$$

where atom i is the center in the three-body interaction, r_{ij} is the distance between atoms i and j, θ_{ijk} is the angle formed by the atoms i, j, and k, B_{ijk} is the strength of the three-body interaction, $\theta_{0,ijk}$ is the covalent bond angle, C_{ijk} is the three-body saturation parameter, γ is the three-body exponent parameter, and r_0 is the three-body range parameter.[49] The values of the three-body interaction parameters are listed in Table I.

The total energy of the system is the summation of all the two- and three-body interactions. To be able to adapt the AlN MDFF for Sc interactions, additional two- and three-body interactions are added to capture the Sc energy. The values of Z_i, λ, $\in$, and r_c for Sc are duplicated from Al as shown in Table I. The values of η, W, and D for the Sc-Sc and Sc-Al interactions are copied from the Al-Al interaction. The value of η for the Sc-N interaction is duplicated from the Al-N interaction. The values of H for the Sc-Sc and Sc-Al interactions as well as H, D, and W for the Sc-N interaction are left as unknown (denoted as P's) and will be optimized. These two-body interactions are shown in Table I. The values of θ_0, C, γ, and r_0 for the new three-body interactions are duplicated from the Al-N-Al interaction. The values of B for these two new interactions are left as unknown. The three-body interactions are also shown in Table I. The unknown P variables must be calibrated to generate an accurate MDFF. The viable value ranges of these

Table I. Vashishta MDFF formulation for the two- and three-body interactions for AlScN MD simulations

Two-body interactions	Z_i, [e]	λ, [Å]	$\in$, [Å]	r_c, [Å]
Al	1.0708	5	3.75	7.6
N	1.0708	5	3.75	7.6
Sc	1.0708	5	3.75	7.6

Two-body interactions	η	H, [eV Å$^\eta$]	D, [eV Å^4]	W, [eV Å^6]
Al-Al	7	507.66860	0	0
Al-N	9	367.05546	24.7978	34.58365
N-N	7	1038.16334	49.5956	0
Sc-Sc	7	P_1	0	0
Sc-N	9	P_2	P_4	P_5
Sc-Al	7	P_3	0	0

Three-body interactions	B, [eV]	θ_0, [°]	C	γ, [Å]	r_0, [Å]
Al-N-Al & N-Al-N	2.1536	109.47	20	1	2.6
Sc-N-Al & Al-N-Sc	P_6	109.47	20	1	2.6
N-Sc-N & Sc-N-Sc	P_7	109.47	20	1	2.6

Table II. List of parameters and limits used for the MOBO

Parameter	Minimum value	Maximum value
P_1, [eV Å$^\eta$]	300	1200
P_2, [eV Å$^\eta$]	300	1200
P_3, [eV Å$^\eta$]	300	1200
P_4, [eV Å^4]	10	60
P_5, [eV Å^6]	10	60
P_6, [eV]	1.5	4
P_7,][eV]	1.5	4

unknown parameters are listed in Table II. These ranges have been selected based on the current values for the two- and three-body interactions between Al and N.

Optimization Set-up for the Calibration of the Force Field

The goal of the calibration is to find the optimal values of variables $\boldsymbol{P} = (P_1, \ldots, P_7)$ such that the piezoelectric coefficient, modulus of elasticity, and lattice constants can be predicted accurately for different Sc dope levels. The computational complexity of this multi-objective optimization problem can grow exponentially as more properties and Sc dope levels are considered. MOBO is utilized to guide the process of selecting the parameters. The specified objectives are to minimize the differences between the predicted and measured properties as

$$\min f_i(\boldsymbol{P}) = \left| \frac{x_i^* - x_i}{x_i^*} \right| \qquad (7)$$

where f_i is the i-th objective, x_i is the i-th property of interest estimated by the MD simulation, and x_i^* is the i-th property of interest observed from experiments.

Here, three AlScN compositions with 23.5%, 33.2%, and 42.7% Sc dope are simulated. Eight different properties for the three compositions are predicted, namely the d_{33} piezoelectric coefficient, modulus of elasticity, and lattice parameters. The properties and the corresponding objectives f_1 to f_8 are listed in Table III. Three versions of the MDFF calibration problem are developed by calibrating the MDFF parameters based on distinct combinations of objectives. In the three-objective problem, the MOBO method is applied to optimize the three objectives (f_1, f_5, f_7) corresponding with the d_{33} coefficient for the three different Sc dope AlScN values. The six-objective problem is to optimize the six objectives $(f_1, f_2, f_5, f_6, f_7, f_8)$ corresponding with the E for the three different Sc dope AlScN values in addition to the objectives in the three-objective problem. The eight-objective problem is to optimize all eight objectives, corresponding with the lattice constants a and c for the 23.5% Sc dope AlScN in addition to the objectives in the six-objective problem. Melting temperature T_m will not be included in the optimization as it was observed to be sufficiently high for AlScN MDFFs. Instead, T_m will be examined after the Pareto front is identified.

MOBO is an active learning method that can be used to guide the calibration process. We previously

Development of Aluminum Scandium Nitride Molecular Dynamics Force Fields with Scalable Multi-Objective Bayesian Optimization

3491

Table III. Comparison of the piezoelectric coefficient, modulus of elasticity, and lattice parameters for 23.5%, 33.2%, and 42.7% Sc dope AlN compared in the three different MOBO setups used to optimize the AlScN MDFF

Sc %	Property, $[x_i]$	f_i	3 Objectives	6 Objectives	8 Objectives
23.5	d_{33}, [pC/N]	f_1	X	X	X
	E, [GPa]	f_2	–	X	X
	a, [Å]	f_3	–	–	X
	c, [Å]	f_4	–	–	X
33.2	d_{33}, [pC/N]	f_5	X	X	X
	E, [GPa]	f_6	–	X	X
42.7	d_{33}, [pC/N]	f_7	X	X	X
	E, [GPa]	f_8	–	X	X

developed a MOBO method called the dimensional overall spread (DOS) and integral hyperarea difference (IHD) method (DOS-IHD), which is scalable to a large number of objectives. The DOS-IHD method is a generalization of the MOBO method developed by Shu *et al.*,[50] where the two-dimensional Pareto front quality metrics are extended to the high-dimensional Pareto front quality metrics DOS and IHD. DOS is the volume of the smallest orthotope that encloses the Pareto front.[51] The IHD is the area under a set of simplexes that define the Pareto front. To do this, n points from the Pareto front are used to perform a linear interpolation to generate $n-1$ dimensional simplexes, which are utilized to find the volume under the Pareto front. A simplex C is defined as

$$C(f) = \left\{ \sum_{i=1}^{n} \lambda_i s^{[i]}(f) \middle| \sum_{i=1}^{n} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \text{ for } i = 1, \ldots, n \right\}$$

(8)

where $f = (f_1, \ldots, f_{n-1})$ are $n-1$ objective values and $s^{[k]}(\cdot)$ returns order statistics as $s^{[k]}(f) = p^{[k]}$ such that $\|f - p^{[k]}\| \leq \|f - p^{[k+1]}\|$ $(k = 1, \ldots, m-1)$ and $\|f - p^{[1]}\| = \min\{\|f - p^{(1)}\|, \ldots, \|f - p^{(m)}\|\}$. $p^{(k)} = \left(p_1^{(k)}, \ldots, p_n^{(k)}\right)$ $(k = 1, \ldots, m)$ is a Pareto solution vector of n objective values and m Pareto solutions. Here $\|\cdot\|$ denotes the Euclidean distance of the $n-1$ objectives in the n-dimensional space, as

$$\|f - p^{(k)}\| = \sqrt{\left(f_1 - p_1^{(k)}\right)^2 + \ldots + \left(f_{n-1} - p_{n-1}^{(k)}\right)^2}.$$

(9)

A simplex evaluation function c is then defined to return the n-th objective value on the simplex C given the other $n-1$ objective values such that

$$c(f) = f_n.$$

(10)

The quality metric IHD is defined based on the estimation of the integral between the Pareto front approximated by the simplexes and $\min\mathcal{P}$, given by

$$\text{IHD} = \int_{\min\mathcal{P}_{(1)}}^{\max\mathcal{P}_{(1)}} \ldots \int_{\min\mathcal{P}_{(n-1)}}^{\max\mathcal{P}_{(n-1)}} (c(f_1, \ldots, f_{n-1}) - \min\mathcal{P}_{(n)}) df_1 \ldots df_{n-1}$$

(11)

where $\max\mathcal{P}_{(i)} = \max\left\{p_i^{(1)}, \ldots, p_i^{(m)}\right\}$ denotes the maximum of the i-th objective values in the set of Pareto solutions and $\min\mathcal{P}_{(i)} = \min\left\{p_i^{(1)}, \ldots, p_i^{(m)}\right\}$ denotes the minimum.

The DOS-IHD MOBO method is utilized here for calibration. However, any high-dimensional MOBO method, such as the expected improvement method (EIM)[52] combining the elements of the EIM into scalar functions, could be used instead of the DOS-IHD method. Another method for calibrating the MDFFs would be to create a single objective problem such that the objective is the weighted sum of the errors of all properties which are to be minimized.[53–55] In this approach, solving the single optimization problem becomes trivial. However, the weights are based on the designer's preferences and are subjective. Optimization results are only useful for the particular design. The multi-objective approach allows for the designer to make their design decisions for each application without having to rerun the optimization problem.

A third Pareto front quality metric, the Gaussian hyperarea difference (GHD), is defined to further examine the quality of the Pareto front of the results from the MOBO method. The GHD is the error associated with the Gaussian process regression model constructed as the surrogate of the Pareto front. Fifty initial samples of P are generated using the Latin hypercube sampling method,[56,57] with their corresponding objectives calculated using MD, to construct the initial surrogate. MOBO will guide the selection of the next P until a maximum of 1000 MDFFs have been generated. The simulations are executed with 12 processors for each MD simulation using servers on the Partnership for an Advanced Computing Environment (PACE) at Georgia Institute of Technology.[58]
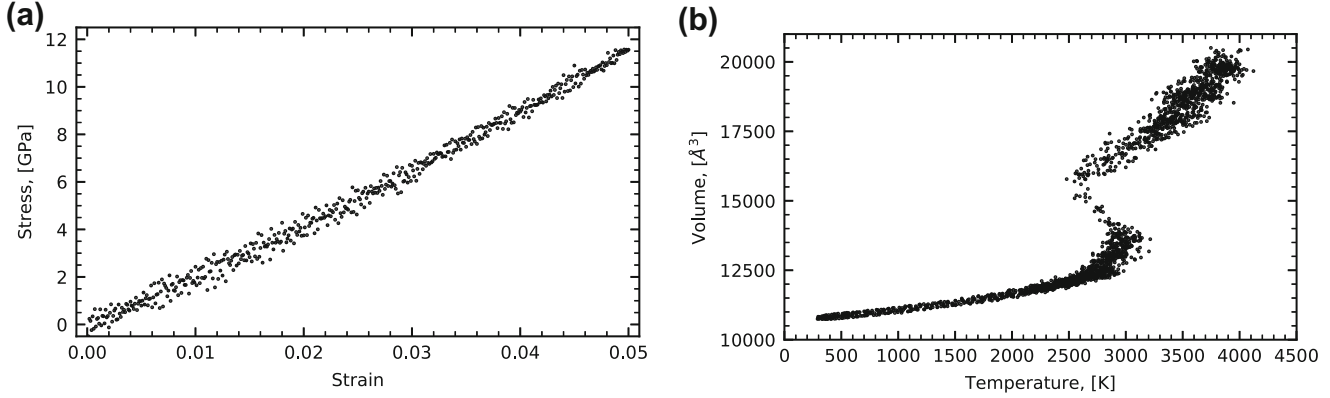
Fig. 1. Selected datasets from the AlN MD simulations to calculate the (a) modulus of elasticity and (b) melting temperature. Data for (a) was generated by deforming an equilibrated MD system at a constant strain rate and measuring the stress induced. Data for (b) were generated by heating up an equilibrated system from 300 K to 2000 K over 200 ps at 0 atm.
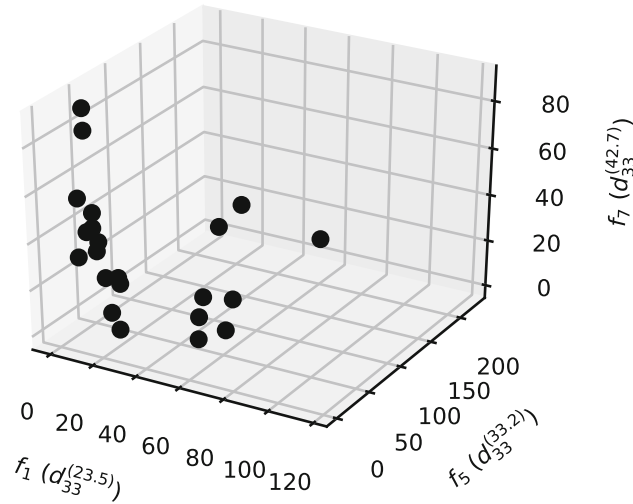

Fig. 2. Pareto front for three-objective MOBO using the DOS-IHD method.

RESULTS AND DISCUSSION

Aluminum Nitride Force Field Characterization

Before Sc is introduced into the AlN Vashishta MDFF, the AlN MDFF is first evaluated with E, d_{33}, lattice constants, and T_m. The tensile test results in a linear correlation between stress and strain, as shown in Fig. 1a, and correspondingly a E value of 233 GPa. The modulus of elasticity is underestimated compared to the experimentally obtained value of 338 GPa.[12] With an electric field of 250 MV/m applied in the z-direction, an average strain of 0.001595 is observed. The resultant d_{33} coefficient is calculated to be 6.38 pC/N, which is in agreement with the experimental measurements, ranging from 5.1 pC/N to 8.4 pC/N.[12,59,60] The lattice parameters are calculated to be 3.119 Å and 4.999 Å for a and c,

respectively. These values are in agreement with the corresponding experimental measurements of a as 3.1114 Å and c as 4.9792 Å.[12] The melting temperature, T_m, is selected from the non-linear region from the MD simulation which occurs between 2500 K and 3200 K, as shown in Fig. 1b. This is in agreement with the experimental value of 3100 K.[61]

Aluminum Scandium Nitride Force Fields

The MOBO method is applied to calibrate the AlScN MDFFs. The three-objective case is first examined, where the d_{33} values for 23.5%, 33.2%, and 42.7% Sc dope AlScN are optimized. One thousand MDFFs are generated, which results in a Pareto front of only 22 MDFFs, as shown in Fig. 2. Among all 22 Pareto solutions, the error in the d_{33} value for one Sc dope level cannot be further reduced without increasing the error in the other Sc dope levels. The values of the Pareto quality metrics are 132.2^3, 59.8^3, and 29.8^3 for the DOS, IHD, and GHD, respectively. A high DOS value and low IHD and GHD values are favorable. The computed values indicate a good-quality Pareto front. As shown in Fig. 2, the Pareto front captures a wide range of optimal solutions. More samples can be generated to further improve the Pareto front.

The 22 MDFFs on the Pareto front are all optimal solutions. Some criteria can be defined to further narrow down the choices. For instance, the selection metrics can be defined as the sum of the errors

$$M_1 = \sum_{i=1}^{n} f_i \tag{12}$$

where M_1 is metric one and n is the number of optimized objectives, the variability in the errors,

$$M_2 = \sum_{i=1}^{n} \sum_{j=i+1}^{n} |f_i - f_j| \tag{13}$$

Table IV. Force field objectives and selection metrics for the three-objective MOBO algorithm

Force field	f_1 where $x_1 = d_{33}^{(23.5)}$	f_5 where $x_2 = d_{33}^{(33.2)}$	f_7 where $x_3 = d_{33}^{(42.7)}$	Metric$_1$	Metric$_2$	Metric$_3$
$FF_{3,1}$	27.32	3.44	7.79	38.55	47.75	86.31
$FF_{3,2}$	19.27	16.38	10.48	46.13	17.58	63.71
$FF_{3,3}$	9.42	2.16	35.13	46.72	65.94	112.66
$FF_{3,4}$	0.05	56.85	28.60	85.50	113.61	199.11
$FF_{3,5}$	0.11	40.21	88.27	128.59	176.33	304.92
$FF_{3,6}$	83.57	1.35	73.41	158.33	164.43	322.76
$FF_{3,7}$	117.30	1.55	66.94	185.78	231.49	417.27
$FF_{3,8}$	43.63	117.64	0.67	161.94	233.95	395.89
$FF_{3,9}$	33.58	104.85	2.03	140.45	205.64	346.09

Table V. Force field objectives for selected MDFFs for six- and eight-objective MOBO

Force Field	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
$FF_{6,1}$	11.12	9.40	–	–	53.11	0.05	13.56	11.16
$FF_{6,2}$	30.68	5.36	–	–	36.45	4.45	28.75	4.65
$FF_{6,3}$	6.47	2.47	–	–	30.92	15.15	47.68	8.26
$FF_{6,4}$	43.95	8.62	–	–	18.34	3.97	24.10	12.42
$FF_{6,5}$	41.34	27.57	–	–	29.12	28.30	29.36	38.89
$FF_{6,6}$	37.53	24.09	–	–	24.94	25.46	36.76	36.79
$FF_{6,7}$	0.11	5.80	–	–	40.21	21.72	88.27	19.53
$FF_{6,8}$	20.32	79.06	–	–	0.08	82.96	37.06	79.14
$FF_{6,9}$	89.37	7.63	–	–	67.83	21.19	0.50	11.47
$FF_{8,1}$	19.45	2.26	1.92	2.72	22.65	6.07	4.04	3.59
$FF_{8,2}$	12.45	9.02	2.60	1.52	53.45	0.42	14.22	10.93
$FF_{8,3}$	2.07	6.94	1.42	3.16	48.27	0.39	42.4	8.26
$FF_{8,4}$	21.24	24.34	1.48	2.65	33.63	24.47	18.36	37.03
$FF_{8,5}$	19.10	13.17	2.00	2.35	26.45	8.73	41.31	18.41
$FF_{8,6}$	0.23	79.28	2.18	8.39	94.65	75.61	46.64	75.25
$FF_{8,7}$	17.70	24.36	1.98	1.58	2.32	27.85	10.64	43.88
$FF_{8,8}$	103.60	10.15	3.29	0.72	97.95	27.35	1.54	19.05

where M_2 is metric two and a combination of the two metrics,

$$M_3 = M_1 + M_2 \qquad (14)$$

where M_3 is metric three. These empirical metrics can be used to rank the MDFFs while maintaining a balance between all objectives.

In this work, the two MDFFs with the lowest metric values for each of the three selection metrics in Eqs. 8–10 are chosen. Due to overlap between the selected MDFFs, a total of three MDFF selections are made, listed as $FF_{3,1}$, $FF_{3,2}$, and $FF_{3,3}$ in Table IV. In addition, MDFFs that have the minimum values of the individual objectives f_1, f_5, and f_7 are also chosen. The MDFFs corresponding to the two lowest values for each of the three objectives are chosen, which are listed as $FF_{3,4}$ through $FF_{3,9}$ in Table IV. The corresponding parameter values of P for each of the MDFFs are tabulated in supplementary Table S-I (refer to online supplementary material).

From Table IV, it is seen that none of the MDFFs can minimize all three objectives. Thus, two out of

the three properties can be predicted accurately, but not all three. Using the three-objective method with the current number of samples, none of the MDFFs were able to predict all the piezoelectric coefficients accurately at the three different Sc dope levels. However, there are MDFFs available that strike a good balance at 23.5% (f_1) and 33.2% (f_5) as well as at 33.2% (f_5) and 42.7% (f_7). There are also suitable MDFFs viable for specific dopant levels. For instance, $FF_{3,1}$ and $F_{3,3}$ can be used in conjunction to capture d_{33} values at 23.5%, 33.2%, and 42.7% dopant levels.

For the six- and eight-objective cases, 1000 MDFFs are generated for each during the search with the MOBO algorithm, which results in 195 and 546 MDFFs on the Pareto front, respectively. The selection metrics in Eqs. 8–10 and individual objectives are similarly used to select MDFFs. These MDFF objectives are tabulated in Table V. For the six-objective case, $FF_{6,1}$, $FF_{6,2}$, $FF_{6,3}$, and $FF_{6,4}$ minimize metric M_1; $FF_{6,4}$ and $FF_{6,5}$ minimize both M_2 and M_3; and $FF_{6,7}$, $FF_{6,8}$, and $FF_{6,9}$ minimize f_1, f_5, and f_7, respectively. For the eight-objective case,

Table VI. Experimental and MD property values for the piezoelectric coefficient, modulus of elasticity, and lattice parameters for 23.5, 33.2, and 42.7% Sc-doped AlN. MD calculations are for six selected MDFFs

Sc %	Property	Experimental[12]	$FF_{3,1}$	$FF_{3,3}$	$FF_{6,7}$	$FF_{6,9}$	$FF_{8,1}$	$FF_{8,7}$
23.5	d_{33}, [pC/N]	20.60	14.97	18.66	20.62	2.19	16.59	16.95
	E, [GPa]	220	174	174	233	237	215	166
	a, [Å]	3.252	3.209	3.136	3.166	3.128	3.190	3.188
	c, [Å]	5.012	5.158	4.962	5.098	5.012	5.148	5.091
	T_m, [K]	–	2300	2800	2400	2600	2400	2300
33.2	d_{33}, [pC/N]	23.60	24.41	23.09	14.11	7.59	18.25	23.05
	E, [GPa]	190	148	140	231	230	202	137
42.7	d_{33}, [pC/N]	27.15	29.26	17.61	3.18	27.29	26.05	24.26
	E, [GPa]	193	122	109	231	215	186	108

Table VII. Piezoelectric coefficients and modulus of elasticities for non-calibrated Sc-doped levels

Sc %	Property	Experimental[12]	$FF_{3,1}$	$FF_{3,3}$	$FF_{6,7}$	$FF_{6,9}$	$FF_{8,1}$	$FF_{8,7}$
29.0	d_{33}, [pC/N]	22	9.03	8.67	0	12.26	1.84	19.59
	E, [GPa]	190	153	159	237	227	209	145
38.0	d_{33}, [pC/N]	24	7.46	16.02	5.51	3.93	7.63	20.37
	E, [GPa]	190	138	124	231	231	196	127

$FF_{8,1}$, $FF_{8,2}$, and $FF_{8,3}$ minimize M_1; $FF_{8,4}$ minimizes M_2; $FF_{8,5}$ minimizes M_3; and $FF_{8,6}$, $FF_{8,7}$, and $FF_{8,8}$ minimize f_1, f_5, and f_7, respectively. The maximum values for f_3 and f_4 out of all 1000 MDFFs are 5.04 and 11.26, respectively. Therefore, most MDFFs predict the lattice constants at the 23.5% dopant level with good accuracy. The inclusion of objectives f_3 and f_4 did not impact the choices of MDFFs.

Similarly, in the six- and eight-objective cases it was not able to minimize d_{33} for all three dopant levels. It is also difficult to minimize both d_{33} and E for the 33.2% dopant level. In contrast, options of MDFFs are available for d_{33} and E for the two other dopant levels.

There are 26 MDFFs to choose from in Tables IV and V. To narrow this list down to preferred MDFFs, two MDFFs from each of the three MOBO algorithms are selected, and their property calculations are tabulated in in Table VI. $FF_{8,1}$ has a strong balance between all of the d_{33} coefficients and has a low error for the E calculations. Due to this, $FF_{8,1}$ is recommended for the best balance between all Sc-doped levels. $FF_{3,3}$ has the best balance between the 23.5 and 33.2% Sc-doped d_{33} calculations and is recommended for piezoelectric applications between 23.5 and 33.2% Sc-doped AlScN. $FF_{3,1}$ has the best balance between the 33.2 and 42.7% Sc-doped d_{33} calculations. If E is of interest, $FF_{6,7}$ and $FF_{8,7}$ perform well between 23.5 and 33.2% Sc-doped. For a strong balance between d_{33} and E values, $FF_{6,7}$, $FF_{8,7}$, and $FF_{6,9}$ are recommended for 23.5, 33.2, and 42.7% Sc-doped levels, respectively. MDFFs for

AlScN that have been developed in this work for different Sc-doped levels for accurate piezoelectric and modulus of elasticity property calculations.

The subset of MDFFs has been examined for only 23.5, 33.2, and 42.7% Sc dope levels. If these MDFFs are to be used at Sc dope levels different from the three tested values, the MDFFs' performance at Sc dope levels between these values must be examined. To do this, the piezoelectric coefficient and modulus of elasticity are examined at different Sc dope levels for the six AlScN MDFFs. The tensile and electric field simulations are performed for 29% and 38% Sc dope levels with the results listed in Table VII. For the piezoelectric coefficient, Fig. 3a shows the $FF_{8,7}$ model accurately trends with the experimental results for the piezoelectric coefficient. For the modulus of elasticity, Fig. 3b shows that $FF_{8,1}$ trends the closest to experimental results.

Though the number of Pareto solutions appears to be much higher for the six- and eight- than the three-objective cases, the solutions are just as sparse in the high-dimensional objective space. To be able to directly compare the quality of Pareto fronts among the three cases, the Pareto quality metrics need to be normalized. The normalized quality metrics are calculated as

$$\widetilde{\mathrm{metric}} = \sqrt[n]{\mathrm{metric}} \qquad (15)$$

for n objectives. The normalized quality metrics calculated from the three cases are tabulated in Table VIII. The normalized DOS and IHD Pareto quality metrics cannot be easily compared for the
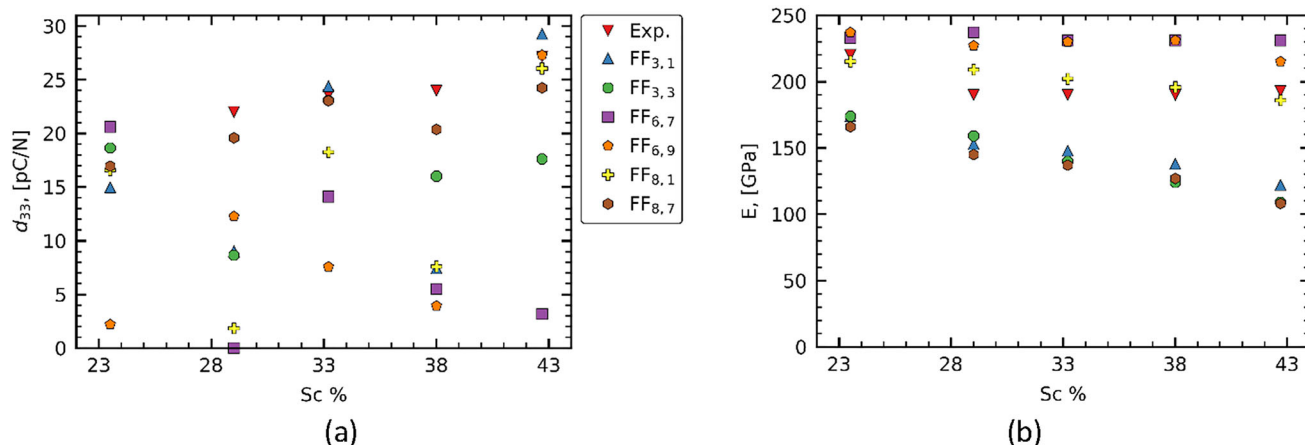
Development of Aluminum Scandium Nitride Molecular Dynamics Force Fields with Scalable Multi-Objective Bayesian Optimization

3495

Fig. 3. (a) Piezoelectric coefficient and (b) modulus of elasticity of the eight selected MDFFs compared to experimental values.

Table VIII. Normalization of Pareto quality metrics for the three MOBOs

Metric	3 Objectives	6 Objectives	8 Objectives
nP	22	195	546
$\widetilde{\text{nP}}$	2.8	2.4	2.2
$\widetilde{\text{DOS}}$	132.2	147.0	83.3
$\widetilde{\text{IHD}}$	59.8	64.8	28.3
$\widetilde{\text{GHD}}$	29.8	68.1	61.5

Fig. 4. Number of objectives in the MOBO algorithm versus the computation time to recommend the next point.

eight-objective case. The low values for f_3 and f_4 skew the normalized values of the Pareto metrics for the eight-objective case.

For the number of Pareto front solutions (nP), as the number of objectives increases, the value for $\widetilde{\text{nP}}$ decreases. This is because higher-dimensional optimization requires more samples to form accurate Pareto fronts. For the normalized Pareto front quality metrics DOS and IHD, there is a slight increase from the three- to six-objective cases and then a significant decrease from the six- to eight-objective cases. The increase from the three- to six-objective cases suggests that the three- and six-objective problems have a similar Pareto front quality. The decrease from the six- to eight-objective cases for $\widetilde{\text{DOS}}$ and $\widetilde{\text{IHD}}$ is due to the inclusion of f_3 and f_4 in the eight-objective problem. The normalized GHD metric is significantly worse between the three- to six-objective problems and performs similarly between the six- and eight-objective problems. With the same number of 1000 samples, the increased six- and eight-objective MOBO algorithms only perform slightly worse than the three-objective MOBO algorithm. This can be mitigated by increasing the number of samples. Overall, the three-, six-, and eight-objective problems are each able to recommend effective MDFFs for accurate piezoelectric
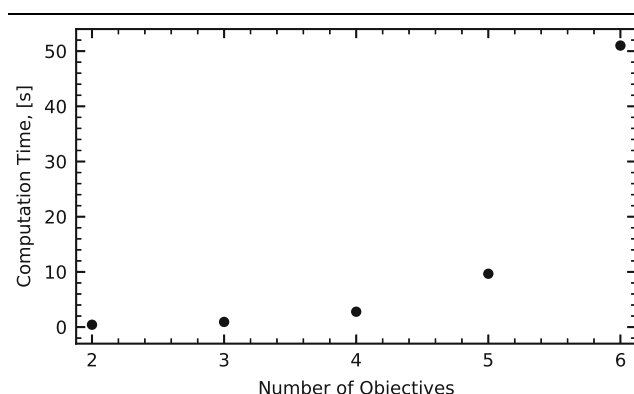
and modulus of elasticity calculations at different Sc dope levels.

The computation time for the MD simulations on 12 processors takes about 19, 42, and 86 s for each equilibration, d_{33} calculation, and E simulation, respectively. The lattice simulation takes about 48 s. This results in a computation time of about 2196, 5292, and 5868 s on one processor for the three-, six-, and eight-objective algorithms, respectively. The computation time to generate the next recommended point using one processor takes about 1.4, 229, and 3,660 s for three-, six-, and eight-objective problems, respectively. The time disparity is due to the increased computation time for higher dimension interpolation and integration. Noticeably, the computational times of the three- and six-objective MOBO algorithms are low to negligible compared to the MD computation time. However, when the number of objectives is increased to eight, the computation time of the MOBO algorithm becomes a significant portion of the overall computation. The increase of computation time can be visualized in Fig. 4, which shows how adding objectives can drastically increase the overall computation time

required to recommend the next set of design values. However, increasing the number of objectives not only increases computational time, but also the number of MDFFs required to form an accurate Pareto front. It is imperative to be mindful of how many objectives to use in the MDFF generation process to ensure the overall algorithm can be run in a reasonable amount of time. This can always be mitigated in the future by optimization and parallelization of the MOBO algorithm. However, the increased computational time will always exist compared to lower objective algorithms.

CONCLUSION

In this work, a MDFF development method was created based on a scalable MOBO scheme, where the calibration of MDFFs can be based on a large number of properties simultaneously. The new DOS-IHD method is applied to develop the MDFFs of AlScN. Up to eight properties, including the piezoelectric d_{33} parameter, modulus of elasticity, and lattice constants, are simultaneously optimized at different Sc dope levels. It has been found that the six- and eight-objective problems perform well, and all three cases can lead to good predictions of the d_{33} value and modulus of elasticity. It is also demonstrated that as the dimension of the problem increases, the quality of the Pareto front can deteriorate if the number of samples remains the same.

As the number of objectives increases, the computational time of the MOBO algorithm also increases. It is shown that for the eight-objective problem, the MOBO algorithm requires a comparable amount of computational time of 3660 s to the MD simulations, which require 5868 s. This suggests that improvements to the MOBO algorithm are necessary for increasing the number of property calculations for which the MDFFs are optimized.

The MD calculations for piezoelectric d_{33} parameter, modulus of elasticity, lattice constants, and melting temperature for this MDFF have been found to be in strong agreement with experimental results, except for the modulus of elasticity, which was underestimated compared to the experimental value. Of the 26 optimized AlScN MDFFs, 6 have been characterized for piezoelectric d_{33} parameters and modulus of elasticities at 23.5, 29.0, 33.2, 38.0, and 42.7% Sc dope levels as well as their lattice constants and melting temperature at 23.5% Sc dope level. All six AlScN MDFFs produced accurate lattice parameters and melting temperatures. Specific AlScN MDFFs are presented for accurate piezoelectric applications at specific Sc-doped levels. For varying Sc-doped levels, one MDFF has been presented for accurate piezoelectric applications and one for accurate modulus of elasticity calculations.

This work lays the groundwork for the optimization of MDFFs to a scalable number of property values. This method can be applied systematically to the development of MDFFs for other material applications. Through this work, a basis of several new AlScN MDFFs is established for utilization in MD studies, such as processes in the synthesis of Sc-AlN compositions and deposition for film growth. When utilizing these MDFFs, it is always important to first characterize the selected MDFF for the specific application. Further work that includes additional properties of this promising material is needed to better understand its potential as a piezoelectric.

DATA AVAILABILITY

Data and code are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1007/s11837-022-05344-7.

REFERENCES

1. S. Yu, P. Hing, and X. Hu, *Compos. Part A Appl. Sci. Manuf.* 33, 289 (2002).
2. S.S. Ogata and H. Kitagawa, *Comput. Mater. Sci.* 15, 435 (1999).
3. K.M. Taylor and C. Lenie, *J. Electrochem. Soc.* 107, 308 (1960).
4. W. Jiang, I.-T. Bae, and W.J. Weber, *J. Phys. Condens. Matter* 19, 356207 (2007).
5. T. Pornphatdetaudom, T. Yano, and K. Yoshida, *Nucl. Mater. Energy* 16, 24 (2018).
6. B. R. Tittmann, B. Reinhardt, and J. Daw, in *AIP Conf. Proc.* (2018), p. 100001.
7. J. Daw, J. Rempe, J. Palmer, P. Ramuhalli, R. Montgomery, H. Chien, B. Tittmann, B. Reinhardt, and P. Keller, *NEET In-Pile Ultrasonic Sensor Enablement-Final Report* (Idaho Falls, ID (United States), 2014).

8. G. Piazza, P.J. Stephanou, and A.P. Pisano, *J. Microelectromechanical Syst.* 15, 1406 (2006).

9. M.A. Dubois and P. Muralt, *Appl. Phys. Lett.* 74, 3032 (1999).

10. N.L. Marana, G.B. Pinhal, J.A.S. Laranjeira, P.G.C. Buzolin, E. Longo, and J.R. Sambrano, *Comput. Mater. Sci.* 177, 109589 (2020).

11. M. Akiyama, T. Kamohara, K. Kano, A. Teshigahara, Y. Takeuchi, and N. Kawahara, *Adv. Mater.* 21, 593 (2009).

12. O. Zywitzki, T. Modes, S. Barth, H. Bartzsch, and P. Frach, *Surf. Coatings Technol.* 309, 417 (2017).

13. M. Park, Z. Hao, R. Dargis, A. Clark, and A. Ansari, *J. Microelectromechanical Syst.* 29, 490 (2020).

14. S. Leone, J. Ligl, C. Manz, L. Kirste, T. Fuchs, H. Menner, M. Prescher, J. Wiegert, A. Žukauskaitė, R. Quay, and O. Ambacher, *Phys. Status Solidi Rapid Res. Lett.* 14, 1900535 (2020).

15. M. Moreira, J. Bjurström, I. Katardjev, and V. Yantchev, *Vacuum* 86, 23 (2011).

16. K. Dukenbayev, A. Kozlovskiy, Z.A. Alyamova, T. Gladkikh, I. Kenzhina, and M. Zdorovets, *J. Mater. Sci. Mater. Electron.* 30, 8777 (2019).

17. D. Parks and B. Tittmann, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 61, 1216 (2014).

18. W. J. Weber, F. Gao, R. Devanathan, W. Jiang, and C. M. Wang, in *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* (2004), pp. 25–35.

19. J. C. Nappé, M. Benabdesselam, P. Grosseau, and B. Guilhot, in *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **269**, 100 (2011).

20. B. Reinhard, B.R. Tittmann, and A. Suprock, *Phys. Procedia* 70, 609 (2015).

21. M.J. Rahman, B. Szpunar, and J.A. Szpunar, *Comput. Mater. Sci.* 166, 193 (2019).

22. C. Shan, L. Lang, T. Yang, Y. Lin, F. Gao, H. Deng, and W. Hu, *Comput. Mater. Sci.* 177, 109555 (2020).

23. J.T. Buchan, M. Robinson, H.J. Christie, D.L. Roach, D.K. Ross, and N.A. Marks, *J. Appl. Phys.* 117, 245901 (2015).

24. J. Fu, Y. Chen, J. Fang, N. Gao, W. Hu, C. Jiang, H.B. Zhou, G.H. Lu, F. Gao, and H. Deng, *J. Nucl. Mater.* 524, 9 (2019).

25. J. Fikar and R. Schaeublin, in *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **255**, 27 (2007).

26. E. Zarkadoula, G. Samolyuk, Y. Zhang, and W.J. Weber, *J. Nucl. Mater.* 540, 152371 (2020).

27. P. Vashishta, R.K. Kalia, A. Nakano, and J.P. Rino, *J. Appl. Phys.* 109, 033514 (2011).

28. M. Tungare, Y. Shi, N. Tripathi, P. Suvarna, and F. Shahedipour-Sandvik, *Phys. Status Solidi Appl. Mater. Sci.* 208, 1569 (2011).

29. K. Choudhary, T. Liang, K. Mathew, B. Revard, A. Chernatynskiy, S.R. Phillpot, R.G. Hennig, and S.B. Sinnott, *Comput. Mater. Sci.* 113, 80 (2016).

30. J. Martinez, T. Liang, S.B. Sinnott, and S.R. Phillpot, *Comput. Mater. Sci.* 139, 153 (2017).

31. S.R. Phillpot, A.C. Antony, L. Shi, M.L. Fullarton, T. Liang, S.B. Sinnott, Y. Zhang, and S.B. Biner, *Comput. Mater. Sci.* 148, 231 (2018).

32. J.L. McDonagh, A. Shkurti, D.J. Bray, R.L. Anderson, and E.O. Pyzer-Knapp, *J. Chem. Inf. Model.* 59, 4278 (2019).

33. A. Dequidt and J.G.S. Canchaya, *J. Chem. Phys.* 143, 084122 (2015).

34. P. Liu, Q. Shi, H. Daumé, and G.A. Voth, *J. Chem. Phys.* 129, 214114 (2008).

35. J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N.E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, *ACS Cent. Sci.* 5, 755 (2019).

36. P. Gkeka, G. Stoltz, A.B. Farimani, Z. Belkacemi, M. Ceriotti, J.D. Chodera, A.R. Dinner, A.L. Ferguson, J.B. Maillet, H. Minoux, C. Peter, F. Pietrucci, A. Silveira, A. Tkatchenko, Z. Trstanova, R. Wiewiora, and T. Lelièvre, *J. Chem. Theory Comput.* 16, 4757 (2020).

37. S. Chmiela, H.E. Sauceda, I. Poltavsky, K.R. Müller, and A. Tkatchenko, *Comput. Phys. Commun.* 240, 38 (2019).

38. J.M. Sestito, M.L. Thatcher, L. Shu, T.A.L. Harris, and Y. Wang, *J. Phys. Chem. A* 124, 5042 (2020).

39. S. Plimpton, *J. Comput. Phys.* 117, 1 (1995).

40. C. Chen, Z. Shang, J. Gong, F. Zhang, H. Zhou, B. Tang, Y. Xu, C. Zhang, Y. Yang, and X. Mu, *ACS Appl. Mater. Interfaces* 10, 1819 (2018).

41. H. Altuntas, C. Ozgit-Akgun, I. Donmez, and N. Biyikli, *J. Appl. Phys.* 117, 155101 (2015).

42. G.B.M. Cervera, and G. Lombera, *Rapid Prototyp. J.* 5, 21 (1999).

43. R.E. Rozas, A.D. Demirağ, P.G. Toledo, and J. Horbach, *J. Chem. Phys.* 145, 064515 (2016).

44. E. Schwegler, M. Sharma, F. Gygi, and G. Galli, *Proc. Natl. Acad. Sci. U. S. A.* 105, 14779 (2008).

45. J.R. Morris and X. Song, *J. Chem. Phys.* 116, 9352 (2002).

46. J.R. Morris, C.Z. Wang, K.M. Ho, and C.T. Chan, *Phys. Rev. B* 49, 3109 (1994).

47. S.W. Watt, J.A. Chisholm, W. Jones, and S. Motherwell, *J. Chem. Phys.* 121, 9565 (2004).

48. H. Xiang, H. Li, and X. Peng, *Comput. Mater. Sci.* 140, 113 (2017).

49. P.S. Branicio, J.P. Rino, C.K. Gan, and H. Tsuzuki, *J. Phys. Condens. Matter* 21, 095002 (2009).

50. L. Shu, P. Jiang, X. Shao, and Y. Wang, *J. Mech. Des.* 142, 1 (2020).

51. S. Cheng, J. Zhou, and M. Li, *J. Mech. Des.* 137, 1 (2015).

52. D. Zhan, Y. Cheng, and J. Liu, *IEEE Trans. Evol. Comput.* 21, 956 (2017).

53. H. Ohno, and S.N. Comput, *Sci.* 2, 214 (2021).

54. S.F. Ghoreishi and D. Allaire, *Struct. Multidiscip. Optim.* 59, 977 (2019).

55. A. Tran, M.S. Eldred, S. McCann, and Y. Wang, *J. Mech. Des.* 144, 1 (2021).

56. J. Park, *J. Stat. Plan. Inference* 39, 95 (1994).

57. G.G. Wang, *J. Mech. Des. Trans. ASME* 125, 210 (2003).

58. PACE, (2017), http://www.pace.gatech.edu. Accessed 1 December 2021.

59. C.M. Lueng, H.L.W. Chan, C. Surya, and C.L. Choy, *J. Appl. Phys.* 88, 5360 (2000).

60. F. Bernardini and V. Fiorentini, *Appl. Phys. Lett.* 80, 4145 (2002).

61. E. G. Gillan, in *Comprehensive Inorganic Chemistry II* (Elsevier, 2013), pp. 969–1000.